FERMILAB-CONF-08-681-E

MEASUREMENT OF THE ELECTROWEAK SINGLE TOP PRODUCTION CROSS SECTION AT THE TEVATRON

Peter Dong for the CDF and D0 Collaborations
University of California, Los Angeles, CA 90095, USA

Abstract

The CDF and D0 collaborations have analyzed 2.2 fb^{-1} and 0.9 fb^{-1}, respectively, of Run II data to search for electroweak single top quark production at the Tevatron. We employ several different analysis techniques to search for a single top signal: boosted decision trees, multivariate likelihood functions, neural networks, and matrix element discriminants. Both experiments see evidence of single top production. D0 measures a combined cross section of 4.7 ± 1.3 pb while CDF measures $2.2^{+0.8}_{-0.7}$ pb. D0 sets a limit at a 95% confidence level of $|V_{tb}| > 0.68$ and measures $|V_{tb}| = 1.3 \pm 0.2$, while CDF calculates $|V_{tb}| > 0.6$ and measures $|V_{tb}| = 0.88^{+0.16}_{-0.14}(\text{stat} + \text{sys}) \pm 0.07(\text{theory})$.

1 Introduction

In 1.96 TeV proton anti-proton collisions at the Tevatron, top quarks are predominantly produced in pairs via the strong force. In addition, the Standard Model predicts that single top quarks can be produced through an electroweak s- and t-channel exchange of a virtual W boson (Figure 1). The production cross sections have been calculated at next-to-leading order (NLO). For a top quark mass of 175 GeV/c^2, the results are 1.98 ± 0.25 pb and 0.88 ± 0.11 pb for the t-channel and s-channel processes, respectively [1]. The combined cross section is about 40% of the top pair production cross section.

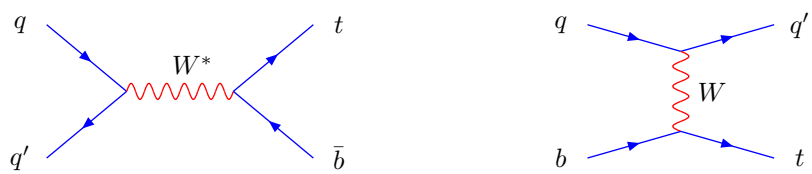

Figure 1: Leading order Feynman diagrams for s-channel (left) and t-channel (right) single top quark production.

The precise measurement of the production cross section allows the direct extraction of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{tb}|$ and offers a source of almost completely polarized top quarks [2]. Moreover, the search for single top also probes exotic models beyond the Standard Model. New physics, like flavor-changing neutral currents or heavy W' bosons, could alter the observed production rate [3]. Finally, single-top processes result in the same final state as the Standard Model Higgs boson process $WH \to Wb\bar{b}$, which is one of the most promising low-mass Higgs search channels at the Tevatron [4]. Nearly all analysis tools developed for the single top search can be used for this Higgs search.

A measurement of this cross section was performed at D0 with 0.9 fb^{-1} of data and at CDF with 2.2 fb^{-1} of data. Each experiment has three analysis techniques that share a common event selection, background estimate, and statistical treatment.

2 Event Selection

The single top event selection exploits the kinematic features of the signal final state, which contains a real W boson, one or two bottom quarks, and possibly additional jets. To reduce multi-jet backgrounds, the W originating from the top quark decay is required to have decayed leptonically. This leads to

a requirement of a high-energy electron or muon and large missing transverse energy ($\not\!\!E_T$) from the undetected neutrino. CDF requires the electron and muon to have $p_T > 20$ GeV and $|\eta| < 1.6$; D0 requires $p_T > 15$ GeV and $|\eta| < 1.1$ for electrons and $p_T > 18$ GeV and $|\eta| < 2.0$ for muons. CDF requires $\not\!\!E_T < 25$ GeV, while D0 requires $15 < \not\!\!E_T < 200$ GeV.

CDF requires two or three jets with $E_T > 20$ GeV and $|\eta| < 2.8$, while D0 requires from two to four jets with $|\eta| < 3.4$. The first jet must have $E_T > 25$ GeV, the second jet must have $E_T > 20$ GeV, and any additional jets must have $E_T > 15$ GeV.

A large fraction of the background is removed by demanding that at least one of these two jets be tagged as a b-quark jet using displaced vertex information from the silicon detector. CDF's secondary vertex tagging algorithm identifies tracks associated with the jet originating from a vertex displaced from the primary vertex indicative of decay particles from relatively long-lived B mesons. D0 uses a neural network which includes seven input variables to distinguish tracks resulting from b quarks, increasing their acceptance for the same rate of mistags compared to a simple secondary-vertex tagger.

The backgrounds surviving these selections are $t\bar{t}$, W + heavy-flavor jets, i.e. $W + b\bar{b}$, $W + c\bar{c}$, $W + c$, Z + heavy-flavor jets, and the diboson processes WW, WZ, and ZZ. Instrumental backgrounds originate from mis-tagged W + jets events (W events with light-flavor jets, i.e. with u, d, or s-quark and gluon content, misidentified as heavy-flavor jets) and from non-W + jets events (multi-jet events in which one jet is erroneously identified as a lepton).

2.1 Background Estimate

Estimating the background contribution after applying the event selection to the single top candidate sample is an elaborate process. NLO cross section calculations exist for diboson and $t\bar{t}$ production, thereby making the estimation of their contribution relatively straightforward. The main background contributions are from $W + b\bar{b}$, $W + c\bar{c}$ and $W + c$ + jets, as well as mis-tagged $W+$ light quark jets. We determine the $W+$ jets normalization from the data and estimate the fraction of the candidate events with heavy-flavor jets using ALPGEN Monte Carlo samples [5], which were calibrated with multi-jet data [6]. The probability that a $W+$ light-flavor jet is mis-tagged is parameterized using a large, generic multi-jet data set. The instrumental background contribution from non-W events is estimated using data in a control region with low $\not\!\!E_T$, containing very little signal, and we subsequently extrapolate the contribution into the signal region with large $\not\!\!E_T$. The expected signal and background yield in the signal sample is shown in Figures 2 and 3 for D0 and CDF, respectively.

The background estimate demonstrates that the expected number of sin-

Process	2 jets	3 jets	4 jets
s-channel	16 ± 3	8 ± 2	2 ± 1
t-channel	20 ± 4	12 ± 3	4 ± 1
$W + bb$	261 ± 55	120 ± 24	35 ± 7
$W + c\bar{c}$	151 ± 31	85 ± 17	23 ± 5
$W+$ light quarks	119 ± 25	43 ± 9	12 ± 2
non-W	95 ± 19	77 ± 15	29 ± 6
$t\bar{t}$	59 ± 14	135 ± 32	154 ± 36
Total prediction	686 ± 131	460 ± 75	29 ± 6
Observed in data	697	455	246

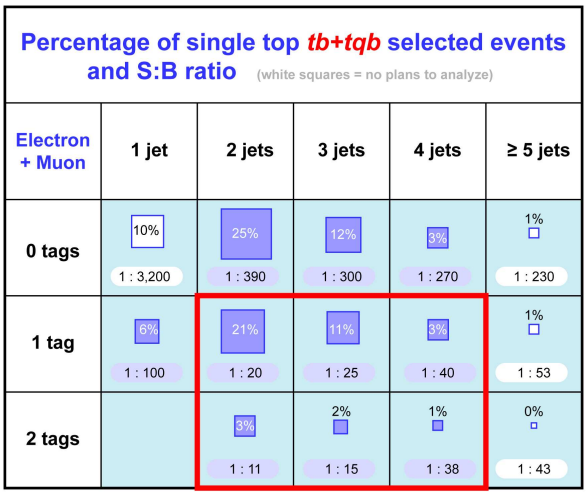

Figure 2: Top: Expected signal and background yield for D0's signal samples. Bottom: Graphical depiction of the expected amount of single-top signal as a function of the W + jets multiplicity and the presence of a b-tagged jet.

Process	2 jets	3 jets
s-channel	41.2 ± 5.9	13.5 ± 1.9
t-channel	62.1 ± 9.1	18.3 ± 2.7
$W + b$	461.6 ± 139.1	141.1 ± 42.6
$W + c$	395.0 ± 121.8	108.8 ± 33.5
$W+$ light quarks	339.8 ± 56.1	101.8 ± 16.9
non-W	59.5 ± 23.8	21.3 ± 8.5
Diboson	63.2 ± 6.3	21.5 ± 2.2
Z + jets	26.7 ± 3.9	11.0 ± 1.6
$t\bar{t}$	146.0 ± 20.9	338.7 ± 48.2
Total prediction	1595.1 ± 269.0	776.6 ± 91.4
Observed in data	1535	712

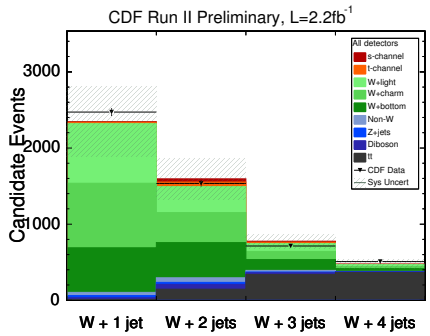

Figure 3: Top: Expected signal and background yield in CDF's signal samples. Bottom: Graphical depiction of sample composition as a function of jet multiplicity.

gle top events is much less than the large amount of expected backgrounds. In fact, the systematic uncertainty on the background estimate is larger than the expected signal, which renders a simple counting experiment impossible. The search for single top quark production requires the best possible discrimination between signal and background processes, thus motivating the use of multivariate analysis tools.

2.2 Neural Network Jet-Flavor Separation

Mistags and $W + c$ events are a large class of background where no real b-quark is present and amount to about 50% of the $W + 2$ jets data sample even after imposing the requirement that one jet is identified by a secondary vertex b-tagger. CDF uses a neural network tool which uses secondary vertex tracking information to distinguish jets from b quarks from jets from c and light quarks. Figure 4 shows the distribution of this jet-flavor-separating neural network in Monte Carlo events. All CDF single-top analyses use this tool to improve their sensitivity.

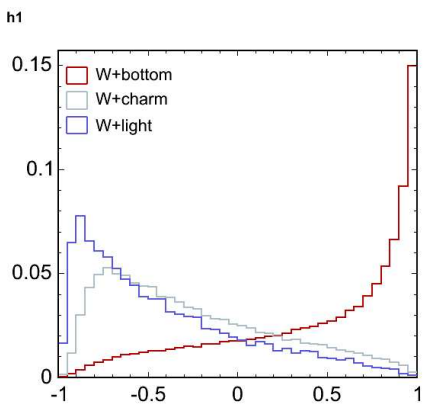

Figure 4: Distribution of the jet-flavor separator used by CDF in $W+$ jets events. Jets likely to come from b quarks are given a high probability, while light-quark jets are assigned a low probability.

3 Analysis Techniques

No single kinematic distribution encodes all conceivable information that can separate signal from background. Sophisticated analysis techniques are needed

to combine information from different variables into a single discriminant distribution which is used to extract the single top rate from the data.

3.1 Boosted Decision Trees (D0)

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 [7]. This average is called a boosted decision tree. One characteristic of decision trees is that because they optimize a cut at each level, their sensitivity is not reduced by the inclusion of unnecessary variables.

D0's boosted decision tree is trained on 49 input variables, the most important of which are the invariant mass of all 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 ($\cos\theta^*_{\ell j}$) [8], and the lepton charge times the pseudorapidity of the untagged jet ($Q \times \eta$) [9].

3.2 Multivariate Likelihood Function (CDF)

A projective likelihood technique [10] is used to combine information from seven input variables to optimize the separation of the single top signal from the backgrounds in the two-jet case. The input variables x_i, $i = 1...7$ are measured for each event and the quantities

$$\mathcal{L}_{2n}(\{x_i\}) = \frac{\prod_{i=1}^{7} p_{i,k=1}}{\sum_{k=1}^{5}\prod_{i=1}^{7} p_{ik}}, \text{ where } p_{ik} = \frac{f_{in_i,k}}{\sum_{m=1}^{5} f_{in_i m}} \tag{1}$$

are computed, where m and k index the five samples (t-channel signal, $Wb\bar{b}$, $Wc\bar{c}+Wc$, Wjj and $t\bar{t}$) and $f_{in_i m}$ is the normalized probability distribution in bin n_i for variable i of a template histogram constructed from a Monte Carlo (or data) model of sample m. Seven to ten input variables are chosen, including the jet-flavor separator neural network output, the mass of the reconstructed top quark, $Q \times \eta$, total scalar sum of transverse energy in the event H_{T}, the invariant mass of all jets in the event, and $\cos\theta^*_{\ell j}$.

3.3 Matrix Element Discriminant (CDF, D0)

The matrix element method relies on the evaluation of event probability densities for signal and background processes based on calculations of the Standard Model fully differential cross sections [12]. These probability densities are calculated for signal and background hypotheses for each event and quantify

how likely the event is to originate either from signal or background. Given a set of observables, x, and underlying partonic quantities, y, the signal and background probability densities are constructed by integrating over the appropriate parton-level differential cross section, $d\sigma(y)/dy$, convolved with parton distribution functions (PDFs) and detector resolution effects:

$$P(x) = \sum_{perm.} \int \frac{d\sigma(y)}{dy} f(q_1) f(q_2) dq_1 dq_2 W(x,y) dy. \tag{2}$$

The PDFs ($f(q_1)$ and $f(q_2)$) take into account the flavors of colliding quark and anti-quark and are given by the CTEQ collaboration [13]. The detector resolution effects are described by a transfer function $W(x,y)$ relating x to y. The momenta of electrons, muons and the angles of jets are assumed to be measured exactly. $W(x,y)$ maps parton energies to measured jet energies after correction for instrumental detector effects. This mapping is obtained by parameterizing the jet response in fully simulated Monte Carlo events, which includes effects of radiation, hadronization, measurement resolution, and energy omitted from the jet cone by the jet-reconstruction algorithm. The definition of the probability densities includes possible permutations of matching jets with partons. The integration is performed over the energy of the partons and p_z^ν. Event probability densities are computed for the s-channel and t-channel signal, as well as $Wb\bar{b}$, $Wc\bar{c}$, Wcj, Wgg, and $t\bar{t}$ background hypotheses. In the specific case of the $t\bar{t}$ matrix element, additional integrations must be performed over the momenta of particles not detected.

The event probability densities are combined into an event probability discriminant, i.e. a distribution which separates signal from background which is used to fit the data: $P_{signal}/(P_{signal} + P_{background})$. To better classify signal events which contain b-jets, CDF incorporates the output of the neural network jet-flavor separator, b, into the discriminant defined as:

$$\frac{b \cdot P_{st}}{b \cdot P_{st} + b(P_{t\bar{t}} + P_{Wbb}) + (1-b)(P_{Wcc} + P_{Wcj})} \tag{3}$$

3.4 Neural Network (CDF)

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

Separate networks are trained to identify different signals in distinct samples. The networks use 11 to 18 input variables, the most important ones being the reconstructed top quark mass, the neural-network jet-flavor separator, the dijet mass, $Q \times \eta$, the cosine of the angle between the lepton and the light quark jet, the transverse mass of the W boson, and H_{T}. The input variables are selected from a large list of investigated variables using an automated evaluation during the preprocessing step before the network training. In an iterative process, we determine how much the total correlation of the ensemble of variables to the target is reduced by the removal of each single variable. For the networks used in this analysis we kept those variables for which the correlation reduction was significant.

3.5 Bayesian Neural Network (D0)

CDF uses a traditional neural network. D0 uses a Bayesian neural network, which is a weighted sum over one hundred different neural networks sampled from the posterior probability density function of the space of network parameters. This protects the network from overtraining and gives the best average neural network for the analysis. The input variables to the Bayesian neural network are similar to those used in the boosted decision tree analysis.

4 Measurement Technique and Results

The cross section is extracted by a Bayesian method in which a posterior probability density function is constructed by integrating the systematic nuisance parameters for a fixed value of the single-top cross section. The maximum value of this density function is the most probable value of the cross section, and the region that contains 68% of the area marks out the uncertainty. All sources of systematic uncertainty are included in this statistical treatment, including the correlation between normalization and discriminant shape changes due to sources of systematic uncertainty (e.g. the jet-energy-scale uncertainty).

The distribution of each of D0's discriminants are shown in Figure 5. Examining 0.9 fb^{-1} of data, D0 measures values of $4.8^{+1.6}_{-1.4}$ pb for the matrix element analysis, $4.4^{+1.6}_{-1.4}$ pb for the Bayesian neural network analysis, and $4.9^{+1.4}_{-1.4}$ pb for the boosted decision tree analysis [14]. The distribution of each of CDF's discriminants are shown in Figure 6. CDF uses 2.2 fb^{-1} of data to measure values of $1.8^{+0.9}_{-0.8}$ pb for the likeihood function analysis, $2.2^{+0.8}_{-0.7}$ pb for the matrix element analysis, and $2.0^{+0.9}_{-0.8}$ pb for the neural network analysis.

D0 combines its results using a best linear unbiased estimator. This takes advantage of the fact that the different analyses, while they use the same data, use different information and are thus not fully correlated. The combination

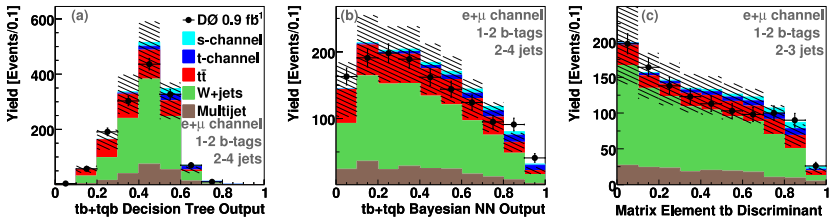

Figure 5: Discriminant distribution in Monte Carlo and data for D0's boosted decision tree analysis (left), Bayesian neural network analysis (center), and matrix element analysis (right).

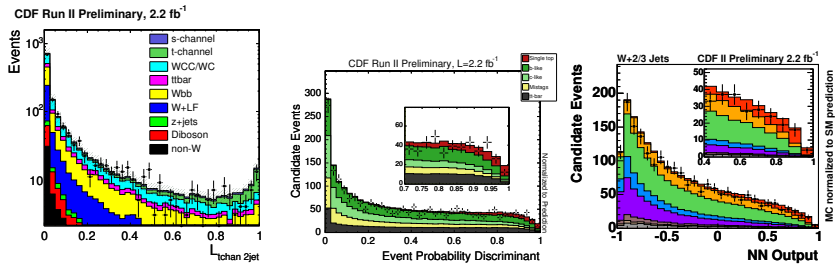

Figure 6: Discriminant distribution in Monte Carlo and data for CDF's multivariate likelihood function analysis (left), matrix element analysis (center), and neural network analysis (right).

thus has a better sensitivity than any of the three analyses. D0's combination measures a cross section of 4.7 ± 1.3 pb.

To quantify the signal significance of a result, pseudo-experiments are generated from events without a single top contribution. The probability (p-value) of the background-only pseudo-experiments to fluctuate to the observed result in data is the signficance of the result. D0 uses the measured cross section as a test statistic. The expected sensitivity the fraction of events which have a cross section higher than the Standard-Model expectation. The expected p-value of D0's combination is 0.011, which corresponds to a 2.3σ signal significance. The observed p-value is 0.00014, which corresponds to a 3.6σ excess (Figure 7).

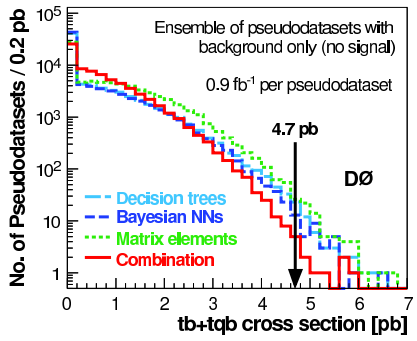

Figure 7: Distribution of cross sections for pseudo-experiments made with no signal included, showing the resulting p-value of 0.011.

5 Measurement of $|V_{tb}|$

The quantity $|V_{tb}|$ can be calculated from the single top cross section, which is directly proportional to $|V_{tb}|^2$. Assuming, based on branching ratio measurements on top quarks [15], that $|V_{td}|^2 + |V_{ts}|^2 << |V_{tb}|^2$, and integrating a flat prior in $|V_{tb}|^2$, requiring $|V_{tb}|$ to be between 0 and 1, D0 obtains $|V_{tb}| > 0.68$ at a 95% confidence level. Using the same method on its matrix element analysis, CDF calculates $|V_{tb}| > 0.6$ at a 95% confidence level (Figure 8). The most probable value of $|V_{tb}|$ is the square root of the cross section divided by the Standard Model prediction. D0 measures $|V_{tb}| = 1.3 \pm 0.2$ and CDF measures $|V_{tb}| = 0.88^{+0.16}_{-0.14}(\text{stat} + \text{sys}) \pm 0.07(\text{theory})$.

6 Conclusions

We have performed searches for electroweak single top quark production at the Tevatron using 0.9 fb^{-1} of data collected with the D0 detector and 2.2 fb^{-1} of data collected with the CDF detector. Both experiments see evidence of single top production. D0 measures a combined cross section of 4.7 ± 1.3 pb while CDF measures cross sections between $1.8^{+0.9}_{-0.8}$ pb and $2.2^{+0.8}_{-0.7}$ pb. D0 sets a limit at a 95% confidence level of $|V_{tb}| > 0.68$ and measures $|V_{tb}| = 1.3 \pm 0.2$, while CDF calculates $|V_{tb}| > 0.6$ and measures $|V_{tb}| = 0.88^{+0.16}_{-0.14}(\text{stat} + \text{sys}) \pm 0.07(\text{theory})$.

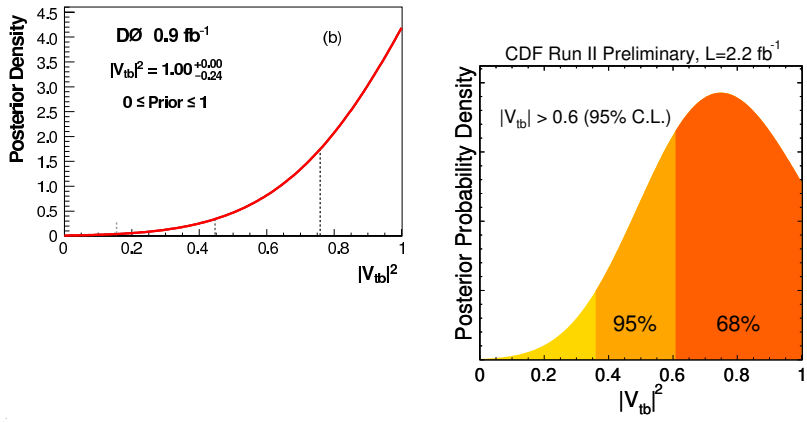

Figure 8: Likelihood curve for $|V_{tb}|^2$, in D0's combined analysis (left) and CDF's matrix element analysis (right), showing the limit on its value.

7 Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions.

References

1. B. W. Harris *et. al.*, *Phys. Rev.* D **66**, 054024 (2002)
   Z. Sullivan *Phys. Rev.* D **70**, 114012 (2004).
2. G. Mahlon, *Phys. Rev.* D **55**, 7249 (1997), hep-ph/0011349.
3. T. M. P. Tait and C.-P. Yuan, *Phys. Rev.* D **63**, 014018 (2002).
4. CDF and DØ Collaborations, FERMILAB-PUB-03/320-E (2003).
5. F. Caravaglios *et al*, Nucl. Phys. B **632**, 343 (2002)
   M. L. Mangano *et al*, JHEP 0307:001 (2003).
6. [CDF Collaboration], D. Acosta *et al*, Phys. Rev. **D71**, 052003 (2005).
7. Y. Freund and R. E. Schapire, in *Machine Learning: Proceedings of the Thirteenth International Conference*, edited by L. Saitta (Morgan Kaufmann, San Francisco, 1996), p. 148.

8. G. Mahlon and S. J. Parke, "Improved spin basis for angular correlation studies in single top quark production at the Tevatron," Phys. Rev. **D55**, 7249 (1997). T. Stelzer, Z. Sullivan and S. Willenbrock, "Single top quark production at hadron colliders," Phys. Rev. D **58**, 094021 (1998). [arXiv:hep-ph/9807340].

9. C. P. Yuan, "A New Method to Detect a Heavy Top Quark at the Tevatron," Phys. Rev. D **41**, 42 (1990).

10. K. Ackerstaff *et al.* (OPAL Collaboration), "Search for the standard model Higgs boson in e+ e- collisions at s**(1/2) = 161-GeV to 172-GeV," Eur. Phys. J. C **1**, 425 (1998). [arXiv:hep-ex/9709003].

11. M. Feindt and U. Kerzel, "The NeuroBayes neural network package," Nucl. Instrum. Methods A **559**, 190-194 (2006).

12. [D0 Collaboration] V. M. Abazov *el al*, "A Precision Measurement of the Mass of the Top Quark," Nature **429**, 638 (2004); F. Canelli, Ph.D. thesis, University of Rochester (2003); B. Stelzer, Ph.D. thesis, University of Toronto (2005); T. Gadfort, Ph.D. thesis, University of Washington (2007); J. Mitrevski, Ph.D. thesis, Columbia University (2007).

13. J. Pumplin *et al.*, J. High Energy Phys. **0709**, 028 (2007).

14. [D0 Collaboration] V.M. Abazov *et al*, "Evidence for Production of Single Top Quarks and First Direct Measurement of $|V_{tb}|$," Phys. Rev. Lett. **98**, 181802 (2007); [D0 Collaboration] V.M. Abazov *et al*, "Evidence for Production of Single Top Quarks," accepted in Phys. Rev. D [arXiv:hep-ex/0803.0739].

15. [D0 Collaboration] V.M. Abazov *et al*, accepted in Phys. Rev. Lett. [arXiv:hep-ex/0801.1326].