7-16-2004

Search for $\mathbf{B_s^0 \mu^+ \mu^-}$ and $\mathbf{B_d^0 \mu^+ \mu^-}$ Decays in $\mathbf{p\bar{p}}$ Collisions at $\sqrt{s} = 1.96$ TeV

Darin Acosta
University of Florida, acosta@phys.ufl.edu

Kenneth A. Bloom
University of Nebraska-Lincoln, kbloom2@unl.edu

Collider Detector at Fermilab Collaboration

Follow this and additional works at: http://digitalcommons.unl.edu/physicsbloom

Part of the Physics Commons

Acosta, Darin; Bloom, Kenneth A.; and Fermilab Collaboration, Collider Detector at, "Search for $\mathbf{B_s^0 \mu^+ \mu^-}$ and $\mathbf{B_d^0 \mu^+ \mu^-}$ Decays in $\mathbf{p\bar{p}}$ Collisions at $\sqrt{s} = 1.96$ TeV" (2004). Kenneth Bloom Publications. 38.
http://digitalcommons.unl.edu/physicsbloom/38

This Article is brought to you for free and open access by the Research Papers in Physics and Astronomy at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Kenneth Bloom Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Optimized search for single-top-quark production at the Fermilab Tevatron


(CDF Collaboration)

1Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China
2Argonne National Laboratory, Argonne, Illinois 60439, USA

0556-2821/2004/69(5)/052003(9)/$22.50 052003-1 ©2004 The American Physical Society
We use a neural-network technique to search for standard model single-top-quark production in the 106 pb$^{-1}$ dataset accumulated by the Collider Detector at Fermilab detector during the 1992–1995 collider run ("run I"). Using a sample of 64 $W^{\pm}+1, 2, 3$ jets events, we set a 95% confidence level upper limit of 24 pb on the $W$-gluon and $W^*$ combined single-top cross section.

DOI: 10.1103/PhysRevD.69.052003 PACS number(s): 14.65.Ha, 12.15.Ji, 13.85.Rm, 87.18.Sn
At the Fermilab Tevatron, top quarks produced in pairs through the strong interaction were observed [1,2]. Within the standard model, top quarks are also expected to be produced singly in the electroweak channel [3], mainly through off mass shell W production ("W*") and W-gluon fusion ("Wg") processes, shown in Fig. 1. The measurement of single-top events is of particular interest because the production cross section is proportional to $|V_{tb}|^2$, where $V_{tb}$ represents the Cabibbo-Kobayashi-Maskawa matrix element relating top and bottom quarks. Assuming $|V_{tb}| = 1$, the next-to-leading order predicted cross sections at $\sqrt{s} = 1.8$ TeV for $W^*$ and Wg channels are 0.76 pb and 1.40 pb, respectively [4]. The DØ Collaboration has published upper limits on single-top production of 22 pb on Wg and 17 pb on W*, both at a 95% confidence level (C.L.) [5]. The Collider Detector at Fermilab (CDF) Collaboration reported lower 95% C.L. limits: 13 pb and 18 pb on the Wg and W* cross sections respectively, and 14 pb for the combined cross section as determined in a separate analysis [6]. In this paper we report on a search for the combined W* and Wg single-top production using a neural-network technique to maximize the discriminating power of seven kinematic variables. This technique is expected to be more sensitive than the method employed in [6]. In addition to using a larger amount of information, the analysis also features marginally higher signal purity obtained by retuning the event selection. The improvement in the average expected upper limit on the single-top cross section is 20% if the SM signal cross section is assumed.

The final state of the W* channel features two $b$-quarks and the decay products of the W boson. Similarly, the Wg channel is characterized by two $b$-quarks and the W decay products plus an additional light quark jet ($u,d$). In addition, initial and final state radiation can increase the jet content of the final state. Our analysis will focus on the channels with lepton+$+$jets" events that we can study using many of the tools developed for the CDF top pair production ($t\bar{t}$) cross section analysis [7].

This analysis uses the data from $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV collected with the Collider Detector at Fermilab between 1992 and 1995. A thorough description of the detector is provided elsewhere [8]. We select the events having an isolated electron (muon) with transverse energy $E_T > 20$ GeV (transverse momentum $p_T > 20$ GeV/c), and missing transverse energy $E_T > 20$ GeV [9]. The $t\bar{t}$ or Z boson decays are removed by rejecting events containing an additional isolated track with $p_T > 15$ GeV/c and charge opposite to that of the primary lepton [10]. Also rejected are Z candidates in which there are two opposite-charge leptons with invariant mass between 75 and 105 GeV/c$^2$. We further require that there are one, two, or three jets with $E_T > 15$ GeV and pseudorapidity $|\eta| < 2.0$ ("tight" jets) in the event. At least one of these jets should be associated with a $b$-quark decay ("B-tagged") as determined by observing a displaced vertex using tracks reconstructed in the silicon vertex detector (SVX) [11]. After these initial selections, the backgrounds can be classified as non-top (mostly QCD multijet) and $t\bar{t}$ production.

We further reduce backgrounds by exploiting the distributions of "soft" jets in the event. These are jets with $E_T > 8$ GeV and $|\eta| < 2.4$ which do not pass the above tight jet criteria. Tight and soft jet multiplicities are denoted by $N_{jt}$ and $N_{js}$. We use $N_{jt}$ to define and label the jet multiplicity bins $W+N_{jt}$ jets. For example, a $W+3$ jets event contains exactly three tight jets and possibly additional soft jets. Figure 2 shows the $N_{jt}$ versus $N_{jt}$ Monte Carlo distributions for $W^*, Wg$, non-top, and $t\bar{t}$ processes. The PYTHIA Monte Carlo program [12] was used, followed by the CDF detector simulation. Optimal signal to background ratio is obtained by demanding $N_{jt} = 1$ in the $W+1$ jet events, and $N_{jt} = 0$ in the $W+3$ jets events. There is no $N_{jt}$ requirement for the $W+2$ jets events. As shown in Table I, the soft jets requirements remove over 50% of the non-top and 40% of the $t\bar{t}$ events passing initial selections. If we assume the theoretical $W^*$ and Wg cross sections [4] we arrive at the signal contributions listed in Table II. The expected numbers of $t\bar{t}$ and

![FIG. 1. Representative Feynman diagrams for single-top-quark production at the Fermilab Tevatron: s-channel W* (left) and t-channel W-gluon fusion (right).](image)

![FIG. 2. $N_{jt}$ versus $N_{jt}$ distribution for simulated signal and background events passing the initial selection described in the text. Non-top backgrounds are suppressed by requiring $N_{jt} = 1$ for $W+1$ jet events ($N_{jt} = 1$). We reduce the $t\bar{t}$ background by requiring $N_{jt} = 0$ for $W+3$ jets events ($N_{jt} = 3$).](image)
non-top events are also given in Table II. The $t\bar{t}$ expectation is obtained using a PYTHIA Monte Carlo calculation normalized to the theory prediction $\sigma_{t\bar{t}}=5.1\pm0.9$ pb [13]. For the non-top background, the primary source (approx. 65%) is the $W$+heavy flavor production process $\bar{q}q'\rightarrow Wg$ with $g\rightarrow b\bar{b}$, $cc'$, and $gq\rightarrow Wq'$ [11]. Other sources include “mistags” (17%), where a light-flavor jet is misidentified as heavy flavor jet, direct $b\bar{b}$ production (11%), $Z$+heavy flavor and $Z\rightarrow \tau\tau$ (5%), and also diboson processes $WW$, $WZ$ (2%). The non-top expectations are based on the calculation performed in the previous CDF single-top analysis [6] which we correct for differences in the selection criteria. To estimate the shape of the non-top background kinematic distributions we use a PYTHIA generated sample of $W$+heavy flavor events.

The estimated signal and background contributions outlined above can be combined to predict a signal to noise ratio of 1/13, which implies a challenging search. We maximize our discriminating power by employing an Artificial Neural Network (ANN) technique [14]. ANN’s employ information from several kinematic variables while accounting for the correlations among them. The goal is to design an ANN to classify events in one of three categories: single-top ($W^*$ and $Wg$), $t\bar{t}$, and non-top. We do not attempt to distinguish between $W^*$ and $Wg$ signal events, as most of the kinematic distributions considered in this analysis are very similar for the two processes (see Fig. 3). The differences between the two signal channels are accommodated by training and testing the network with $W^*$ and $Wg$ events in the proportion expected from SM (Table II). We will subsequently demonstrate that our method is rather insensitive to the precise $W^*$ -- $Wg$ mixing proportion within a range of $\pm50\%$ of its SM value.

The network is a feed-forward perceptron with one intermediate (hidden) layer and three output nodes. The advantages of using one output node for each class of events are detailed in Ref. [15]. For training we use 30000 Monte Carlo events, and require an output of (0,1,0) for signal, (0,0,1) for $t\bar{t}$, and (1,0,0) for non-top background. The weights are updated according to the “Manhattan” algorithm in JETNET [16] with default parameters.

To select the inputs of the ANN, we started from a set of 18 variables with good signal-background separation potential [6,17,18]: $E_T^{\ell\nu}$, $E_T^{\ell\nu}$, $E_T^\tau$, $H_T$, $\sqrt{s}$, $M^{\ell\nu}$, $M^{\ell\nu}$, $P_T^\ell$, $P_T^\ell$, $P_T^\ell$, $Q\times \eta$, $\cos(\phi)$, $R_{min}$, $N_{j\ell}$, $N_{j\ell}$, $N_{B-tags}$. Here $j_1$ and $j_2$ are the leading jets in the event, $H_T$ is the total transverse energy defined as $E_T+\sum E_T^\nu$ where the last term includes both the tight and the soft jets, $\sqrt{s}$ is the total energy in the center-of-mass system, $\ell\nu$ refers to the lepton, neutrino, and leading $B$-tagged jet system, $jj$ refers to the $j_1$ -- $j_2$ system, $Q\times \eta$ is the product between the primary

---

**TABLE I.** $N_{j\ell}$ cut efficiencies for signal and background. $\epsilon_{N_{j\ell}}(W+1\text{ jet})$ represents the fraction of $W+1\text{ jet}$ events with $N_{j\ell}=1$, after the initial selections were imposed. Similarly, $\epsilon_{N_{j\ell}}$ is the fraction of $W+1\text{ jet}$, 2, and 3 jet events passing the $N_{j\ell}$ selections. The overall $\epsilon_{tot}$ results from multiplying the efficiencies of the initial and the $N_{j\ell}$ selections.

<table>
<thead>
<tr>
<th>Efficiency</th>
<th>$W^*$</th>
<th>$Wg$</th>
<th>non-top</th>
<th>$t\bar{t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_{N_{j\ell}}(W+1\text{ jet})$</td>
<td>43.4%</td>
<td>39.7%</td>
<td>23.9%</td>
<td>42.7%</td>
</tr>
<tr>
<td>$\epsilon_{N_{j\ell}}(W+3\text{ jets})$</td>
<td>72.9%</td>
<td>75.2%</td>
<td>73.5%</td>
<td>42.8%</td>
</tr>
<tr>
<td>Combined $\epsilon_{N_{j\ell}}$</td>
<td>83.6%</td>
<td>74.1%</td>
<td>47.6%</td>
<td>59.7%</td>
</tr>
<tr>
<td>Overall $\epsilon_{tot}$</td>
<td>2.4%</td>
<td>1.6%</td>
<td>0.02%</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

---

**FIG. 3.** Monte Carlo distributions for the seven variables used in the ANN. In the left plots, the open (shaded) histograms correspond to the $W^*$ ($Wg$) channel. Similarly, in the middle-column plots open (shaded) histograms correspond to $t\bar{t}$ ($W+\text{ jets}$). All Monte Carlo distributions are normalized to unit area for comparison. The histograms in the right column correspond to the run 1 data events.
lepton charge and the pseudorapidity of the highest-$E_T$ untagged jet ($q$), $\hat{E}_q$ is the angle between the direction of the lepton and that of the $q$ jet, and $R_{\text{min}}$ is the minimum separation $\sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$ among all possible pairs of jets in the event. We considered a large number of combinations of variables that can be drawn from this 18-variable set. For each combination we minimized a typical mean squared error function [15]:

$$E = \frac{1}{N} \sum_{k=1}^{N} |\hat{O}^k - T^k|^2$$

where $k$ is the event index, $T^k$ is the target output corresponding to the correct event category, and $\hat{O}^k$ is the actual ANN output. For the input combinations having the lowest error function values we calculate the expected average upper limit on the single-top cross section. The lowest limit is obtained for the following input set: $E_T^j$, $E_T^q$, $E_T^l$, $P_T^j$, $H_T$, and $Q \times \eta$. The distributions of these variables are shown in Fig. 3. We note that the two backgrounds $tt\bar{t}$ and non-top are kinematically situated on different sides of the signal. Finally, in the range of 7–20 nodes in the intermediate layer, the error $E$ has a weak minimum for 17 hidden nodes. The 7-17-3 configuration of nodes corresponds to 190 free parameters adjusted by training. As shown in Ref. [15], the output nodes $O_1$, $O_2$, $O_3$ estimate the Bayesian posterior probabilities for the three classes of events: non-top, signal, and $tt\bar{t}$, respectively. This implies that $O_1 + O_2 + O_3 = 1$, so that all events tend to lie in the same plane in the output space. We indeed found that the output sum peaks at 1.0 with a maximum deviation of 0.1 for the three Monte Carlo samples. Consequently, we reduce the output space to two dimensions $(x, y)$ by projecting all output points onto the plane of equation $O_1 + O_2 + O_3 = 1$, as shown in Fig. 4. The $(x, y)$ signal and background distributions are presented in Fig. 5, along with the data. We employ a maximum likelihood fit to these distributions to estimate the signal content of the run I dataset. We note that Fig. 5 shows improved separation between signal and background compared to the individual input variables of Fig. 3. To quantify this separation one can for example define a “signal region” as the locus of the output points with $O_2 > O_1, O_3$. This signal region contains 67% of the signal, 27% of the non-top, and 24% of the $tt\bar{t}$ Monte Carlo events, respectively.

The performance of this method is tested a priori by constructing simulated experiments using Monte Carlo generated event samples (“pseudo-run I” datasets). A simulated experiment contains $N_s$ signal, $N_{nt}$ non-top, and $N_{tt}\bar{t}$ $tt\bar{t}$ events, where the number of events in each category is drawn from a Poisson distribution using the expected mean values in Table II. We propagate these events through the network and form the $(x, y)$ output distribution. The latter is fitted using a background-constrained binned likelihood:

$$\mathcal{L}(n_s, n_{nt}, n_{tt\bar{t}}) = \mathcal{L}_{\text{background}} \times \mathcal{L}_{\text{shape}}$$

$$= G_1(n_{nt}) G_2(n_{tt\bar{t}}) \prod_{i=1}^{N_{\text{bins}}} \frac{e^{-n_i} n_i^{d_i}}{d_i!}$$

where $n_s$, $n_{nt}$, $n_{tt\bar{t}}$ are the parameters of the fit, representing the numbers of signal, non-top, and $tt\bar{t}$ events respectively present in the sample. Moreover, $n_i = n_{i,nt} + n_{i,tt\bar{t}}$ is the expected number of events in the $i$th bin, and $f_{i,nt}$, $f_{i,tt\bar{t}}$ are the fractions of Monte Carlo single-top, non-top, or $tt\bar{t}$ appearing in bin $i$. By $d_i$ is denoted the number of events in the simulated event that populate the $i$th bin. The Gaussian functions $G_1(n_{nt})$, $G_2(n_{tt\bar{t}})$ constrain the non-top and $tt\bar{t}$ backgrounds to the expected values: 43.3 ± 8.4 non-top and 7.4 ± 2.2 $tt\bar{t}$ events, respectively.

Different scenarios regarding signal expectation were also investigated. Specifically, we considered signal cross sections ranging from 0 pb to 20 pb. For each case, we performed 10000 simulated experiments. In Fig. 6 we show the
The mean values of $n_s$ along with the 16 and 84 percentile points are presented in Fig. 7. We note that the mean of the fitted cross sections is consistent with the input cross section for all cases. We further tested the sensitivity of our method to the particular ratio of $W_g$ and $W^*$ cross sections ($R_{W_g/W^*}$). Two situations were considered: $\sigma_s = 2$ pb and $\sigma_s = 10$ pb. Simulated experiments were constructed with one of seven different values of $R_{W_g/W^*}$, but fitted to the standard templates of Fig. 5. The results are shown in Fig. 8, and show that the mean of the fitted cross sections varies by less than 11% across the $R_{W_g/W^*}$ range studied.

The systematic uncertainties for this analysis are divided into two groups. The first group consists of systematic effects which modify only the rates of events accepted, and not the shapes of the distributions of input variables. The luminosity of $10^6$ pb$^{-1}$ has an uncertainty of 4.1% [19]. The uncertainty on the trigger and lepton identification efficiency has been estimated to be 10%. Moreover, the efficiency for identifying jets containing $B$-hadrons has an uncertainty of 10% [7]. These uncertainties can be expressed in number of events by simply multiplying by the particular single-top content (Table III).

The second group of systematic uncertainties includes the effects that impact both the shapes of the Monte Carlo templates of Fig. 5 and the rates of events accepted. To illustrate how these systematics are extracted, let us consider the uncertainty associated with the signal generator (SG). We start by generating new $W^*$ and $W_g$ samples using the HERWIG [20] program instead of PYTHIA. Among the differences between the two generators, we note the hadronization approach and the underlying event modeling. The new samples are run through the ANN, and simulated experiments are constructed based on the recalculated acceptances and output shapes. Each experiment is then fitted to the standard templates of Fig. 5. We define the uncertainty $\delta_{SG}$ as the absolute value of the shift in the mean fitted signal contribution $n_s$. 

FIG. 6. Results from simulated experiments with input single-top cross sections of 2 pb (top) and 10 pb (bottom). In both cases the mean fitted number of signal events $n_s$ agrees with the expected value $N_s^{exp}$.

FIG. 7. Test of the ANN fitting technique under different hypotheses for signal cross section. As in Fig. 6, we note good agreement between the input and fitted signal cross sections. The theoretically calculated value is $\sigma_s^{SM} = 2.2$ pb [4]. The ends of the error bars mark the 16 and 84 percentile points for each fitted $\sigma_s$ distribution.

FIG. 8. Test of the ANN fitting technique under different hypotheses for $W_g$ to $W^*$ cross section ratio $R_{W_g/W^*}$. This ratio is expressed as the fraction $f$ of the SM value $R_{W_g/W^*}^{SM} = 1.8$. Two values for the combined signal cross section are considered: $\sigma_s = 2$ pb and $\sigma_s = 10$ pb.
TABLE III. Systematic uncertainties (in number of events). The second column corresponds to the theoretical prediction $\mu_{SM}=3.9$ signal events. The third column lists the uncertainties estimated at the measured value $n_s=23.9$ events. The overall uncertainties $\delta s^{\text{norm}}$ and $\delta s^{\text{shape}}$ are obtained by adding in quadrature the individual effects.

<table>
<thead>
<tr>
<th></th>
<th>$\delta s^{\text{norm}}$</th>
<th>$\delta s^{\text{shape}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminosity</td>
<td>0.16</td>
<td>0.98</td>
</tr>
<tr>
<td>Trigger and lepton identification</td>
<td>0.39</td>
<td>2.39</td>
</tr>
<tr>
<td>$B$-tag efficiency</td>
<td>0.39</td>
<td>2.39</td>
</tr>
<tr>
<td>Total $\delta s^{\text{norm}}$</td>
<td>0.57</td>
<td>3.52</td>
</tr>
</tbody>
</table>

Shape and normalization effects

<table>
<thead>
<tr>
<th></th>
<th>$\delta s^{\text{BG}}$</th>
<th>$\delta s^{\text{ISR}}$</th>
<th>$\delta s^{\text{PDF}}$</th>
<th>$\delta s^{\text{JES}}$</th>
<th>$\delta s^{\text{FSR}}$</th>
<th>$\delta s^{\text{IIS}}$</th>
<th>$\delta s^{\text{JES}}$</th>
<th>$\delta s^{\text{IIS}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal generator ($\delta s^{\text{SG}}$)</td>
<td>0.12</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background generator ($\delta s^{\text{BG}}$)</td>
<td>0.15</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jet energy measurement ($\delta s^{\text{JES}}$)</td>
<td>1.49</td>
<td>2.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial and final state radiation ($\delta s^{\text{FSR}}$)</td>
<td>0.51</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parton distribution functions ($\delta s^{\text{PDF}}$)</td>
<td>0.16</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top quark mass ($\delta s^{\text{M_{top}}}$)</td>
<td>0.17</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total $\delta s^{\text{shape}}$</td>
<td>1.59</td>
<td>3.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The uncertainty $\delta s^{\text{BG}}$ related to the background generator is similarly calculated. In this case, the non-top sample was a mixture of two subsamples HERWIG $Wb\bar{b}$, PYTHIA $Wc\bar{c}$, and $Wc$, while the $t\bar{t}$ background was generated with HERWIG.

Figure 9 shows a comparison between the HERWIG $Wb\bar{b}$ events and the default $W+$jets sample generated with PYTHIA. A good level of agreement regarding the shapes of the kinematic distributions can be observed. We note that $\delta s^{\text{BG}}$ accounts for a small fraction of the total $\delta s^{\text{shape}}$. As shown in Table III, the largest contribution to $\delta s^{\text{shape}}$ comes from the uncertainty in the measurement of jet momenta $\delta s^{\text{JES}}$. A change in the jet momentum scale simultaneously impacts five of the seven kinematic variables used in our analysis, which can lead to significant changes on an event by event basis. As detailed in Ref. [9], we apply $+1\sigma$ and $-1\sigma$ shifts in the $P_T$ scale of the jets, and define $\delta s^{\text{JES}}$ as the average difference: $(\delta s + 1\sigma - \delta s - 1\sigma)/2$. To study the uncertainty associated to the initial state radiation (ISR) we turn off ISR in PYTHIA and regenerate signal and background samples. We take $\delta s^{\text{ISR}}$ to be one half the shift in the mean fitted signal contribution. To isolate the effects of final state radiation (FSR) we start from the no_ISR PYTHIA samples and select the (no_ISR, no_FSR) subset of events in which every jet matches to a final state parton within a $(\eta, \phi)$ distance of 0.4. The uncertainty $\delta s^{\text{FSR}}$ is defined to be $(\delta s^{\text{ISR,FSR}} - \delta s^{\text{ISR}})/2$. Combined systematic uncertainty $\delta s^{\text{JES}}$ on the initial and final state radiation is obtained by adding in quadrature $\delta s^{\text{ISR}}$ and $\delta s^{\text{FSR}}$. We evaluate the uncertainty $\delta s^{\text{PDF}}$ due to the parton distribution function set by switching to the CTEQ 3L [21] set from the default GRV 94L [22] choice in PYTHIA. The last systematic effect studied is the top quark mass. We vary the top quark mass from the default $M_{top}=175$ GeV to 170 and 180 GeV respectively, and generate new $W^+$, $Wg$, and $t\bar{t}$ samples. We take $\delta s^{\text{M_{top}}}$ to be the larger of the shifts $\delta s^{\text{170}}$ and $\delta s^{\text{180}}$.

Finally, the magnitude of the systematic uncertainties depends on the particular signal content used in performing simulated experiments. To exemplify this, let us consider the jet energy scale effect, which accounts for the largest fraction of the total $\delta s^{\text{shape}}$. The variation of $\delta s^{\text{JES}}$ with the input signal mean $n_s$ is presented in Fig. 10, where the fit shown is a parabola. Consequently, the values listed in the second column of Table III ($n_s=3.9$ events) will be used in deriving the $a\text{ priori}$ single-top results, while the third column values ($n_s=23.9$ events) will be used in expressing the signal cross section measured from the CDF data.

Simulated experiments based on the SM expectations of Table II result in a distribution of $n_s$ having a mean of 3.9 signal events and standard deviation of 5.9 events. Given this significant uncertainty, we focus on calculating the expected limit for single-top production, using a standard Bayesian procedure. For each simulated experiment, $L(n_s, n_{\mu}, n_{\tau})$ is

![FIG. 9. Distributions of four of the ANN input variables for HERWIG $Wb\bar{b}$ events (open histograms) and the default PYTHIA $W+$jets sample (shaded histograms). All histograms are normalized to unit area for comparison.](image-url)
are integrated out to yield the posterior density $p_M l$ the
obtain the probability density $L n$
widths equal to 6
grated to obtain the 95% C.L. limit $\sim 64$ events pass the selection criteria 14%.
the individual $L p$ends on the parameters in the likelihood the given simulated experiment, $d$
integrated out with respect to $n nt$
ous CDF combined single-top study $n s$
mean expected $n s$
0. In addition to $n s$, the density $L(n_s)$ implicitly depends on the parameters in the likelihood $L_{shape}$ that pertain to normalization and shape uncertainties. These parameters are accounted for by using Gaussian priors of unit means and widths equal to $\delta s^{norm}$ and $\delta s^{shape}$, respectively, and they are integrated out to yield the posterior density $p(n_s)$. For the given simulated experiment, $p(n_s)$ is numerically integrated to obtain the 95% C.L. limit on $n_s$. The mean value of the individual $n s$'s distribution is 10.6 pb and defines the mean expected (or “a priori”) limit on the single-top cross section in the presence of the signal. Compared to the previous CDF combined single-top study [6], the neural-network method features an improvement of 21% in the a priori confidence limit. Roughly 7% of this improvement comes from returning the selection criteria, with $N_{js}$ selection replacing the $M_{1b}$ window cut. Using a multivariate technique (seven variables rather than $H_T$ alone) accounts for the remaining 14%.

We have applied this method to the run I dataset, where 64 events pass the selection criteria (Table II). The overlap with the 65-event sample of the search reported in Ref. [6] is 35 events. Figure 5 shows the distribution of data events in the $O_1 + O_2 + O_3 = 1$ plane. We maximize the likelihood of Eq. (2) to extract a signal contribution of $n_s = 23.9 \pm 7.7$(stat) $\pm 4.7$(syst) events, or equivalently $13.5 \pm 5.1$ pb, including systematic uncertainties. This can be compared to the expected value of 2.2 pb. It can be seen in Fig. 5 that a significant fraction of the data events is indeed consistent with the simulated signal distribution. The numbers of back-

![Graph 1](image1.png)

**FIG. 10.** Jet energy scale systematics as a function of signal content $n_s$. The circles show the shifts $\delta s^{+1\sigma}$ in the mean fitted signal contribution for $a+1\sigma$ increase in jet transverse momenta [9]. The squares represent the shifts $\delta s^{-1\sigma}$, while the triangles correspond to the combined $\delta s^{\text{AS}}$ defined as $(\delta s^{+1\sigma} - \delta s^{-1\sigma})/2$.

integrated out with respect to $n_{st}$, $n_{ij}$ for all values $n_s$ to obtain the probability density $\mathcal{L}(n_s)$. We further assume a uniform prior distribution, and restrict to the physical range $n_s>0$. In addition to $n_s$, the density $\mathcal{L}(n_s)$ implicitly depends on the parameters in the likelihood $\mathcal{L}_{shape}$ that pertain to normalization and shape uncertainties. These parameters are accounted for by using Gaussian priors of unit means and widths equal to $\delta s^{norm}$ and $\delta s^{shape}$, respectively, and they are integrated out to yield the posterior density $p(n_s)$. For the given simulated experiment, $p(n_s)$ is numerically integrated to obtain the 95% C.L. limit on $n_s$. The mean value of the individual $n s$'s distribution is 10.6 pb and defines the mean expected (or “a priori”) limit on the single-top cross section in the presence of the signal. Compared to the previous CDF combined single-top study [6], the neural-network method features an improvement of 21% in the a priori confidence limit. Roughly 7% of this improvement comes from returning the selection criteria, with $N_{js}$ selection replacing the $M_{1b}$ window cut. Using a multivariate technique (seven variables rather than $H_T$ alone) accounts for the remaining 14%.

We have applied this method to the run I dataset, where 64 events pass the selection criteria (Table II). The overlap with the 65-event sample of the search reported in Ref. [6] is 35 events. Figure 5 shows the distribution of data events in the $O_1 + O_2 + O_3 = 1$ plane. We maximize the likelihood of Eq. (2) to extract a signal contribution of $n_s = 23.9 \pm 7.7$(stat) $\pm 4.7$(syst) events, or equivalently $13.5 \pm 5.1$ pb, including systematic uncertainties. This can be compared to the expected value of 2.2 pb. It can be seen in Fig. 5 that a significant fraction of the data events is indeed consistent with the simulated signal distribution. The numbers of back-

![Graph 2](image2.png)

**FIG. 11.** The x and y neural-network output distributions for the data events (black line) and for the Monte Carlo events mixed in the proportions returned by the fit (gray line), respectively.

Several cross checks of the results have been done. Due to the large expected non-top contribution in the data, the non-top background model is perhaps the most important factor determining the ANN fit result. As described in the previous sections, our non-top model is a PYTHIA sample of $W$+heavy flavor jets events. Using HERWIG $Wb\bar{b}$, $t\bar{t}$, and PYTHIA $Wc\bar{c}$, $Wc$ samples we derived the systematic uncertainty listed in Table III. To further test how the shape of the non-top ANN output distribution depends on the particular Monte Carlo generator, we have studied a WBBGEN [23] sample of $Wb\bar{b}$ events. This sample was run through the ANN, and the resulting distribution was used to fit the data, along with the default signal and $t\bar{t}$ distributions of Fig. 5. The fit yields a signal contribution $\sigma_s = 11.1 \pm 5.2$ pb (stat+syst), consistent with the 13.5 $\pm 5.1$ pb value obtained using the PYTHIA background estimation. Another case considered was the extreme alternative of replacing the default non-top sample with a PYTHIA sample of $W$+light flavor jets events where a jet is mistagged as a B-jet. We have found that the ANN input and output distributions are very similar for the mistags and the default non-top samples, confirming that the mistags are modeled well in our analysis. Finally, we performed a “goodness of fit” test by employing a simple $\chi^2$ fit. For this study, the

\[
\sigma(W^+ + Wg) < 23.8 \text{ pb at } 95\% \text{ C.L. (stat+syst).}
\]
(x,y) output space was divided into 10 bins with roughly equal data populations. We fit the data as a weighted sum of the signal and background templates (10-bin histograms) to obtain $\sigma_s = 15.0 \pm 5.9$ pb (stat+syst), with a $\chi^2$ of 3.2 for 6 degrees of freedom, indicating reasonable agreement between data and Monte Carlo output distributions.

In summary, we have searched for single-top production using a neural-network method. We constructed a network whose outputs estimate signal and background posterior probabilities for every given event. The method presented here improves the previous CDF search strategy reported in [6]. By analyzing the run I dataset, we found an upper limit of 24 pb (at 95% C.L.) on the single-top cross section.

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 Korea Science and Engineering Foundation (KoSEF); the Korea Research Foundation; and the Comision Intermensil- rial de Ciencia y Tecnologia, Spain.