



## Single-Top-Quark Search with $0.7 \text{ fb}^{-1}$

The CDF Collaboration  
URL <http://www-cdf.fnal.gov>  
(Dated: April 18, 2006)

We report results from a search for single-top-quark production in a  $0.7 \text{ fb}^{-1}$  dataset accumulated with the CDF II detector. We employ two different multivariate techniques to analyze the data, one using neural-networks discriminants and the other using a multivariate likelihood. Neither method finds a significant evidence of a signal, and 95% confidence level limits are set under various scenarios.

*Preliminary Results for Summer 2006 Conferences*

## I. INTRODUCTION

According to the Standard Model, in  $p\bar{p}$  collisions at the Tevatron top quarks can be created in pairs via the strong force, or singly via the electroweak interaction. The latter production mode is referred to as “single-top-quark” production and takes place mainly through the  $s$ - or  $t$ -channel exchange of a  $W$  boson (Figure 1). The CDF and DØ collaborations have published single-top results at  $\sqrt{s} = 1.8$  TeV and  $\sqrt{s} = 1.96$  TeV [1, 2]. None of these studies established single-top evidence, and 95% confidence level (C.L.) upper limits on the single-top production cross section were set.

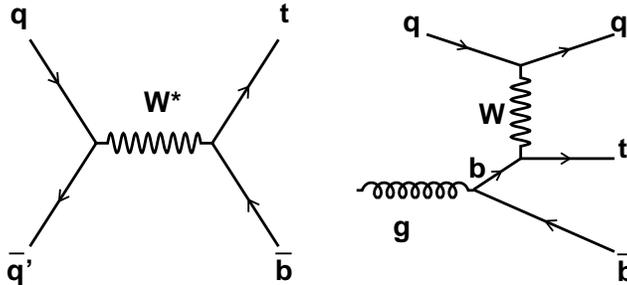


FIG. 1: Representative Feynman diagrams for single-top-quark production at the Tevatron: s-channel  $W^*$  (left) and t-channel  $W$ -gluon fusion (right).

Studying single-top production at hadron colliders is important for a number of reasons. First, it provides the only window into measuring the CKM matrix element  $|V_{tb}|^2$ , which in turn is closely tied to the number of quark generations. Second, measuring the spin polarization of single-top quarks can be used to test the V-A structure of the top weak charged current interaction. Third, single-top events represent an irreducible background to several searches for SM or non-SM signals, as for example Higgs boson searches. Fourth and last, the presence of various new SM and non-SM phenomena may be inferred by observing deviations from the predicted rate of the single-top signal.

The theoretical single-top production cross section is  $\sigma_{s+t} = 2.9$  pb for a top mass of  $175$  GeV/ $c^2$  [3]. Despite this small rate, the main obstacle in finding single-top is in fact the large associated background. After all selection requirements are imposed, the signal to background ratio is close to 1/10. This challenging, background-dominated dataset is the main motivation for using multivariate techniques. The following sections present the event selection, the signal and background estimations, an extended  $b$ -tagger and a kinematic fitter used to improve signal identification, the statistical techniques and the expected and observed single-top cross section results, and a brief summary of these results.

## II. SELECTION REQUIREMENTS

### A. Trigger

Events for this analysis are selected with a high- $E_T$  electron trigger or a high- $P_T$  muon trigger, both at 18 GeV and measured in the central subdetectors ( $|\eta| < 1$ ). A new addition to the analysis was the inclusion of a forward electron trigger which selects events with one high- $E_T$  electron (20 GeV) and  $1 < |\eta| < 2$ , and large missing transverse energy  $\cancel{E}_T$  (15 GeV). While the plug calorimeters extend beyond  $|\eta| > 3$ , the pseudorapidity range is restricted due to the additional requirement that the calorimeter energy cluster matches to a track measured in the silicon detector, which only covers the  $|\eta| < 2$  region.

### B. Event Selection

Our selection exploits the kinematic features of the signal final state, which contains a top quark, a bottom quark, and possibly additional light quark jets. To reduce multijet backgrounds, the  $W$  originating from the top quark is required to have decayed leptonically. We demand therefore a high-energy electron or muon ( $E_T(e) > 20$  GeV, or  $P_T(\mu) > 20$  GeV/ $c$ ) and large missing energy from the undetected neutrino  $\cancel{E}_T > 20$  GeV. We reject dilepton events from  $t\bar{t}$  and  $Z$  decays by requiring the dilepton mass to be outside the range:  $76$  GeV/ $c^2 < M_{\ell\ell} < 106$  GeV/ $c^2$ . The

Process	2 jets
$t$ -channel	$16.71 \pm 1.72$
$s$ -channel	$11.52 \pm 0.91$
$t\bar{t}$	$40.34 \pm 3.45$
$WW$	$3.81 \pm 0.40$
$WZ$	$6.09 \pm 0.55$
$ZZ$	$0.21 \pm 0.02$
$Z \rightarrow \tau\tau$	$2.59 \pm 0.27$
$Z \rightarrow \mu\mu$	$4.44 \pm 0.48$
Total MC-based background	$57.48 \pm 3.56$

TABLE I: Number of  $W + 2$  jets events expected from signal,  $t\bar{t}$ , diboson, and  $Z \rightarrow \tau\tau, \mu\mu$  in the  $0.7 \text{ fb}^{-1}$  dataset.

Process	2 jets
Non- $W$	$119.5 \pm 40.4$
Mistags	$164.3 \pm 29.6$
$Wb\bar{b}$	$170.7 \pm 49.2$
$Wc\bar{c}$	$64.5 \pm 17.3$
$Wc$	$69.4 \pm 15.3$
Total data-based (all above)	$588.4 \pm 96.0$
Single-top	$28.2 \pm 2.6$
Total MC-based (no single-top)	$57.5 \pm 3.6$
Total Expected	$674.1 \pm 96.1$
Observed	689

TABLE II: Expected and observed numbers of  $W + 2$  jets events after all selection requirements have been imposed.

backgrounds surviving these selections can be classified as “non-top” and  $t\bar{t}$ . The non-top backgrounds are:  $Wb\bar{b}$ ,  $Wc\bar{c}$ ,  $Wc$ , mistags (light quarks misidentified as heavy flavor jets), non- $W$  (events where a jet is erroneously identified as a lepton), and diboson  $WW$ ,  $WZ$ , and  $ZZ$ . We remove a large fraction of the non-top and  $t\bar{t}$  backgrounds by demanding exactly two jets with  $E_T > 15 \text{ GeV}$  and  $|\eta| < 2.8$  be present in the event. At least one of these two jets should be tagged as a  $b$ -quark jet by using displaced vertex information from the silicon vertex detector (SVX). The non- $W$  content of the selected dataset is further reduced by requiring the angle between the  $\cancel{E}_T$  vector and the transverse momentum vector of the leading jet to satisfy:  $0.5 < \Delta\Phi < 2.5$ .

### III. SIGNAL AND BACKGROUND ESTIMATIONS

Depending on the method in which their contributions are estimated, the different processes can be classified into two categories: Monte Carlo-based or data-based estimations. For example, the  $t\bar{t}$ , diboson ( $WW$ ,  $WZ$  and  $ZZ$ ) contributions,  $Z \rightarrow \tau\tau$  and  $Z \rightarrow \mu\mu$  belong to the first category. The same can be said about signal estimations. For all these processes, the contributions are estimated using a combination of Monte Carlo-generated samples (to extract acceptance and efficiency factors) and the theoretical cross sections (to normalize the rates).[4] Table I shows the expected yields in the  $0.7 \text{ fb}^{-1}$  dataset.

The other category contains those background processes whose estimations require the use of CDF data:  $W$ +heavy flavor ( $Wb\bar{b}$ ,  $Wc\bar{c}$ ,  $Wc$ ), mistags, and non- $W$  events. Their contributions are obtained using a similar method with that employed in Ref.[5], with two differences. One difference is the larger  $\eta$  range for the jet definition ( $|\eta| < 2.8$ ) used in this search. The other difference is the inclusion of the forward electron events, for which a different method is used in estimating the non- $W$  component. The expected and observed event yields corresponding to the  $0.7 \text{ fb}^{-1}$  dataset are given in Table II.

## IV. SPECIAL EVENT VARIABLES

### A. ANN extended $B$ -tagger

An Artificial Neural Network (ANN) [6] was developed to increase the  $b$ -quark purity of the sample selected by the standard  $b$ -tagging algorithm. The latter is based on measuring displaced (secondary) vertices, and in addition to  $b$ -jets it also selects a significant fraction of  $c$ - and light flavor jets as well (as much as 50%). The extended (ANN) tagger is applied to jets selected by the standard  $b$ -tagger, and exploits mainly the long lifetime (1.6 ps) of  $b$ -hadrons. Other features used by the ANN are the high  $b$ -quark mass, the high decay multiplicity, and the decay into leptons. For illustration, Fig. 2 shows good agreement between the ANN output distributions for the  $W + 2$  jet data and a fit to the individual background components.

As a measure of the power provided by this tool, we note that using the ANN output to select events in the single-top analysis would lead to a reduction of more than 60% of non- $b$  vertices while keeping about 82% of real  $b$  vertices. In the analysis however we will not cut on the ANN tagger output, but rather use this output as an event variable.

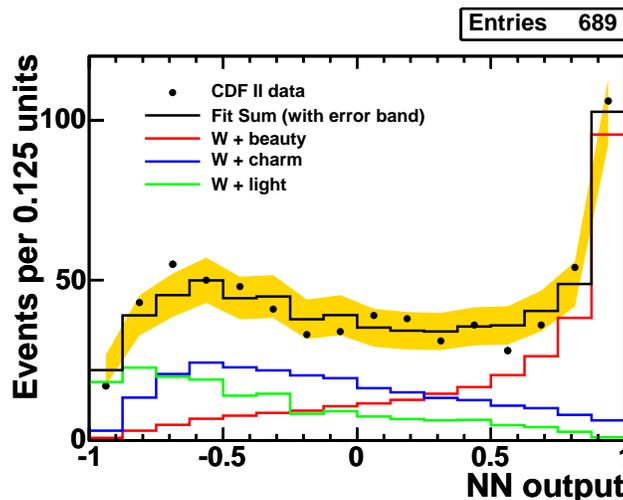


FIG. 2: The ANN tagger output distributions for the CDF  $W+2$  jets events (points) compared to the fit of the Monte Carlo distributions of the three primary components.

### B. Kinematic Fitter

Some of the variables used in our analysis require knowledge of the top rest frame. Other variables, such as the MADGRAPH matrix element are very sensitive to  $M_{\ell\nu b}$ , which is poorly measured compared to the width of the pole in the propagator in the matrix element. If we use the measured four-vectors for the two jets, the missing transverse energy and the lepton directly in the matrix element, we will get a random contribution from the mismeasured top mass which is mistakenly interpreted to be far off-shell. One of the main purposes of the kinematic fit is to find the four-vectors of the  $b$ -decay jet and the neutrino which are the most consistent with the measured values and which have the top quark on shell. A similar problem arises with the measured mass of the lepton and neutrino four-vectors. If these depart from the  $W$  mass, the matrix element can be very small due to the  $W$  propagator. The mass of the  $\ell\nu$  system is forced to  $m_W$ , but this fails in about 30% of events because of the values of the lepton momentum and missing transverse energy vectors do not allow a real solution of  $P_z$  for the neutrino. The kinematic fit  $\chi^2$  function includes a term which steers the fit to find real solutions for the neutrino momentum.

There are five quantities which are quite imprecisely measured:  $(P_x(b), P_y(b), P_z(b))$  and  $(P_x(\nu), P_y(\nu))$ . These variables can be bijectively mapped to  $(P_b, \eta_b, \Phi_b)$  and  $(\cancel{E}_T, \cancel{\Phi}_\nu)$ . Assuming that the  $b$  quark direction is reasonably well measured, then there are only three quantities we will allow to float in the kinematic fit:  $(P_b, \cancel{E}_T, \cancel{\Phi}_\nu)$ .

The following ideas form the basis for the kinematic fitter:

1. Allow  $P_b, \cancel{E}_T, \cancel{\Phi}_\nu$  to float within their uncertainties around their measured values.

2. Constrain the mass of the lepton- $\nu$  pair to  $M_W = 80.4 \text{ GeV}/c^2$ , and derive the two neutrino solutions  $P_{z_1}$  and  $P_{z_2}$  (analytical functions of  $\cancel{E}_T$  and  $\cancel{\Phi}_\nu$ ).
3. Construct a  $\chi^2$  function:

$$\chi^2 = \frac{(P_b - P_b^{obs})^2}{\sigma_{P_b}^2} + \frac{(\cancel{E}_T - \cancel{E}_T^{obs})^2}{\sigma_{\cancel{E}_T}^2} + \frac{(\cancel{\Phi}_\nu - \cancel{\Phi}_\nu^{obs})^2}{\sigma_{\cancel{\Phi}_\nu}^2} + \frac{(M_{\ell\nu b} - M_t)^2}{\sigma_{M_t}^2} + Y(\text{Im}(P_z))^2$$

4. We minimize the  $\chi^2$  above with respect to  $P_b$ ,  $\cancel{E}_T$ ,  $\cancel{\Phi}_\nu$ , under four scenarios:
  - a.  $\chi_1^2$ ,  $\chi_3^2$  in which we assume jet  $j_1$  comes from top decay and use the two neutrino solutions  $P_{z_1}$  and  $P_{z_2}$  in calculating  $M_{\ell\nu b}$ .
  - b.  $\chi_2^2$ ,  $\chi_4^2$  in which we assume jet  $j_2$  comes from top decay and use the two neutrino solutions  $P_{z_1}$  and  $P_{z_2}$  in calculating  $M_{\ell\nu b}$ .

The  $\chi^2$  values above will be used to select the  $b$ -jet from top decay, to select the  $P_z$  solution for the neutrino, or simply as event variables.

## V. LIKELIHOOD FUNCTION ANALYSIS

No single variable encodes all conceivable signal-background separation, and so a likelihood function [7] is proposed to combine several variables together into a discriminant which can be used to compute limits or to discover a signal.

The likelihood function  $\mathcal{L}$  is constructed by first forming histograms of each variable ( $n_i$  bins per variable), separately for the signal distributions and for the several background distributions, denoted  $f_{ijk}$  for bin  $j$  of variable  $i$  for the event class  $k$ . For the signal,  $k = 1$ , and in this note, four background classes are considered:  $Wb\bar{b}$ ,  $t\bar{t}$ ,  $Wc\bar{c}/Wc$ , and mistags, which are event classes 2, 3, 4 and 5. These histograms are normalized such that  $\sum_{j=1}^{n_i} f_{ijk} = 1$  for all  $i$  and all  $k$ . The likelihood function for an event is computed by evaluating in which bin  $j_i$  in which the event falls in the distribution of variable  $i$ , and computing

$$p_{ik} = \frac{f_{ij_i k}}{\sum_{m=1}^5 f_{ij_i m}}, \quad (1)$$

which is used to compute

$$\mathcal{L}_k(\{x_i\}) = \frac{\prod_{i=1}^{n_{var}} p_{ik}}{\sum_{m=1}^5 \prod_{i=1}^{n_{var}} p_{im}}. \quad (2)$$

The signal likelihood function is the one which corresponds to the signal class of events,  $\mathcal{L}_1$ .

Two likelihood functions are computed – one using the  $t$ -channel single-top signal in the signal reference histograms  $\mathcal{L}_t$ , and one using the  $s$ -channel single-top signal in the signal reference histograms  $\mathcal{L}_s$ .

### A. $t$ -channel Likelihood Function.

The  $t$ -channel likelihood function  $\mathcal{L}_t$  uses seven variables, and assumes the  $b$ -tagged jet comes from top decay:

1.  $H_T$ , the scalar sum of the  $E_T$ 's of the two jets, lepton  $E_T$ , and  $\cancel{E}_T$ .
2.  $Q \times \eta$ , the charge of the lepton times the pseudorapidity of the jet which is not  $b$ -tagged.
3.  $M_{\ell\nu b}^{hyb}$ , a hybrid reconstructed top mass formed using the raw reconstructed value of  $P_b$ , while taking the  $P_z(\nu)$  from the kinematic fit (the solution corresponding to the lower of the two  $\chi^2$  values).
4.  $\cos\theta_{t\text{-chan}}$ , the cosine of the angle between the lepton and the untagged jet in the top decay frame.
5.  $M_{jj}$ , the invariant mass of the two jets.
6.  $\text{ME}_{t\text{-chan}}$ , the MADGRAPH matrix element computed using the constrained four-vectors of the  $b$ ,  $\ell$  and  $\nu$ .
7. ANN  $b$ -tag output.

We show the  $\mathcal{L}_t$  likelihood function resulting from combining the above seven variables in Fig. 3. A good signal-background separation is apparent.

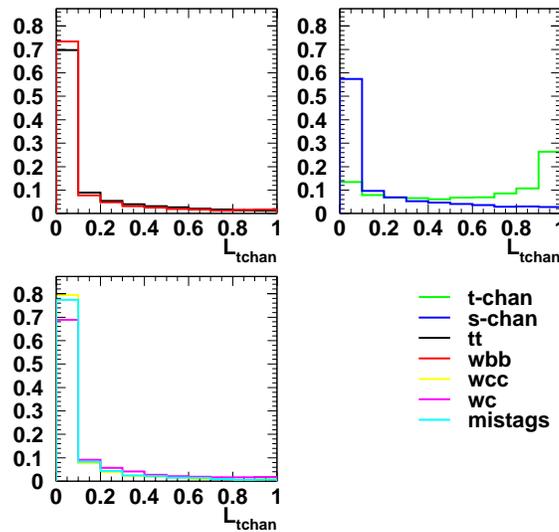


FIG. 3: The distributions of the  $t$ -channel likelihood function  $\mathcal{L}_t$  normalized to unit area.

### B. $s$ -channel Likelihood Function.

The  $s$ -channel likelihood function  $\mathcal{L}_s$  uses six variables and it is constructed in the same way as the  $t$ -channel likelihood. In fact, three of the variables are common to  $\mathcal{L}_t$  (numbers 1, 6, and 7 from the list given in the previous subsection). The other three variables are:

1.  $M_{\ell\nu b}^{hyb2}$ , a hybrid reconstructed top mass formed using the raw reconstructed value of  $P_b$ , while taking the  $P_z(\nu)$  from the kinematic fit, and choosing the  $b$ -jet from top using a simple ANN function based on the difference of kinematic fitter  $\chi^2$  and the difference between the pseudorapidities of the jets.
2.  $\cos\theta_{s\text{-chan}}$ , the cosine of the angle between the neutrino and the beam in the top quark rest frame. This variable has no special motivation, other than the fact we noticed it increases the separation.
3.  $E_T(j1)$ , the transverse energy of the leading jet.

The  $\mathcal{L}_s$  likelihood function distributions are shown in Fig. 4. The  $\mathcal{L}_s$  distributions for signal reflect the fact that none of the above six variables separates  $s$ - from  $t$ - channel components well.

### C. Statistical Method. Results.

In the previous two subsections we showed how multiple variables are combined to form the two likelihood functions  $\mathcal{L}_t$  and  $\mathcal{L}_s$ . Now we will describe how these two variables can be used to make statements about the single-top content of the CDF data. Our statistical approach follows the one presented in Ref. [8]. The CDF data is compared against two models at a time. One is the null hypothesis ( $H0$ ) which asserts that the Standard Model processes *without* single-top describes the data, while the other, referred to as the test hypothesis ( $H1$ ), asserts that the data are modeled by the SM processes *including* Standard Model single top. Our test statistic is:

$$Q(\vec{d}) = -2 \cdot \ln \frac{P(\vec{d}|H1)}{P(\vec{d}|H0)}$$

where  $N_{bin}$  denotes the total number of bins in the likelihood function,  $\vec{d} = (d_1, d_2, \dots, d_{N_{bin}})$  is the observed data distribution.

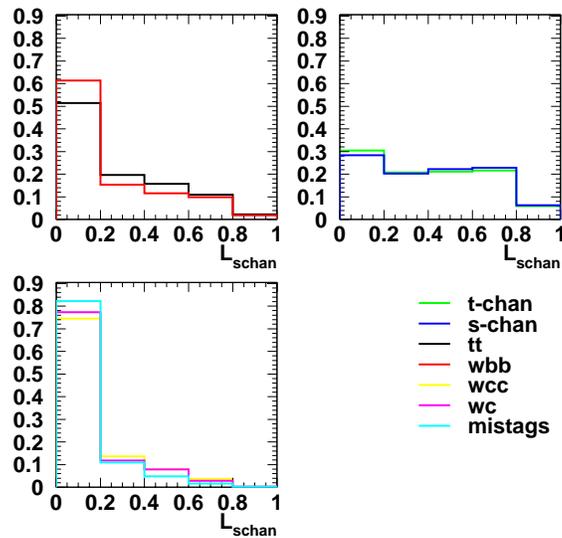


FIG. 4: The distributions of the  $s$ -channel likelihood function  $\mathcal{L}_s$  normalized to unit area. There is almost no separation between the  $s$ - and  $t$ - channel single-top distributions (upper right plot).

A large number of simulated experiments  $\vec{d}$  are drawn either from the  $H1$  or the  $H0$  hypotheses, and the corresponding  $Q(\vec{d})$  values are stored in two separate histograms  $f^{H1}(Q)$  and  $f^{H0}(Q)$ , respectively. The systematic uncertainties are accounted for as described in Ref. [8]. Both the rate and shape uncertainties are taken into account by choosing nuisance parameters randomly on each pseudoexperiment.

If for a given pseudoexperiment (or CDF data) we measure a value  $Q_{obs}$ , then the  $f^{H1}(Q)$  and  $f^{H0}(Q)$  distributions can be integrated in the region  $Q > Q_{obs}$ , and the following quantities can be defined:

$$CL_{s+b}(Q_{obs}) = \text{Prob}(Q \geq Q_{obs}|H1) = \int_{Q_{obs}}^{\infty} f^{H1}(Q) \cdot dQ \quad (3)$$

$$CL_b(Q_{obs}) = \text{Prob}(Q \geq Q_{obs}|H0) = \int_{Q_{obs}}^{\infty} f^{H0}(Q) \cdot dQ \quad (4)$$

$$CL_s(Q_{obs}) = \frac{CL_{s+b}}{CL_b} \quad (5)$$

$CL_s$  will be used to define the 95% C.L. exclusion limits, and  $1-CL_b$  is in statistics language a  $p$ -value, i.e. the probability that the  $H0$  model fluctuated up to more than the pseudo-experiment data.

There are two ways one can perform the single-top measurement: i) use the two-dimensional distributions of  $\mathcal{L}_t$  vs.  $\mathcal{L}_s$  to measure the  $s+t$  signal combined rate, or ii) use the individual  $\mathcal{L}_t$ , and  $\mathcal{L}_s$  distributions to set separate limits on  $s$ - or  $t$ - signal channel rates.

*Combined and Separate Search Results.* In Fig. 5 we present the  $\mathcal{L}_t$  and  $\mathcal{L}_s$  distributions of the data events compared to the Monte Carlo prediction (contributions from Table II). The combined search results are shown in the second column of Table III. These results assume no signal (backgrounds only) in the null hypothesis  $H0$ . The test hypothesis  $H1$  assumes backgrounds plus SM  $s+t$  channel signal. The interpretation of the  $s+t$  combined search results in Table III is as follows:

1. The median  $CL_s$  in  $H0$  pseudo-experiments is 9.2% indicating that a priori we do not have the sensitivity required (5%) to exclude the  $H1$  hypothesis at 95% C.L. The observed  $CL_s$  of 19.4% is even farther away.
2. The median  $1-CL_b$  in  $H1$  pseudo-experiments is 3.9%, indicating that a priori 50% of the  $H1$  pseudoexperiments give this  $p$ -value or less (approx.  $2\sigma$  excess). The a posteriori result  $1-CL_b = 25.6\%$  shows good consistency between the data and the  $H0$  hypothesis.

3. The expected 95% C.L. limit  $\sigma_{95} = 3.4$  pb is obtained from testing different  $H1$  hypotheses by modifying the signal rate from its SM value, until the expected  $CL_s$  becomes 5%. The observed 95% C.L. limit of  $\sigma_{95} = 4.3$  pb is slightly higher than expected, mostly due to the slight excess seen in the signal region of the  $\mathcal{L}_s$  distribution.

A similar interpretation applies to the separate search results (second and third columns in Table III). However, there is one difference. For the  $t$ -channel separate search, the null hypothesis is SM background plus SM  $s$ -channel single top *and* no  $t$ -channel single top. Similar for  $s$ -channel, where the null hypothesis consists of backgrounds plus SM  $t$ -channel signal and no  $s$ -channel signal.

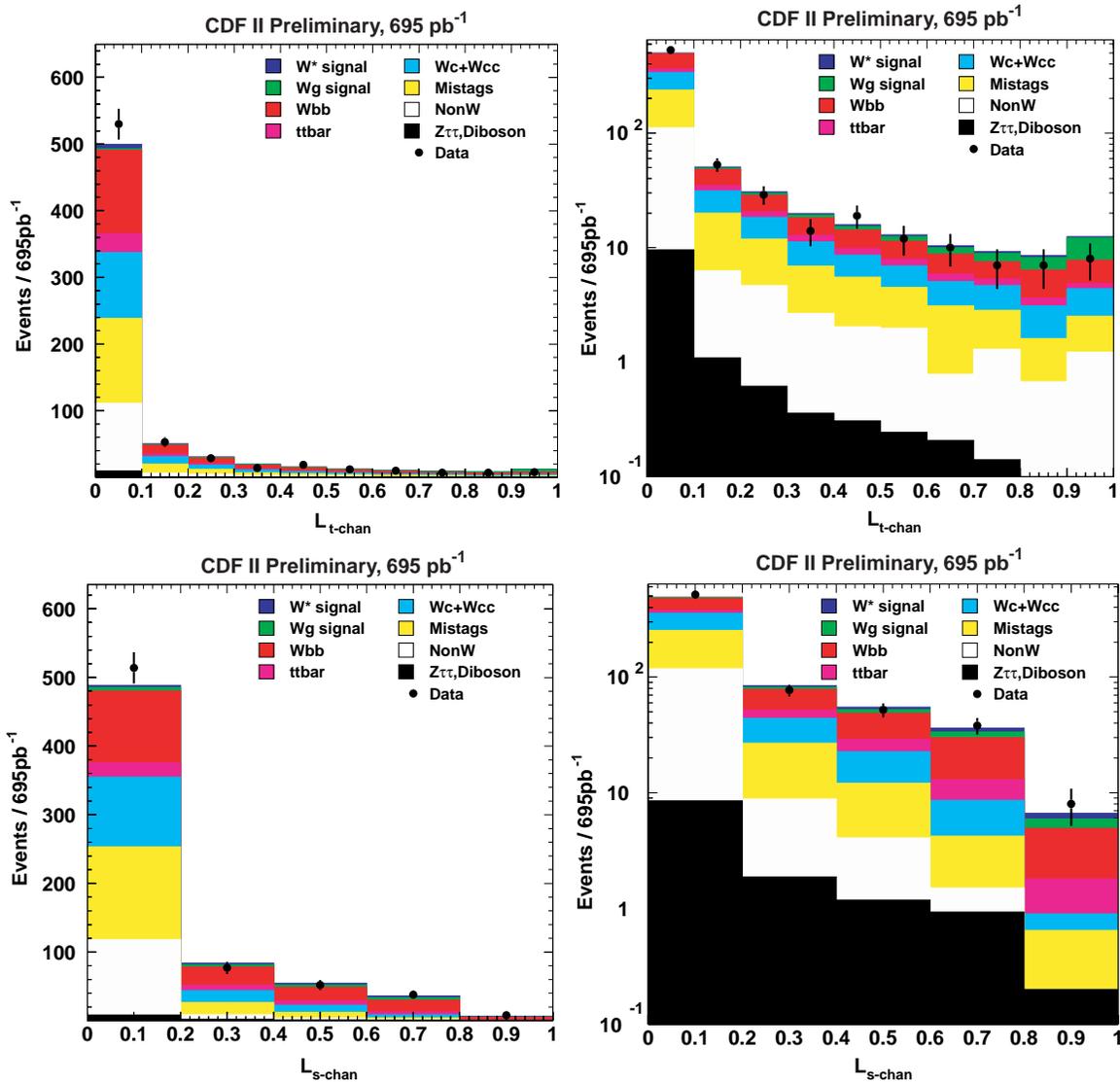


FIG. 5: The distributions of the  $t$ -channel (upper) and  $s$ -channel (lower) likelihood functions for CDF data compared to the Monte Carlo distributions normalized to the expected contributions. A linear (logarithmic) scale is used for the left (right) plots.

## VI. NEURAL NETWORK ANALYSIS

The information from multiple event variables can also be combined using the Artificial Neural Network method, which has the advantage of taking into account the correlations among the variables.

The neural networks used here are three layer perceptrons, implemented using the NeuroBayes<sup>®</sup> package [9] provided by the company Phi-T. The output layer contains one node which gives a continuous output in the interval  $[-1,1]$ .

	$s + t$ combined	$t$ -channel	$s$ -channel
Expected $CL_s$	0.092	0.127	0.733
Observed $CL_s$	0.194	0.167	0.796
Expected $1 - CL_b$	0.039	0.053	0.372
Observed $1 - CL_b$	0.256	0.394	0.375
Expected limit $\sigma_{95}$ (pb)	3.40	2.63	5.66
Observed limit $\sigma_{95}$ (pb)	4.32	2.89	5.10

TABLE III: Observed and expected results for the combined  $s + t$  search and the  $s$ - and  $t$ -channel separate searches. For comparison, the theoretical values are  $\sigma_t = 1.98$  pb, and  $\sigma_s = 0.89$  pb, respectively.

The nodes of two consecutive layers are connected with variable weights ( $\omega_{ji}$ ). For each node  $j$ , a biased weighted sum of the values of the previous layer  $x_i$  is calculated

$$a_j(\mathbf{x}) = \sum_i \omega_{ji} x_i + \mu_{0,j} \quad (6)$$

and passed to the transfer function which gives the output of the node. This transfer function is a transformed sigmoid function:

$$S(\mathbf{x}) = \frac{2}{1 + e^{-a(\mathbf{x})}} - 1 \quad (7)$$

This function has a binary, switch-like character, and the weights  $\omega$  and biases  $\mu$  are iteratively adjusted (“trained”) such that the ANN output is close to 1 for signal events, and close to -1 for background events. A more detailed description of the neural network package and its mathematical background can be found elsewhere [9].

#### A. Input variables.

As in the likelihood function analysis case, we will want to make statistical statements about the cross section of the individual  $s$ - and  $t$ - signal channels, as well as on the combined rate of  $s + t$  single-top production. There are therefore three distinct neural networks used in this analysis, one for each case mentioned. Each training sample contains about  $17k$  events. In the training of the  $s$ -channel network, the  $t$ -channel events are treated as background and vice versa, although this has a negligible effect. All networks have 14 input variables:

1.  $M_{jj}$ , the dijet invariant mass
2.  $M_{\ell\nu b}$ , the invariant mass of the lepton-neutrino- $b$  system
3.  $\log_{10} \Delta_{34}$ , calculated by the  $K_T$  jet cluster algorithm
4.  $Q \times \eta$ , the charge of the lepton times the pseudorapidity of the jet which is not  $b$ -tagged.
5.  $P_T(\ell)$ , lepton transverse momentum
6.  $\sum \eta_j$ , the sum of the pseudoerapidities of all jets in the event
7.  $\eta_W$ , the reconstructed  $W$  boson ( $\ell + \nu$ ) pseudorapidity
8.  $E_T(j1)$ , the leading jet  $E_T$
9.  $E_T(j2)$ , the second leading jet  $E_T$
10. ANN  $b$ -tag output for the highest- $P_T$   $b$ -tagged jet
11.  $\chi_1^2 - \chi_2^2$ , the  $\chi^2$  difference from the kinematic fitter
12.  $\chi_3^2$  from the kinematic fitter
13.  $\sum \eta(j^t j^l)$ , the pseudorapidity sum using the tight and loose jets
14.  $\cos \theta_{t\text{-chan}}$ , the cosine of the angle between the lepton and the untagged jet in the top decay frame.

## B. $s + t$ Combined Search Results

Based on the NN output shape similarity of the different MC distributions, we group the different physics processes into four classes: single-top ( $t$ - and  $s$ -channel),  $t\bar{t}$ , charm-like ( $Wc\bar{c}$ ,  $Wc$ ,  $WW$  and mistags) and bottom-like ( $Wb\bar{b}$ ,  $WZ$ ,  $ZZ$  and  $Z \rightarrow \tau\tau$ ). The resulting four different templates can be seen in Fig. 6 (left). The templates are normalized to unit area for comparison. The two single-top sample outputs and the resulting template for single-Top is shown in Fig. 6 (right). They are also normalized to unit area.

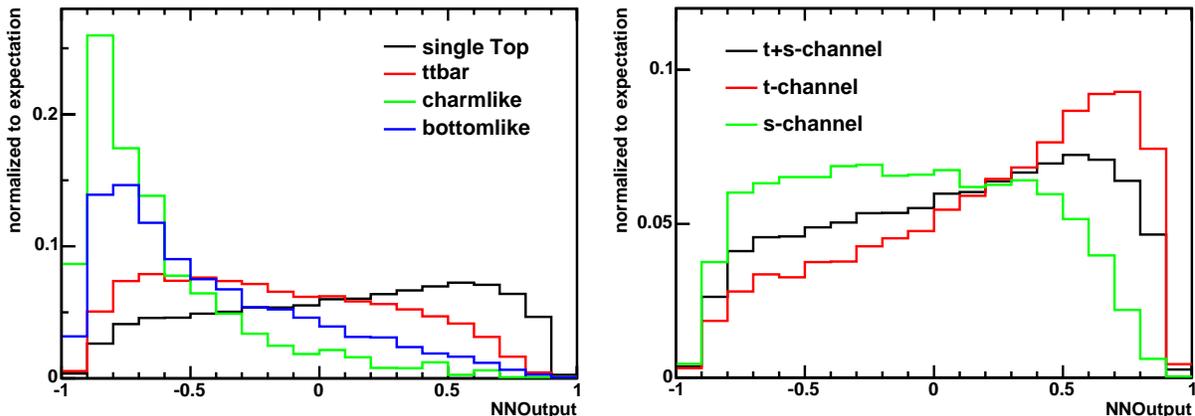


FIG. 6:  $s+t$  NN output distributions: single-top,  $t\bar{t}$ , charm-like, and bottom-like events (left); the individual signal components, along with the combined template obtained using the SM expected proportions (right).

The statistical method is fully Bayesian and it is described in the 2005 single-top PRD article [1]. The Bayesian likelihood function consists of Poisson terms for the individual bins of the fitted histograms. Systematic uncertainties are included as factors modifying the expectation value of events in a certain bin. The shape uncertainties are estimated from the shifted NN output histograms corresponding to MC samples in which the particular sources for systematic uncertainties are varied. This is simultaneously included along with the rate uncertainties which are listed in table IV.

Syst. Source	$t$ -channel	$s$ -channel	$t$ - and $s$ -channel combined
ISR	$\pm 2\%$	$\pm 1\%$	$\pm 2.7\%$
FSR	$\pm 3\%$	$\pm 1\%$	$\pm 2.1\%$
Jet Energy	+7% / -5%	$\pm 4\%$	+5% / -1%
PDF	$\pm 3\%$	$\pm 1.5\%$	$\pm 1.8\%$
NN b tagger	$\pm 1\%$	$\pm 1\%$	$\pm 6\%$
non- $W$ modeling	$\pm 2\%$	+170% / -90%	+26% / -31%
Total	+8.7% / -7.2%	+170.1% / -90.1%	+27.4% / -31.8%

TABLE IV: Estimate of the systematic uncertainty on the central value of the likelihood fits to the neural network outputs. The  $t$ - and  $s$ -channel values refer to the separate search, the number of  $t$ - and  $s$ -channel combined to the combined search.

We generate pseudo-experiments from the SM background+SM signal samples, and use the Bayesian procedure mentioned above to determine the most probable value of the signal cross section along with the 95% C.L. limits. Our expected limit corresponds to the median of the individual 95% C.L. limits distribution, which is 5.7 pb (Fig. 7, left plot).

In case of the CDF data, the likelihood fit yields a best value for the  $s+t$  cross section of  $0.8_{-0.8}^{+1.3}$  (stat.)  $_{-0.3}^{+0.2}$  (syst.) pb. The resulting upper limit on the  $s+t$  cross section is 3.4 pb at 95% C.L. The posterior probability density is shown in the right plot of Fig. 7. The fit result is illustrated in Fig. 8 (left hand-side). The data and the expectation in the signal region are shown on the right hand-side of Fig. 8.

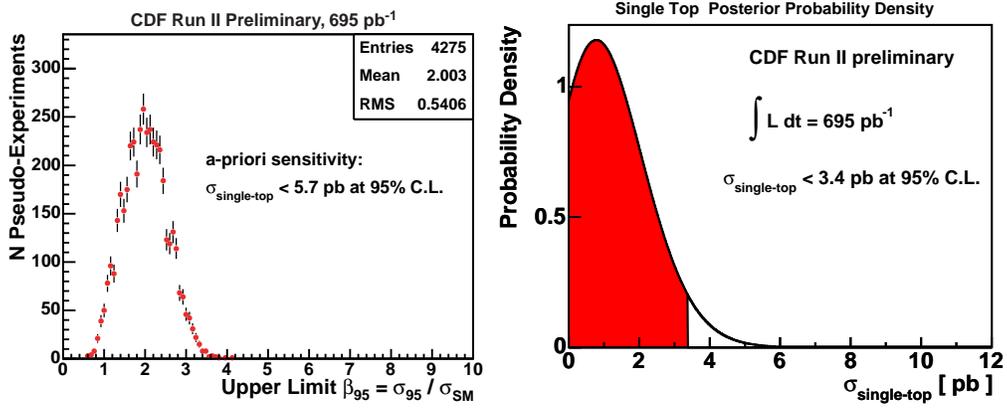


FIG. 7: Expected distribution of 95% C.L. upper limits on  $\sigma_{s+t}$  (left) and observed posterior probability density (right) for the combined search using the  $s + t$  NN output. In the left plot, the cross section is shown in units of  $\sigma_{s+t}^{SM} = 2.9 \text{ pb}$ .

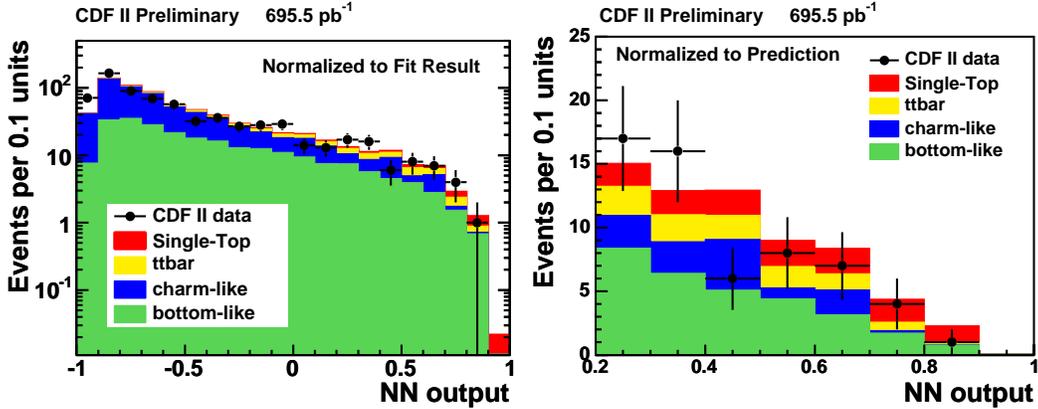


FIG. 8: Data distribution of the  $s + t$  NN output in the entire output region (left hand-side) and the signal region (right hand-side). On the left hand-side the fitted component are shown. On the right hand-side the expectation is displayed.

### C. $s-$ and $t-$ Channel Separate Search

As already mentioned, we use two distinct neural networks, one trained for  $s$ -channel and the other one for  $t$ -channel, which provide the opportunity to search for both channels individually and simultaneously. The training and input variables of the networks are similar to the one described in the previous section except that now each net is optimized for one ( $s-$  or  $t-$ ) channel alone. The 2D templates can be seen in Fig. 9. Repeating the same procedure as done in the previous subsection, we perform pseudoexperiments using the SM signal contributions, and fitted to the 2D templates shown in Fig. 9. The expected 95% C.L. limits on the  $t-$  and  $s-$  cross sections are 4.2 pb and 3.7 pb, respectively.

Using the CDF data, our likelihood fitting procedure returns  $\sigma_{t-ch} = 0.6^{+1.9}_{-0.6}(\text{stat.})^{+0.1}_{-0.1}(\text{syst.}) \text{ pb}$  and  $\sigma_{s-ch} = 0.3^{+2.2}_{-0.3}(\text{stat.})^{+0.5}_{-0.3}(\text{syst.}) \text{ pb}$ . This is graphically shown in Fig. 10. At the 95% C.L. the resulting upper limits on the  $t$ - and  $s$ -channel cross sections are 3.1 pb and 3.2 pb, respectively.

## VII. CONCLUSION

We analyzed the  $0.7 \text{ fb}^{-1}$  dataset in search of single-top-quark signal, using multivariate techniques. We present the multivariate likelihood technique and the neural networks technique, along with the expected and observed results in each case. Neither analysis finds any evidence for a signal. The 95% observed limit on the  $s + t$  combined single-top cross section using the 2D ( $\mathcal{L}_s, \mathcal{L}_t$ ) likelihood functions is 4.3 pb. Using the  $\mathcal{L}_s, \mathcal{L}_t$  functions to separately search

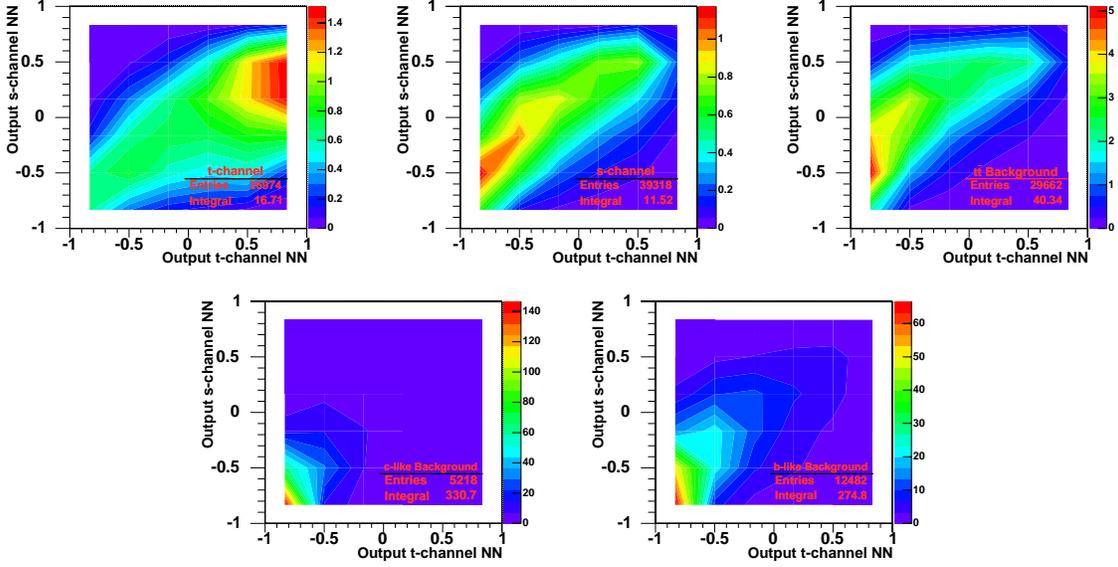


FIG. 9: Upper row: 2D NN output templates for  $t$ -channel,  $s$ -channel, and  $t\bar{t}$  events. The lower plots show the 2D templates for the charm-like and bottom-like events, respectively.

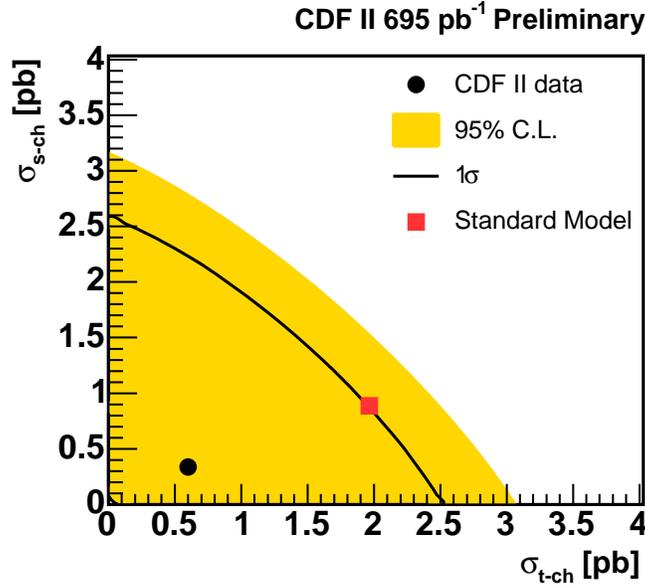


FIG. 10: The separate search results, showing the most probable value (circle) the SM prediction (square) and the region excluded at 95% C.L. by our data (yellow).

for the two corresponding signal channels, we obtain limits of 2.9 pb and 5.1 pb for the  $t$ -channel and  $s$ -channel, respectively. In the case of the neural-networks analysis, the observed 95% C.L. limit on the  $s + t$  cross section is 3.4 pb. For the separate search (using neural networks) the observed 95% C.L. limits are 3.1 pb and 3.2 pb on the  $t$ -channel and  $s$ -channel single-top, respectively. These results are summarized in Table V.

	$s + t$ combined	$t$ -channel	$s$ -channel
Multivariate Likelihood analysis: $\sigma_{95}$ (pb)	4.3	2.9	5.1
Neural-Networks analysis: $\sigma_{95}$ (pb)	3.4	3.1	3.2

TABLE V: Summary of the 95% C.L. upper limits on single-top production. For comparison, the theoretical values are  $\sigma_t = 1.98$  pb, and  $\sigma_s = 0.89$  pb, respectively.

### Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

- 
- [1] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **65**, 091102 (2002); D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **69**, 052003 (2004); D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **71**, 012005 (2005).
- [2] B. Abbott *et al.* (DØ Collaboration), Phys. Rev. D **63**, 031101 (2001); V. M. Abazov *et al.* (DØ Collaboration), Phys. Lett. B **517**, 282 (2001); V. M. Abazov *et al.* (DØ Collaboration), Phys. Lett. B **622**, 265 (2005).
- [3] B. W. Harris *et al.*, Phys. Rev. D **66**, 054024 (2002); Z. Sullivan, Phys. Rev. D **70**, 114012 (2004).
- [4] S. Budd *et al.*, “Event Detection Efficiency for Single-Top Events in CDF II”, CDF Note 7863, 2006 (*internal*).
- [5] The CDF Collaboration, “Measurement of the  $t\bar{t}$  Production Cross Section in Vertex-Tagged Lepton+Jets Events” CDF Note 8810, 2006 (*public*).
- [6] M. Feindt, S. Richter, and W. Wagner, “A Neural Network  $b$  Tagger for Single-Top Analyses”, CDF note 7816, 2006 (*internal*).
- [7] R. Barlow, “Statistics: a Guide to the Use of Statistics in the Physical Sciences” John Wiley and Sons, LTD, West Sussex, England (1989).
- [8] T. Junk, “Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematics”, CDF note 8128 (*internal*). Also see: A. L. Read, J. Phys. G: Nucl. Part. Phys. **28**, 2693 (2002); P. Bock *et al.* (the LEP Collaborations), CERN-EP-98-046 (1998) and CERN-EP-2000-055 (2000).
- [9] M. Feindt, e-Print Archive physics/0402093 (2004).