



## Search for Electroweak Single-Top-Quark Production using Neural Networks with $2.2 \text{ fb}^{-1}$ of CDF II data

The CDF Collaboration  
URL <http://www-cdf.fnal.gov>  
(Dated: February 27, 2008)

We report on a search for electroweak single-top-quark production with CDF II data corresponding to  $2.2 \text{ fb}^{-1}$  of integrated luminosity. We apply neural networks to construct discriminants that distinguish between single-top and background events. We combine  $t$ - and  $s$ -channel events to one single-top signal assuming the ratio of the two processes is given by the standard model (SM) and assuming a top quark mass of  $175 \text{ GeV}/c^2$ . A binned likelihood fit to the data yields a cross section of  $2.0_{-0.8}^{+0.9} \text{ pb}$  for single top-quark production.

*Preliminary Results for Winter 2008 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. The CDF and DØ collaborations have published single-top results at  $\sqrt{s} = 1.8$  TeV and  $\sqrt{s} = 1.96$  TeV [1–3]. The most recent result from the DØ collaboration [3] has seen evidence for single top quark production and measured  $\sigma_{s+t} = 4.9 \pm 1.4$  pb.

The theoretical single-top production cross section is  $\sigma_{s+t} = 2.9 \pm 0.4$  pb for a top mass of  $175$  GeV/ $c^2$  [4]. Despite this small rate, the main obstacle in finding single-top is in fact the large associated background. After all section requirements are imposed, the signal to background ratio is approximately 1/20. This challenging, background-dominated dataset is the main motivation for using multivariate techniques.

## II. COMMON EVENT SELECTION

The CDF event 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. One therefore demands a single high-energy electron or muon ( $E_T(e) > 20$  GeV, or  $P_T(\mu) > 20$  GeV/ $c$ ) and large missing transverse energy  $\cancel{E}_T > 25$  GeV from the undetected neutrino.

The backgrounds belong to the following categories:  $Wb\bar{b}$ ,  $Wc\bar{c}$ ,  $Wc$ , mistags (light quarks misidentified as heavy flavor jets), top pair production  $t\bar{t}$  events (one lepton or two jets are lost due to detector acceptance), non- $W$  (QCD multijet events where a jet is erroneously identified as a lepton),  $Z \rightarrow ll$  and diboson  $WW$ ,  $WZ$ , and  $ZZ$ . We remove a large fraction of the backgrounds by demanding two or three jets with  $E_T > 20$  GeV and  $|\eta| < 2.8$  be present in the event. At least one of these jets has to be tagged as a  $b$  quark jet by using displaced vertex information from the silicon vertex detector (SVX) of CDF [5]. The non- $W$  content of the selected electron dataset is further reduced by several requirements to transverse mass of the  $W$ -boson candidate, the missing transverse energy significance, the angle between the  $\cancel{E}_T$  vector and the transverse momentum vector and the angle between the charged lepton and the momentum vector of the jets. The numbers of expected and observed events are listed in table I.

Process	Number of Events			
	2-jet-bin		3-jet-bin	
	1-tag-bin	2-tag-bin	1-tag-bin	2-tag-bin
$t\bar{t}$	$94.8 \pm 13.3$	$21.1 \pm 3.5$	$204.1 \pm 28.5$	$60.3 \pm 9.9$
$Wb\bar{b}$	$376.2 \pm 113.4$	$49.7 \pm 15.5$	$106.7 \pm 32.2$	$17.6 \pm 5.5$
$Wc\bar{c}/Wc$	$361.4 \pm 111.4$	$4.8 \pm 1.6$	$92.7 \pm 28.5$	$2.4 \pm 0.8$
Mistags	$308.3 \pm 51.1$	$1.2 \pm 0.4$	$88.6 \pm 14.8$	$0.9 \pm 0.3$
Non- $W$	$55.8 \pm 22.3$	$1.5 \pm 0.6$	$21.3 \pm 8.5$	$0.2 \pm 0.1$
Diboson	$52.4 \pm 5.2$	$3.2 \pm 0.3$	$16.7 \pm 1.7$	$1.1 \pm 0.1$
$Z$ +jets	$19.1 \pm 2.8$	$0.9 \pm 0.1$	$7.1 \pm 1.0$	$0.5 \pm 0.1$
$t$ -channel	$50.6 \pm 7.4$	$1.4 \pm 0.2$	$13.1 \pm 1.9$	$2.1 \pm 0.3$
$s$ -channel	$26.3 \pm 3.7$	$7.6 \pm 1.2$	$8.2 \pm 1.2$	$2.7 \pm 0.4$
total background	$1268.0 \pm 319.5$	$82.4 \pm 22.0$	$537.2 \pm 115.2$	$83.0 \pm 16.8$
total single-top	$76.9 \pm 11.1$	$9.0 \pm 1.4$	$21.3 \pm 3.1$	$4.8 \pm 0.7$
total prediction	$1345.0 \pm 231.9$	$91.3 \pm 17.6$	$558.7 \pm 68.8$	$87.8 \pm 11.6$
observation	1312	82	491	95

TABLE I: Expected number of signal and background events and total number of events observed in  $2.2$  fb $^{-1}$  in the CDF single-top dataset. Note, the NN analysis does not use loose muons.

### III. NEURAL NETWORK INPUT VARIABLES

Using neural networks many kinematic or event shape variables are combined to a powerful discriminant. In the search for single top-quark production, four different neural networks are trained in different jet and tag bins:

- 2 jets, 1 tag:  $t$ -channel
- 2 jets, 2 tags:  $s$ -channel
- 3 jets, 1 tag:  $t$ -channel
- 3 jets, 2 tags:  $t$ -channel

The 2jets, 1tag sample is the largest sub sample and dominates so the search for single top. The four most important variables of the the NN in the 2 jets, 1tag sample are the reconstructed top quark mass, the KIT flavor separator, the invariant mass of the two jets and the product of the lepton-charge and the pseudorapidity of the light quark, see figure 1. For each variable the signal and background shapes, a data Monte-Carlo comparison and a check of the background shape in the zero-tag sample are shown in figure 1. The KIT flavor separator gives an additional handle to reduce the large background components where no real  $b$  quarks are contained, mistags and charm-backgrounds. Both of them amount to about 50% in the  $W+2$  jets data sample even after imposing the requirement that one jet is identified by the secondary vertex tagger of CDF [5].

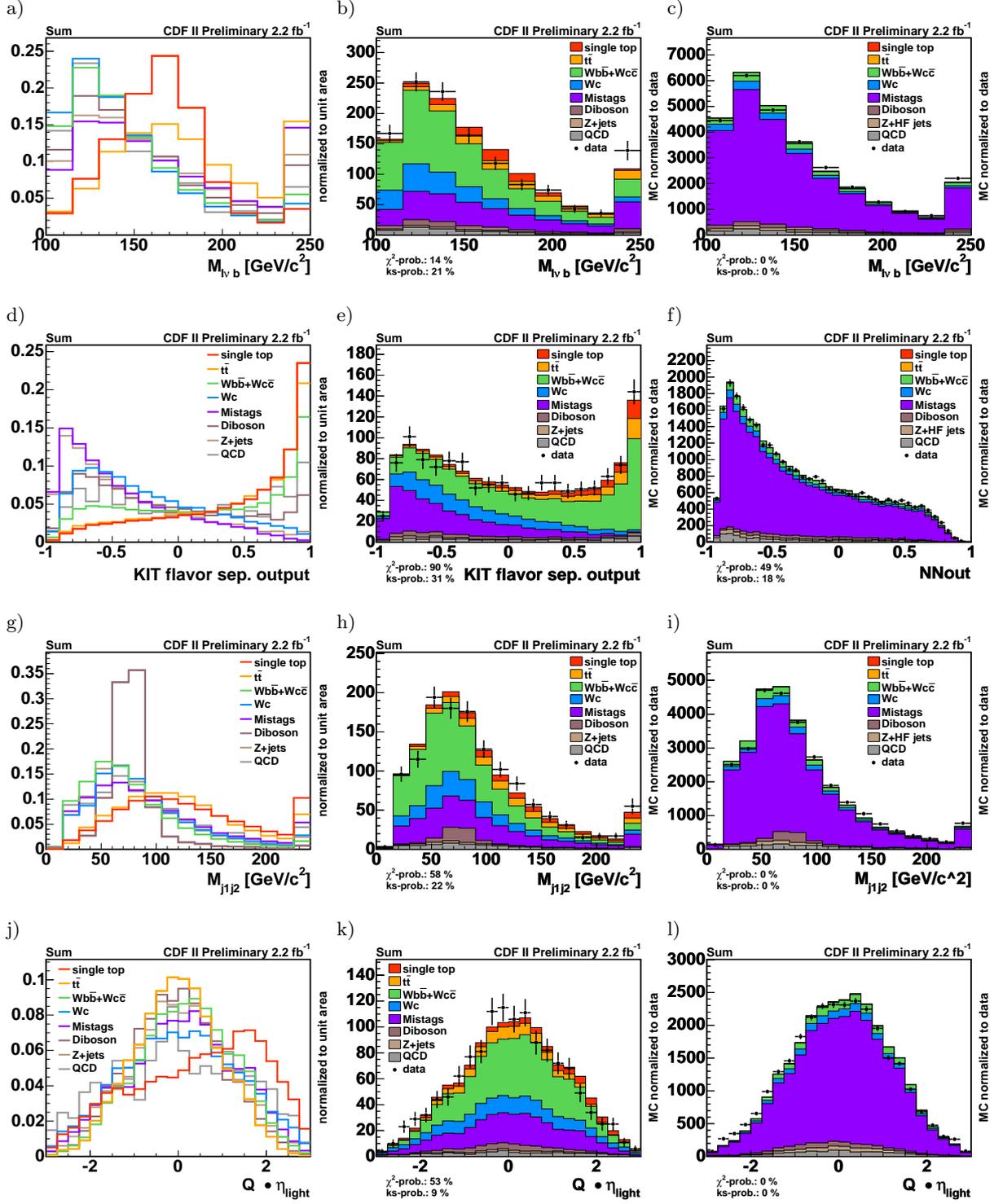


FIG. 1: Some of the most powerful variables in the 2jet, 1tag sample. Left: signal and background shapes, Middle: Data-Monte Carlo comparison, Left: Check of background shapes in the W+2jet zero tag sample.

#### IV. TEMPLATES FOR BINNED LIKELIHOOD FIT

The training of a neural network results in one output variable continuously distributed between  $-1$  and  $1$ . The output of the different neural networks is used to create signal and background templates which are to be fitted to the output distribution of observed events. We perform a combined single top search meaning that the output distributions of both  $t$ - and  $s$ -channel events are combined into one single distribution, where the ratio between the two processes is as predicted by the standard model. In the fit, all considered sub sample (2 jets 1tag, 2 jets 2tags, 3 jet 1 tag and 3jet 2 tags) are fitted simultaneously to determine the combined single-top cross section.

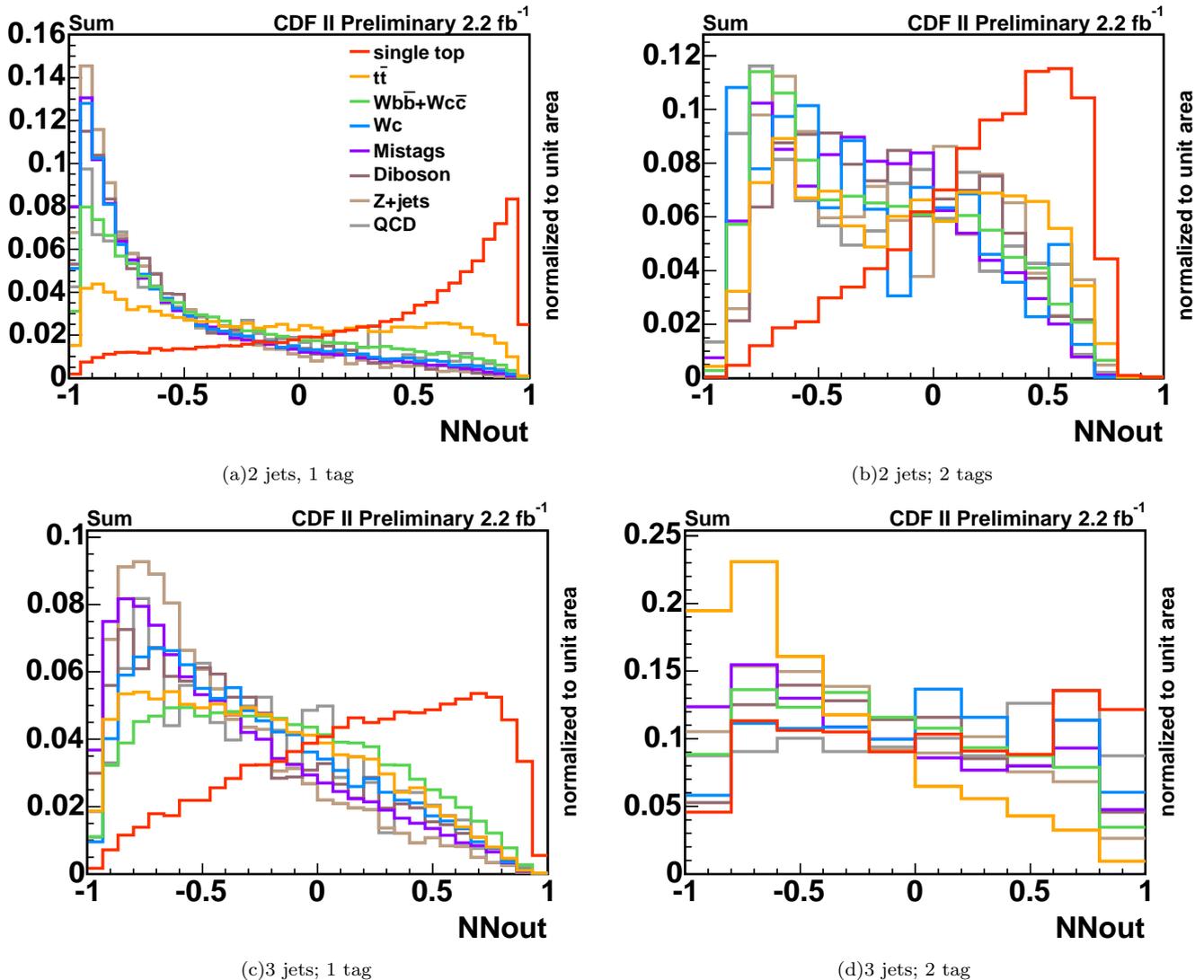


FIG. 2: The templates for the combined search are built using the following networks: the (a)  $t$ -channel neural network in the 2-jet bin with 1  $b$ -tag, (b)  $s$ -channel neural network in the 2-jet bin with 2  $b$ -tags, and the (c)  $t$ -channel neural network in the 3-jet bin with 1  $b$ -tag and with 2  $b$ -tags (d). The output of  $t$ - and  $s$ -channel events is added with a ratio corresponding to the standard-model prediction.

Figure 2 shows the templates of the combined single top search. The  $t$ -channel templates in the 2-jet bin with 1  $b$ -tag and in the 3-jet bin with 1 and 2  $b$ -tags are presented in figure 2(a), 2(c) and 2(d), while the  $s$ -channel templates in the 2-jet bin with 2  $b$ -tags are presented in figure 2(b).

## V. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties can cause a shift in the event detection efficiency for events of different physics processes, but can also cause a change in the shape of the template distributions.

The rate uncertainties are summarized in tables II-V separately for the four sub samples. They are only determined for single-top-quark and  $t\bar{t}$  events because the rates of the main backgrounds,  $W$ +jets and non- $W$  events, are estimated based on the observed rate of events before  $b$  tagging or by a fit to the observed missing transverse energy distribution, respectively.

The following sources of systematic shape uncertainties are considered: the jet energy scale (JES), initial state gluon radiation (ISR), final state gluon radiation (FSR), parton distribution functions (PDFs), neural net  $b$  tagger, the factorization and renormalization scale for  $W$  + heavy flavor processes, the reweighting of the mismodelled variables  $\Delta R_{j1j2}$  and  $\eta_{j2}$ , the modeling of mistag events, the flavor composition and modeling of non- $W$  events.

The shape uncertainties are determined by altering the respective effects within their uncertainties. In this way two shifted distributions are obtained for first five sources (see three examples in figure 3), one plus and one minus distribution. For the last five systematic sources one alternative model is considered. Therefore, only one systematic shape is obtained for these effects.

Source	$t$ -channel	$s$ -channel	single-top	$t\bar{t}$
ISR less/more	2.8/-0.2 %	0.3/6.7 %	1.9/2.1 %	-2.6/-7.1 %
FSR less/more	4.2/-1.3 %	5.9/0.4 %	4.8/-0.7 %	-5.1/-2.6 %
PDF	3.4/-3.4 %	2.2/-2.2 %	3.0/-3.0 %	1.8/-1.8 %
MC	2.0/-2.0 %	1.0/-1.0 %	1.7/-1.7 %	-2.7/2.7 %
$\epsilon_{\text{evt}}$	4.2/-4.2 %	2.3/-2.3 %	3.6/-3.6 %	2.9/-2.9 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.6/-12.6 %	12.4/-12.4 %
$M_{\text{top}}$ 170/180	1.3/-0.8 %	2.4/-1.7 %	1.7/-1.1 %	-3.1/1.4 %
	Diboson	Z+jets		
$\epsilon_{\text{evt}}$	7.6/-7.6 %	8.3/-8.3 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Cross section	1.9/-1.9 %	10.8/-10.8 %		

TABLE II: Systematic rate uncertainties for 2 jets and 1  $b$  tag

Source	$t$ -channel	$s$ -channel	single-top	$t\bar{t}$
ISR less/more	-4.9/-6.9 %	1.3/9.2 %	0.4/6.7 %	0.5/-9.5 %
FSR less/more	3.9/-6.6 %	8.1/2.2 %	7.5/0.8 %	-8.1/-1.8 %
PDF	2.0/-2.0 %	2.0/-2.0 %	2.0/-2.0 %	1.7/-1.7 %
MC	2.0/-2.0 %	1.0/-1.0 %	1.2/-1.2 %	4.6/-4.6 %
$\epsilon_{\text{evt}}$	10.0/-10.0 %	8.7/-8.7 %	8.9/-8.9 %	9.0/-9.0 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.5/-12.5 %	12.4/-12.4 %
$M_{\text{top}}$ 170/180	-4.7/-4.1 %	2.1/0.1 %	1.0/-0.5 %	0.4/3.0 %
	Diboson	Z+jets	Mistags	
$\epsilon_{\text{evt}}$	9.8/-9.8 %	10.6/-10.6 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Double tag			23.4/-23.4%	
Cross section	1.9/-1.9 %	10.8/-10.8 %		

TABLE III: Systematic rate uncertainties for 2 jets and 2  $b$  tags

Source	$t$ -channel	$s$ -channel	single-top	$t\bar{t}$
ISR less/more	-6.8/-0.0 %	2.4/-12.6 %	-3.3/-4.8 %	-0.6/-4.6 %
FSR less/more	-1.5/-3.1 %	-6.0/-4.8 %	-3.3/-3.8 %	-3.5/-2.2 %
PDF	2.7/-2.7 %	2.3/-2.3 %	2.6/-2.6 %	1.8/-1.8 %
MC	1.9/-1.9 %	1.5/-1.5 %	1.7/-1.7 %	-1.7/1.7 %
$\epsilon_{\text{evt}}$	3.5/-3.5 %	2.3/-2.3 %	3.0/-3.0 %	2.3/-2.3 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.6/-12.6 %	12.4/-12.4 %
$M_{\text{top}}$ 170/180	1.5/-2.8 %	6.0/-2.7 %	3.2/-2.7 %	-0.7/0.8 %
	Diboson	Z+jets		
$\epsilon_{\text{evt}}$	7.8/-7.8 %	7.8/-7.8 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Cross section	1.9/-1.9 %	10.8/-10.8 %		

TABLE IV: Systematic rate uncertainties for 3 jets and 1  $b$  tag

Source	$t$ -channel	$s$ -channel	single-top	$t\bar{t}$
ISR less/more	7.8/3.2 %	4.3/-11.2 %	5.8/-4.9 %	-0.5/-6.6 %
FSR less/more	15.0/1.3 %	-7.4/-5.0 %	2.4/-2.2 %	-3.4/-2.7 %
PDF	1.5/-1.5 %	2.1/-2.1 %	1.9/-1.9 %	1.7/-1.7 %
MC	1.9/-1.9 %	1.5/-1.5 %	1.7/-1.7 %	2.0/-2.0 %
$\epsilon_{\text{evt}}$	9.1/-9.1 %	8.8/-8.8 %	8.9/-8.9 %	9.1/-9.1 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.5/-12.5 %	12.4/-12.4 %
$M_{\text{top}}$ 170/180	4.2/3.0 %	1.6/-6.8 %	2.7/-2.5 %	-0.6/-1.0 %
	Diboson	Z+jets	Mistags	
$\epsilon_{\text{evt}}$	10.8/-10.8 %	11.1/-11.1 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Double tag			23.4/-23.4 %	
Cross section	1.9/-1.9 %	10.8/-10.8 %		

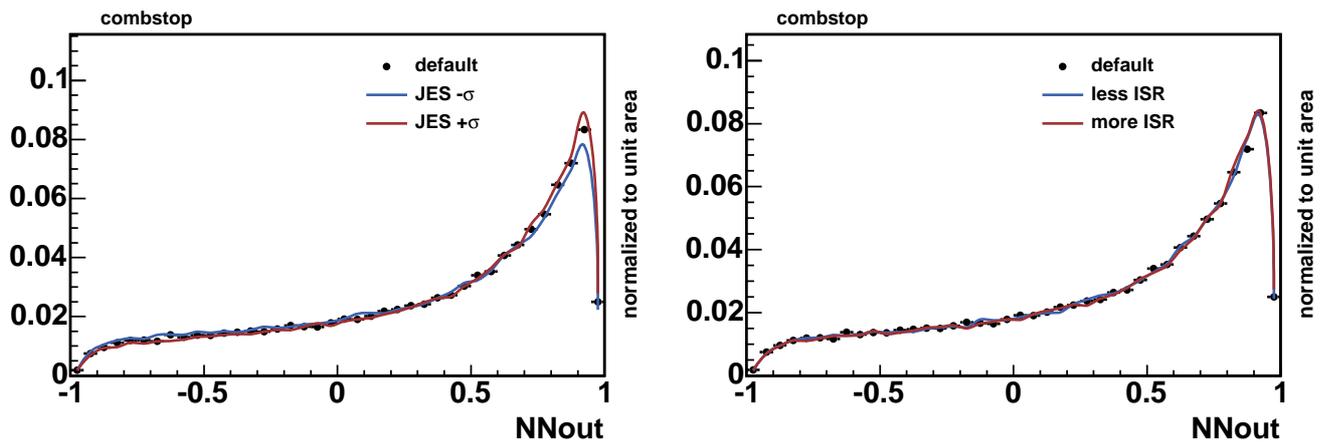
TABLE V: Systematic rate uncertainties for 3 jets and 2  $b$  tags

FIG. 3: Example of shape uncertainties. Shown is the comparison between the default distribution and the shifted distribution. Left: Shape systematics due the uncertainty on the jet energy correction for single-top. Right: Shape systematics due the uncertainty in the final state radiation for single-top.

## VI. LIKELIHOOD FUNCTION

The likelihood function consists of Poisson terms for the individual bins of the fitted histograms, Gaussian constraints to the background rates, and Gaussian constraints to the strengths of systematic effects.

$$L(\beta_1, \dots, \beta_5; \delta_1, \dots, \delta_S) = \prod_{k=1}^B \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \cdot \prod_{j=2}^5 G(\beta_j, 1.0, \Delta_j) \cdot \prod_{i=1}^S G(\delta_i, 0.0, 1.0); \quad (1)$$

Systematic uncertainties are included as factors modifying the expectation value  $\mu_k$  of events in a certain bin  $k$ .

$$\mu_k = \sum_{j=1}^5 \beta_j \cdot \nu_j \cdot \mathcal{L}_{\text{int}} \cdot \left\{ \prod_{i=1}^S (1 + \delta_i \cdot \epsilon_{ji}) \right\} \cdot \alpha_{jk} \cdot \left\{ 1 + \sum_{i=1}^S (\delta_i \cdot \kappa_{jik}) \right\} \quad (2)$$

The index  $j$  runs over the different physics processes that occur in the likelihood function. The cross section of process  $j$  is  $\sigma_j$ . In the likelihood function we use the parameter  $\beta_j$ , which is the cross section normalized to its standard model prediction. The event detection efficiency of process  $j$  is named  $\nu_j$ . The normalized content of bin  $k$  of the template histogram for process  $j$  is  $\alpha_{jk}$ .

Rate uncertainties as well as uncertainties in shape are considered. The sources of systematic uncertainties are indexed with  $i$ . The relative acceptance uncertainties due to these sources are named  $\epsilon_{ji}$ . The relative uncertainties in the bin content of bin  $k$  of the template histograms are called  $\kappa_{jik}$ . The variation in strength of a systematic effect  $i$  is measured with the variable  $\delta_i$ .

The shape uncertainties are calculated from the systematically shifted histograms  $\alpha_{jik}^+$  and  $\alpha_{jik}^-$  according to

$$\kappa_{jik} = \frac{\alpha_{jik}^+ - \alpha_{jik}^-}{2\alpha_{jk}} \quad (3)$$

By construction the  $\kappa_{jik}$  satisfy the normalization condition

$$\sum_{k=1}^B \kappa_{jik} = 0. \quad (4)$$

The systematically shifted template that takes into account the shifts caused by all systematic effects with strengths  $\{\delta_i\}$  is given by

$$\alpha'_{ji} = \alpha_{jk} \cdot \left\{ 1 + \sum_{i=1}^S \delta_i \cdot \kappa_{jik} \right\} \quad (5)$$

Due to (4) the shifted histogram  $\alpha'_{ji}$  is properly normalized:

$$\sum_{k=1}^B \alpha'_{ji} = 1. \quad (6)$$

The background rates (cross sections) and the parameters describing the strength of systematic excursions ( $\delta_i$ ) are constrained by additional Gaussian terms in the likelihood. The background rates are constraint within the uncertainties of the prediction,  $\Delta_j$ . The strengths of the systematic effects are constraint to 0.0 with a standard deviation of 1.0. The single-top content (cross section) is measured by fitting the parameters ( $\beta_j$  and  $\delta_j$ ) of the likelihood function to the observed data. This is achieved by minimizing the log likelihood with respect to these parameters using the program MINUIT.

Using this technique one can compute the likelihood as a function of the single-top cross section,  $\beta_1$ , only. The log likelihood is minimized at a fixed value of  $\beta_1$  with respect to all other variables which are also often called *nuisance* parameters. The resulting one-dimensional function is called the reduced likelihood,  $\mathcal{L}_{\text{red}}(\beta_1)$ . This method is often called *profiling* the likelihood function.

## VII. EXPECTED SENSITIVITY AND SIGNIFICANCE

We use ensemble tests to compute the sensitivity of our analysis. An ensemble test consists of a set of pseudo experiments. For each pseudo experiment we determine first the number of events  $N_j$  of each process by drawing a random number from a Poisson distribution with a mean  $\mu_j$ . In a second step we draw random numbers from the template distributions of the neural network output.

We perform two ensemble tests: one with single-top events included at the predicted standard model rate, one without any single-top events. For each pseudo experiment, the single-top-quark cross section is determined using the same method as used for observed events. Based on the ensemble test with single-top-quark events included, the root mean square (RMS) of the resulting single-top-quark cross-section distribution is defined as the expected uncertainty of the measurement. We find a value of 0.75 pb including all systematic uncertainties, which is 26.3% of the predicted cross section as illustrated in figure 4.

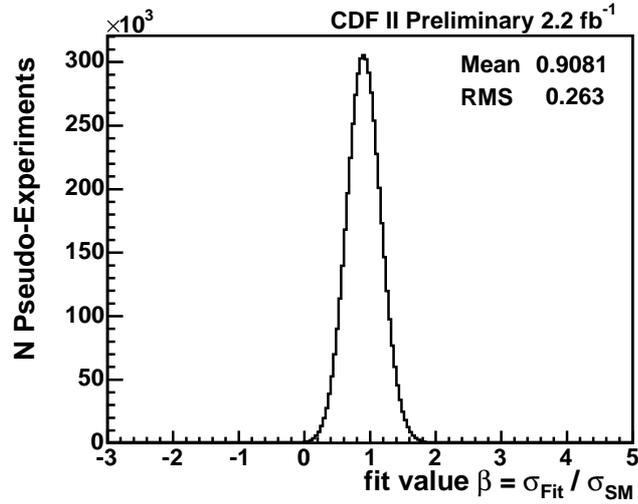


FIG. 4: The distribution of expected measurements in the combined search is shown. The fitted cross sections are based on the ensemble test with single-top-quark events included. The RMS value of this distribution is defined as the expected uncertainty of the measurement.

## VIII. BINNED LIKELIHOOD FIT TO DATA

The neural networks are applied to the observed events. The predicted and measured output distribution of all four neural networks used in the combined search are depicted in figure 5. In figure 6 the distributions of all four neural networks are added together.

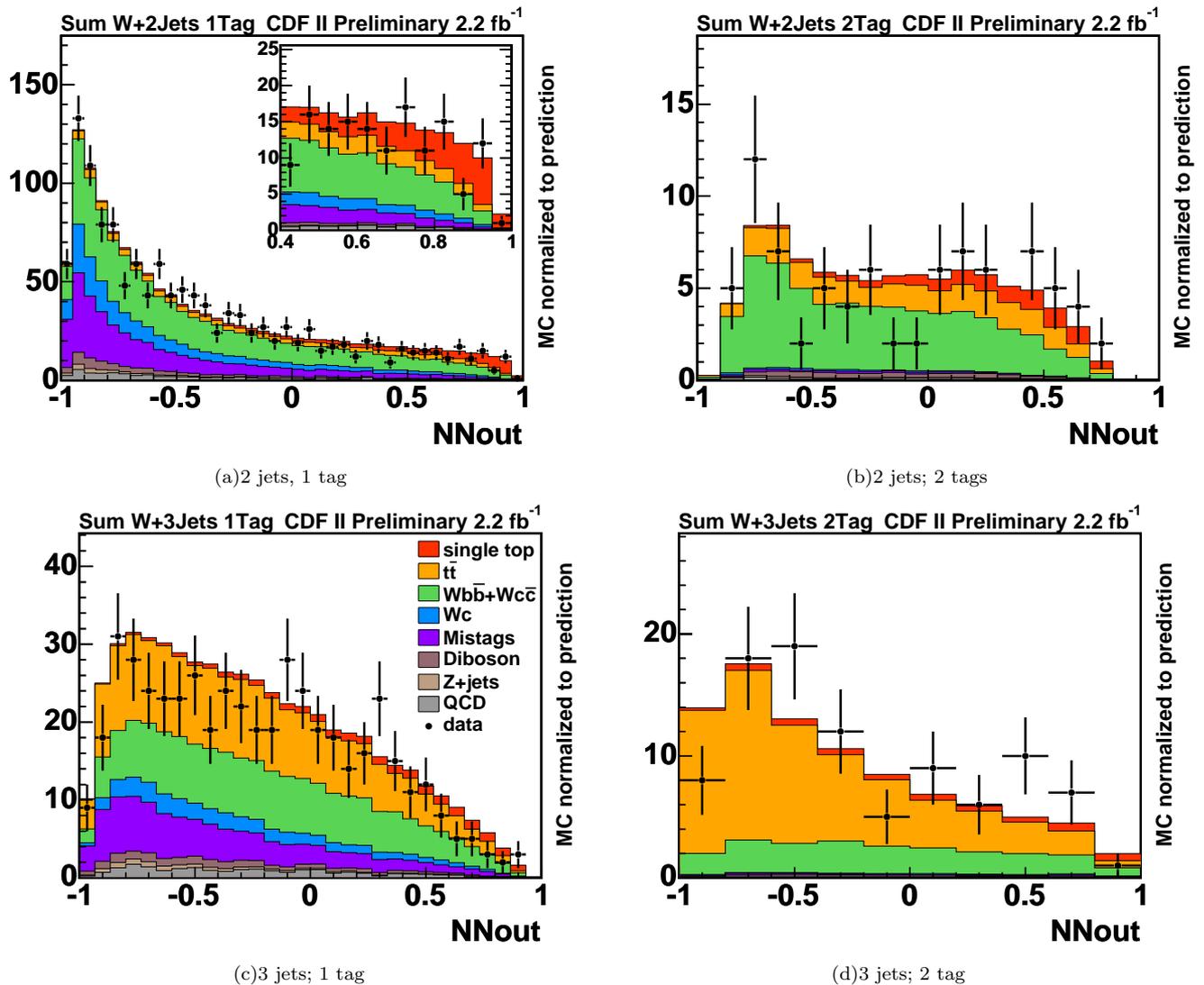


FIG. 5: The predicted and measured distributions of the combined search are built using the following networks: the (a)  $t$ -channel neural network in the 2-jet bin with 1  $b$ -tag, (b)  $s$ -channel neural network in the 2-jet bin with 2  $b$ -tags, and the (c)  $t$ -channel neural network in the 3-jet bin with 1  $b$ -tag and with 2  $b$ -tags (d). The output of  $t$ - and  $s$ -channel events is added with a ratio corresponding to the standard-model prediction.

Finally, the templates are fitted to the observed distributions to determine the single-top-quark cross section. The fit yields a single top cross section of  $2.0_{-0.8}^{+0.9}$ pb. As depicted in figure 7, the fitted distributions describe the observed output distributions of the four sub samples well. In figure 8 the distributions of all four neural networks are added together.

The distribution of the reconstructed top mass as well as the product of the lepton-charge and the pseudo-rapidity of the light-quark jet are presented in figure 9 for the high NN-output regions  $NN > 0.4$  and  $NN > 0.8$ .

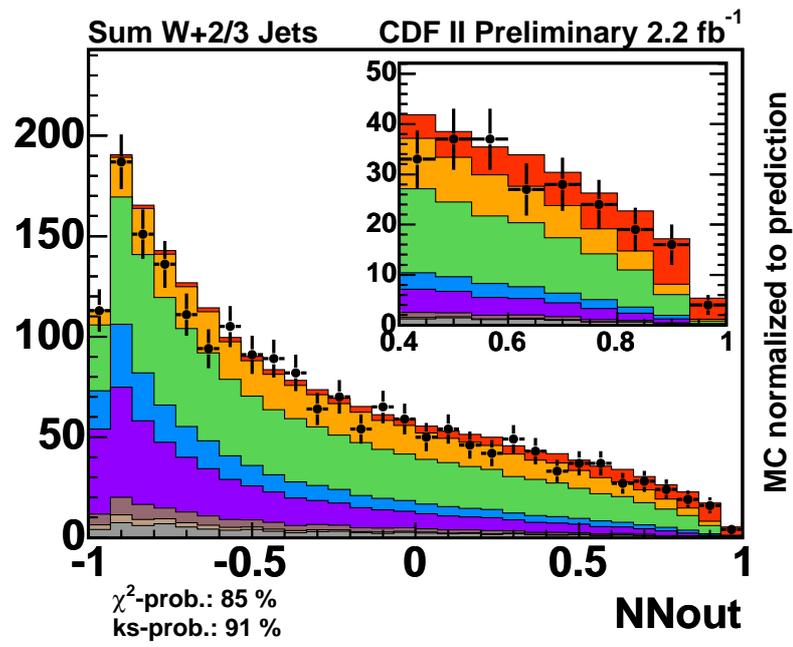


FIG. 6: The predicted and measured distributions of all four networks used in the combined search. The output of  $t$ - and  $s$ -channel events is added with a ratio corresponding to the standard-model prediction.

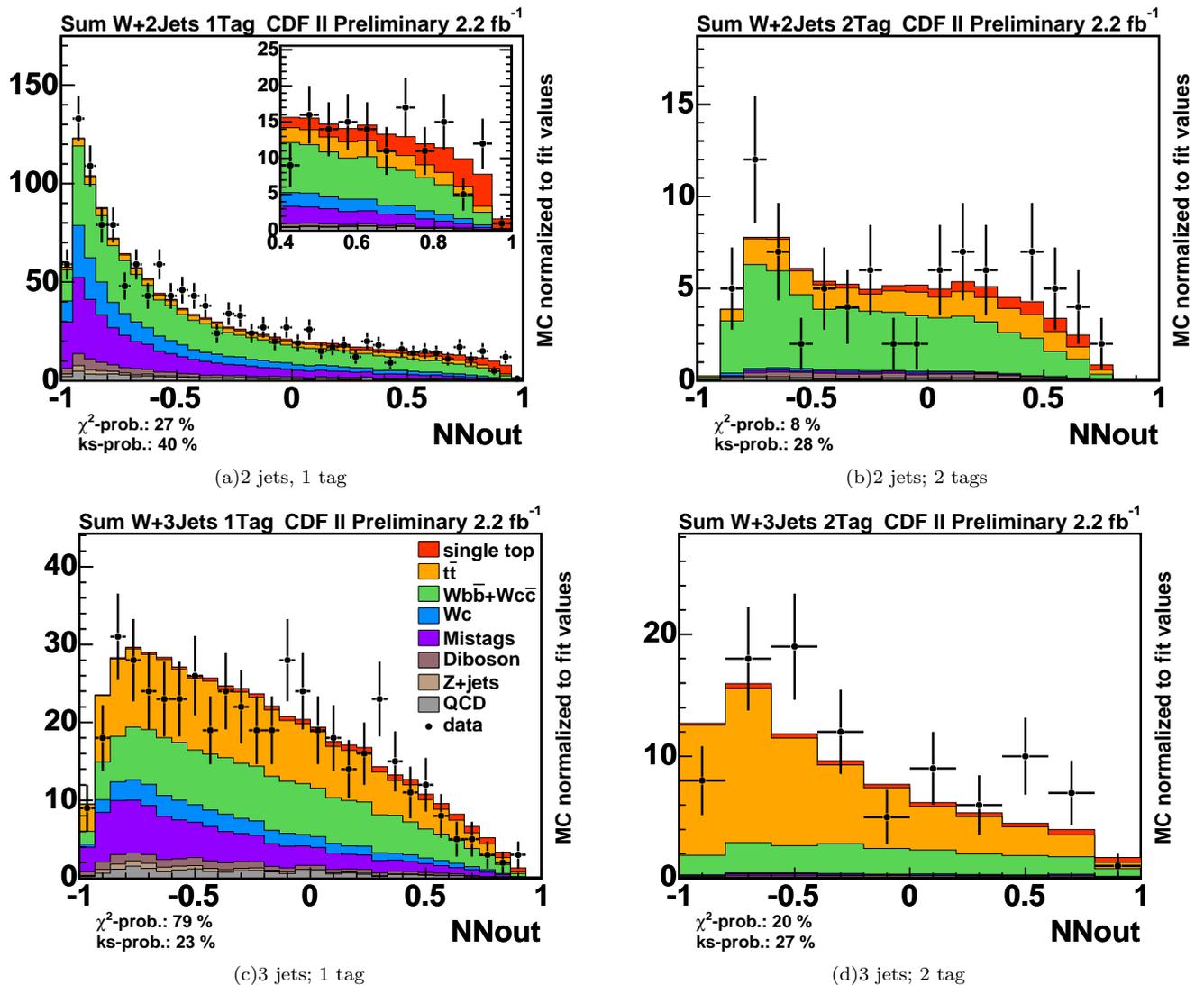


FIG. 7: The normalized to the fit results and measured distributions of the combined search are built using the following networks: the (a)  $t$ -channel neural network in the 2-jet bin with 1  $b$ -tag, (b)  $s$ -channel neural network in the 2-jet bin with 2  $b$ -tags, and the (c)  $t$ -channel neural network in the 3-jet bin with 1  $b$ -tag and with 2  $b$ -tags (d). The output of  $t$ - and  $s$ -channel events is added with a ratio corresponding to the standard-model prediction.

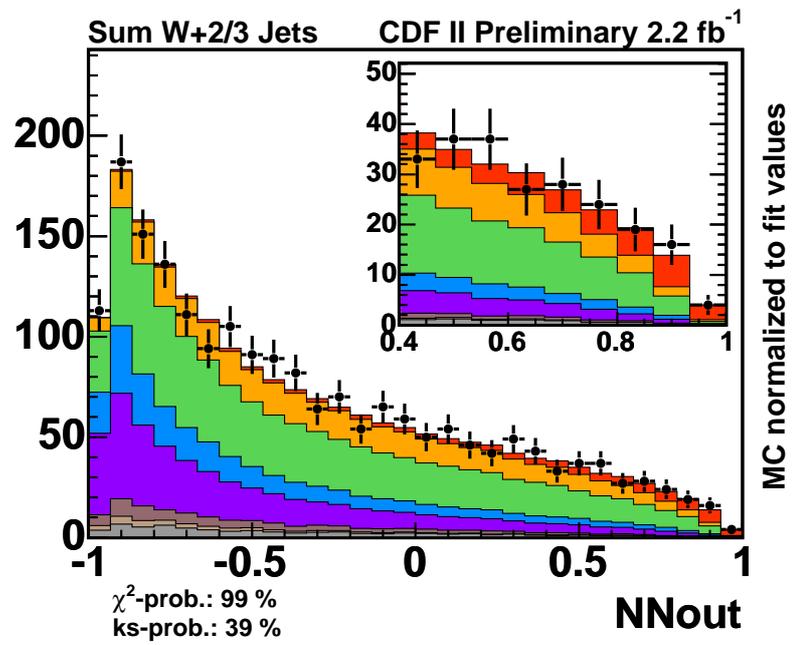


FIG. 8: The normalized to the fit results and measured distributions of all four networks used in the combined search. The output of  $t$ - and  $s$ -channel events is added with a ratio corresponding to the standard-model prediction.

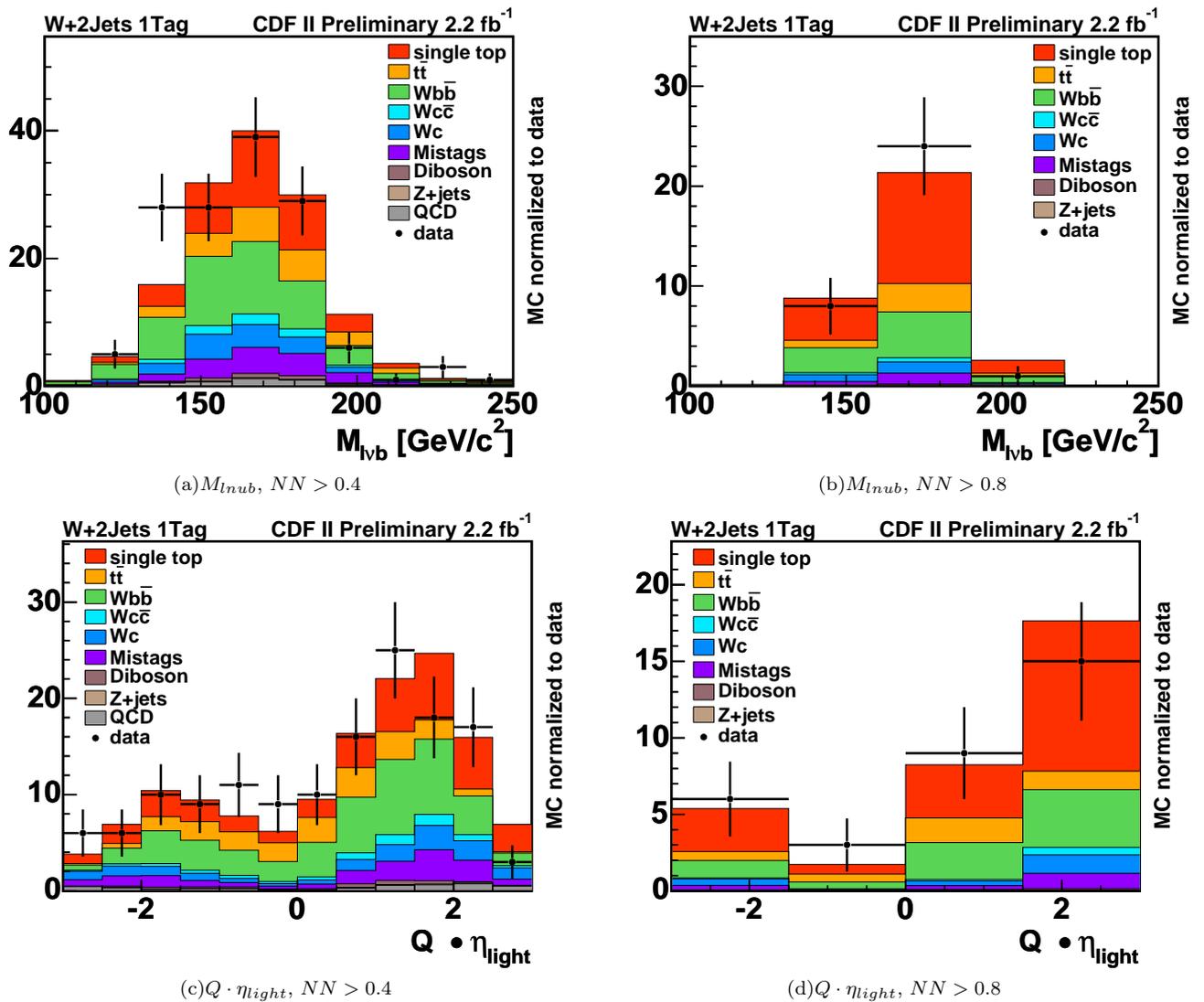


FIG. 9: High NN-output region. Shown are the reconstructed top mass and the product of the lepton-charge and the pseudo-rapidity of the light-quark jet. Left:  $NN > 0.4$ , right:  $NN > 0.8$ .

## IX. CONCLUSIONS

We present a search of single-top quark production in a CDF II data set corresponding to  $2.2 \text{ pb}^{-1}$ , assuming a top quark mass of  $175 \text{ GeV}/c^2$ . We employ neural networks to construct a discriminant between single-top events and background events. In data we see evidence for single top. We measure  $\sigma_{s+t} = 2.0_{-0.8}^{+0.9}$ .

- 
- [1] D. Acosta *et al.* (CDF Collaboration), *Phys. Rev.* **D65**, 091102 (2002); D. Acosta *et al.* (CDF Collaboration), *Phys. Rev.* **D69**, 052003 (2004); D. Acosta *et al.* (CDF Collaboration), *Phys. Rev.* **D71**, 012005 (2005).
  - [2] B. Abbott *et al.* (DØ Collaboration), *Phys. Rev.* **D63**, 031101 (2001); V. M. Abazov *et al.* (DØ Collaboration), *Phys. Lett.* **B517**, 282 (2001); V. M. Abazov *et al.* (DØ Collaboration), *Phys. Lett.* **B622**, 265 (2005).
  - [3] V. M. Abazov *et al.* (DØ Collaboration), *Phys. Rev. Lett.* **98**, 181802 (2007).
  - [4] B. W. Harris *et al.*, *Phys. Rev.* **D66**, 054024 (2002); Z. Sullivan, *Phys. Rev.* **D70**, 114012 (2004). J. Campbell, R.K. Ellis, F. Tramontano, *Phys. Rev.* **D70**, 094012 (2004).
  - [5] D. Acosta *et al.* (CDF Collaboration), *Phys. Rev.* **D71**, 052003.