DØ note 4428

Measurement of the $t\bar{t}$ production cross section in the all-jets channel

The DØ Collaboration
(Dated: April 8, 2004)

We have measured the top and anti-top quark ($t\bar{t}$) pair cross section using an event sample corresponding to 162 pb^{-1}, which was recorded by DØ in the period 2002/2003. The analysis concentrates on events where both t and $\bar{t}$ decay to a b-quark and a W-boson that decays hadronically. We isolate the $t\bar{t}$ events from the background using a secondary vertex tag and several kinematic and topological event quantities. These quantities are combined using artificial neural networks to achieve optimal discrimination between $t\bar{t}$ signal and background events. We observe 220 events with an expected background of $186 \pm 5(\text{stat}) \pm 12(\text{syst})$ events, with a signal efficiency of $\epsilon \cdot \text{BR} = 0.0275 \pm 0.0009(\text{stat})^{+0.0086}_{-0.0079}(\text{syst})$, which corresponds to a cross section of $\sigma(t\bar{t}) = 7.7^{+3.4}_{-3.3}(\text{stat})^{+4.7}_{-3.8}(\text{syst}) \pm 0.5(\text{lumi})$ pb.

Preliminary Results for Winter 2004 Conferences

I. INTRODUCTION

This note describes the analysis of the $t\bar{t} \rightarrow$ all-jets production cross section with a dataset taken between July 2002 and January 2004, in $p\bar{p}$ collisions with $\sqrt{s} = 1.96$ TeV.

This analysis concentrates on the $t\bar{t}$ pair production in the 'all-jet' channel. The top quarks decay to a b quark and a W-boson, which subsequently decays hadronically. This channel constitutes 46% of the total $t\bar{t}$ production cross section, larger than any other channel with one or both W-bosons decaying to leptons.

One of the experimental difficulties in this $t\bar{t}$ decay channel is the overwhelming QCD background, which is three to four orders of magnitude larger than the expected signal. The identification of the b-jets using secondary vertex tagging (SVT) is by itself not sufficient to reduce the background. The analysis strategy consists of the determination of the background content in our b-tagged sample, after which we do a multivariate analysis with artificial neural nets to extract our signal.

These neural networks are trained using Monte Carlo simulated $t\bar{t}$ events and a background sample which consists of a fraction of the untagged data events. A $t\bar{t}$ enriched sample is obtained by selecting with a high neural network output probability. The final background content in this sample is determined from untagged data. Note that we use $t\bar{t}$ Monte Carlo *only* for neural network training, while the background content is entirely derived from data.

First, we will discuss our tagged sample and the parametrization of the QCD background with tag rate functions (TRF), after which we will continue with the kinematic variables, which are then used in a chain of top-discriminating neural networks. We will shortly discuss the neural network training, after which we use the neural networks to extract the $t\bar{t}$ production cross section.

II. MONTE CARLO SIMULATION

We use ALPGEN 1.2 [1] and parton distribution function CTEQ 6.1M [2] to generate our top-anti top pair production, after which we use PYTHIA 6.2 [3] for the simulation of gluon radiation and hadronization processes. We separately study $t\bar{t}$ events where both W-bosons decay hadronically, and where one of the two W-bosons in the event decays to a lepton-neutrino pair. Our simulated signal is passed through a full detector simulation. We use the same reconstruction for data and Monte Carlo events.

III. DATA SAMPLE AND EVENT SELECTION

A. Trigger

The data is collected with a dedicated trigger at an average trigger efficiency of 74%. This number is derived by application of the jet-level turn-on curves (measured on electron+jets data) to our $t\bar{t}$ Monte Carlo simulation. Only Monte Carlo simulated events that passed the preselection are used. We make the assumption that our jet algorithms are not sensitive to the number of jets in our event. Although we expect these correlations to be relatively small, we plan to study them in the near future. Our trigger requirements depend on the version of the trigger list, and are listed in table I. The integrated luminosity for our trigger is 162 pb^{-1}, after rejection of events in luminosity blocks with bad calorimeter or tracking performance.

B. Pre-selection

We use events with six or more jets, reconstructed using a standard η-ϕ jet-finding cone algorithm with radius 0.5 . We then make additional cuts on jet level:

- Jet $p_T > 15$ GeV
- Jet $|\eta| < 2.5$

	Luminosity per trigger		
	v8	**v9-11**	**v12**
name	4TJ10	4JT10	4JT12
L1	CJT(4,5)	CJT(4,5)	CJT(3,5)
L2	JET(3,8)	JET(3,8) HT(90)	JET(3,8) HT(50)
L3	L3FJet(SCJET_9,4,10)	L3FJet(SCJET_9,4,10) L3FJet(SCJET_9,2,20)	L3FJet(SC5JET_9_PV1,4,12) L3FJet(SC5JET_9_PV1,3,15) L3FJet(SC5JET_9_PV1,2,25)
luminosity	20.8 pb^{-1}	102.7 pb^{-1}	38.9 pb^{-1}

TABLE I: *The $t\bar{t} \to$ all-jets triggers for different trigger list versions.*

We reject events that contain isolated leptons or have a primary vertex with $N_{tracks} < 3$ or $|z| > 60$ cm. We use the electron and muon identification criteria as in the $t\bar{t} \to l +$ jets analyses, to obtain an orthogonal dataset. The lepton requirements reduce our data sample by less than 1% and have even less influence in signal Monte Carlo.

After pre-selection, our dataset consists of 292k events.

IV. SVX TAGGING

We use the Secondary Vertex Tagger (SVT) to identify b-jets. We define a jet to be tagged when there is a secondary vertex with a signed vertex lifetime significance greater than 7.0 associated with the jet. We require exactly one tagged jet in our events, because the probability to tag a second jet is larger because of the change in flavor content of our sample. The studies needed to also include double-tagged events will be done in the near future.

The efficiencies to tag b, c and light quarks and gluon jets are measured using data. These efficiencies are parametrized as a function of jet p_T for different η bins. When these parametrizations are applied to the $t\bar{t}$ MC, we find the efficiency for a hadronic $t\bar{t}$ event to have only one tag to be 46%.

The uncertainties on the parametrizations are taken into account in our systematic error calculation.

V. BACKGROUND MODELING

We parametrize our background by the application of tag rate functions (TRFs). As both our tagged *and* untagged sample is still completely dominated by background, we use them to determine the probability that a 'background' jet will be tagged:

$$P_{jet}(p_T, \eta) = N f(p_T) g(\eta).$$

As $f(p_T)$ and $g(\eta)$ are observed to be uncorrelated, we can make this factorization. We parametrize $P_{jet}(p_T, \eta)$ in four different bins of H_T. N is used to normalize our probability density function to unity. After this normalization we are able to accurately predict the number of tagged jets. This is illustrated in figure 1a, which shows the number of observed and predicted SVX tagged jets.

The probability for an event to have one single tag is then

$$P_{event}(tags = 1) = \sum_{i=1}^{i=N_{jet}} P_{jet}(i) \prod_{j}^{i \neq j} (1 - P_{jet}(j)),$$

where we presume that the tagging probabilities are uncorrelated between jets.

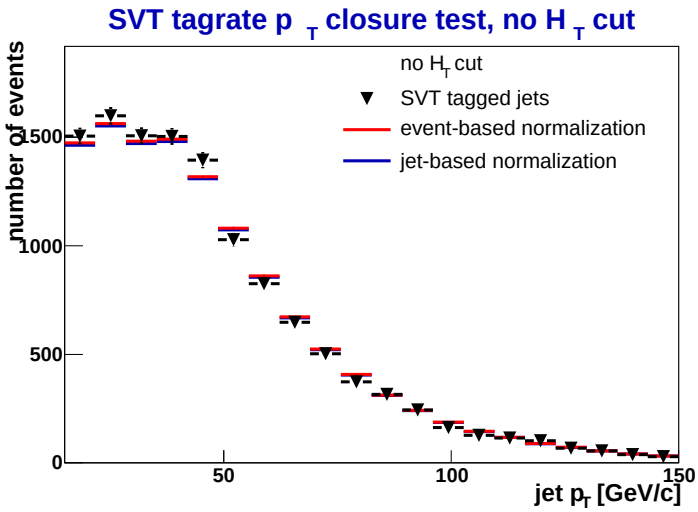

FIG. 1: *Test of the normalization of our tag rate functions: Closure test as a function of jet* p_T.

We studies the dependence of the normalization factor on the selection criteria, and determined a value of

$$N = 1.04 \pm 0.07.$$

The error of the measured normalization factor gives a handle on the TRF accuracy, and is used as the uncertainty on our background predictions.

VI. EVENT SELECTION USING KINEMATIC QUANTITIES

The variables used to discriminate hadronic top signal from QCD multi-jet background can be distinguished into five categories[4]. The selection is based on the quantities used in the DØ Run 1 analysis, and was not optimized for use in TeVatron Run 2:

(i) Energy scale. QCD background tends to have an overall lower transverse energy distribution, jets are less energetic and the total invariant mass of the event is smaller than in $t\bar{t}$ events. Even though the average jet energy is smaller in QCD events, the leading jets tend to be more energetic in QCD than in $t\bar{t}$ events. The variables used here are H_T, the sum of all jet p_T's and $\sqrt{s}$, the invariant mass of the event.

(ii) Soft non-leading jets. As the QCD background mainly consists of hard 2-jet processes with extra soft gluon jets, the additional jets are expected to be softer in QCD background than in $t\bar{t}$ signal. The variables we use are H_T^{3j}, the sum of all jet p_T's except the leading two, $E_{T_{4,5}}$, the geometric mean of the fifth and sixth jet, and N^A_{jets}, the p_T weighted jet multiplicity.

(iii) Event Shape. These quantities describe the behavior of the angles and sizes of jets in the event as a whole. Top events have a different shape than QCD background. The jets are almost spherically distributed in top events, while QCD events usually have a more back-to-back jet distribution. We use *aplanarity* and *sphericity* to quantify this difference between signal and background.

(iv) Rapidity distribution. These quantities are used to identify where the set of jets in the event was observed in the detector. Because of their typical hard scatter origin, the jets in QCD background events are expected to be more back-to-back than top signal, while QCD events are also more likely to be boosted in the direction of the beam-line. This has as a consequence that not all the jets in the event are expected to be central. We use *centrality*, the ratio of the total transverse energy over the total energy in the event, and $< \eta^2 >$, the p_T weighted jet variance from $\eta = 0$.

	H_T	$\sqrt{s}$	H_T^{3j}	N_{jets}^A	$E_{T_{5,6}}$	*Aplan*	*Spher*	*Cent*	$<\eta^2>$	$M_{t\bar{t}}$	M_W	M_{12}^{min}	M_{34}^{min}
H_T	1	0.63	0.74	0.84	0.49	-0.04	0.08	0.23	-0.04	0.10	0.11	0.34	0.42
$\sqrt{s}$		1	0.52	0.56	0.35	-0.40	-0.42	-0.45	0.33	0.19	0.27	0.31	0.40
H_T^{3j}			1	0.79	0.77	0.13	0.10	0.12	0.02	-0.00	0.03	0.37	0.41
N_{jets}^A				1	0.57	0.03	0.10	0.19	-0.02	-0.01	0.06	0.43	0.51
$E_{T_{5,6}}$					1	0.12	0.07	0.07	0.03	-0.02	0.03	0.42	0.43
Aplan						1	0.69	0.58	-0.27	-0.12	-0.15	0.03	0.01
Spher							1	0.74	-0.38	-0.12	-0.19	0.03	0.02
Cent								1	-0.37	-0.09	-0.16	-0.02	-0.05
$<\eta^2>$									1	0.08	0.14	0.01	0.02
$M_{t\bar{t}}$										1	0.20	-0.06	-0.07
M_W											1	0.00	0.01
M_{12}^{min}												1	0.71
M_{34}^{min}													1

TABLE II: *Average correlations among the kinematic parameters for data events*

(v) Typical top properties. The second neural network ($NN2$) is trained on properties which are very typical for top event structure, like the presence of W-bosons and b-quarks. The variables used are the W and top χ^2, $\mathcal{M}_{WW}$ and $\mathcal{M}_{t\bar{t}}$, and the minimal di-jet masses $M_{min}^{1,2}$ and $M_{min}^{3,4}$.

Figure 2 shows the distributions for background and signal Monte Carlo for all the variables mentioned above. The absolute predicted background is consistent with the observed tagged data.

The kinematic variables used in this analysis are correlated. To account for these correlations in order to achieve optimal discrimination between signal and background, we use artificial neural networks (ANN). The correlation matrix for data events is shown in table II.

VII. NEURAL NETWORK ANALYSIS

A. Training

We use feed-forward neural networks (NNs), trained by back propagation as implemented in the JETNET[5] program. All the NNs have 1 output node and 1 middle layer with a number of nodes double the number of the input layer. The NNs are trained on a small, randomly chosen fraction (2500 events, $\approx 1\%$) of our background sample and the same number of MC simulated $t\bar{t}$ events, also randomly chosen. Tagged data events are never used in the training of the NNs.

We use the *probability* that an event is tagged to select events for our random sample selection. Events with high tagging probability are more likely to be used for neural network training. This way our background training sample contains taggable events that are similar to the tagged events in our signal sample.

B. Neural network chain

The kinematic variables introduced in VI are combined in a chain of artificial neural networks:

- $NN0$. This neural network is used for rejection of obvious background. Events with a $NN0$ discriminant smaller than 0.05 are rejected. $NN0$ uses H_T, $\sqrt{s}$, H_T^{3j}, N_{jets}^A, *sphericity*, *aplanarity* and *centrality*. $NN0$ does not have a major influence on our signal, but does reject a significant fraction of our background.
- The next neural net, $NN1$, includes the same variables as $NN0$, with additional information from $E_{T_{4,5}}$ and $<\eta^2>$. As this neural net is trained only on events that passed the $NN0 > 0.05$ cut, it is a lot more sensitive to the difference between top-like background and hadronic $t\bar{t}$ signal. We do not cut on $NN1$, but use it as an input to the final neural net.

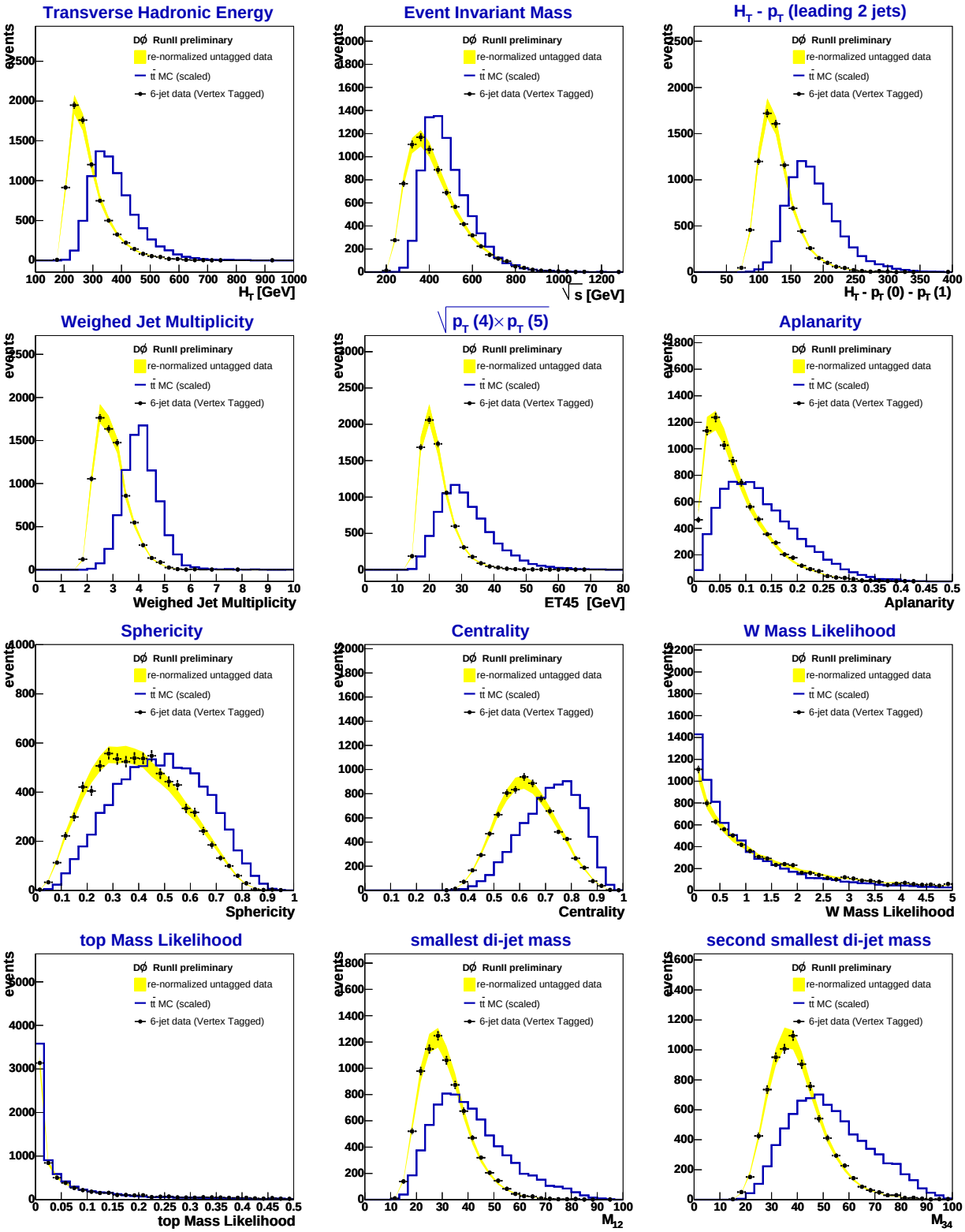

FIG. 2: *All kinematic variables used in this analysis, for tagged events (points) and the predicted background in those tagged events(band), and for informative purposes the distribution shape for hadronic $t\bar{t}$ signal Monte Carlo.*

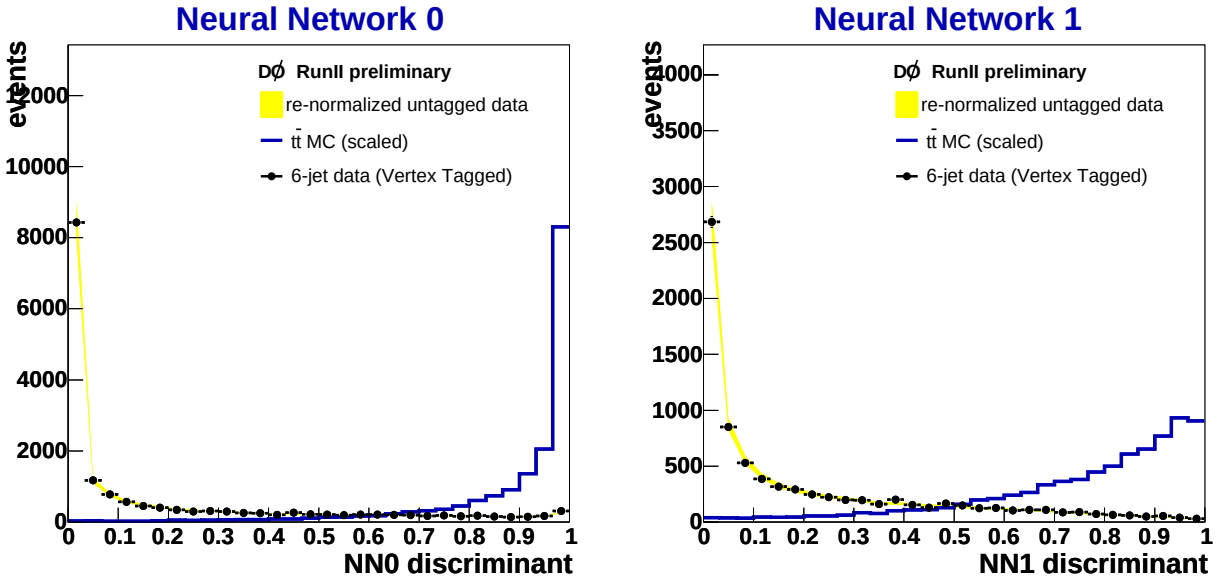

FIG. 3: *Predicted and observed numbers of tagged events for the lower level neural networks $NN10$ and $NN1$.*

efficiencies

	$t\bar{t} \to$ all − jets		$t\bar{t} \to$ lepton + jets	
	marginal	cumulative	marginal	cumulative
pre-selection	0.388 ± 0.003	0.388 ± 0.003	0.038 ± 0.001	0.038 ± 0.001
trigger	0.737 ± 0.008	0.286 ± 0.002	0.498 ± 0.019	0.018 ± 0.001
$NN0 > 0.05$	0.997 ± 0.011	0.285 ± 0.002	0.983 ± 0.042	0.018 ± 0.001
$N_{SVT} = 1$	0.456 ± 0.005	0.131 ± 0.001	0.481 ± 0.020	0.009 ± 0.000
$NN2 > 0.75$	0.436 ± 0.007	0.058 ± 0.001	0.220 ± 0.018	0.002 ± 0.000

TABLE III: *Efficiencies for all analysis cuts.*

- The final neural net, $NN2$, combines the information from $NN1$ with variables which are sensitive to high-mass objects in the event, or to the $t\bar{t}$ hypothesis. The input variables for $NN2$ are $NN1$, $\mathcal{M}_{WW}$, $\mathcal{M}_{t\bar{t}}$, $M^{1,2}_{min}$ and $M^{3,4}_{min}$.

Figure 3 shows the expected and observed distributions for the neural network discriminants $NN0$ and $NN1$.

C. NN2 cut

We have studied the optimal $NN2$ cut value using our (untagged) background events and the $t\bar{t}$ MC simulation. The statistical fractional error was determined to be minimal for a cut of $NN2 > 0.75$, which is what is therefor used in this analysis. Figure 4 shows the expected distributions for background, signal+background and the actual observed tagged data distribution.

The efficiencies for all analysis cuts are listed in table III. The analysis presented in this note is tuned to find hadronic $t\bar{t}$ events, but we find we are also marginally sensitive to events from the lepton+jets decay channels. The measured efficiencies are:

$$\epsilon_{\text{all-jets}} = 0.058 \pm 0.001(\text{stat})$$

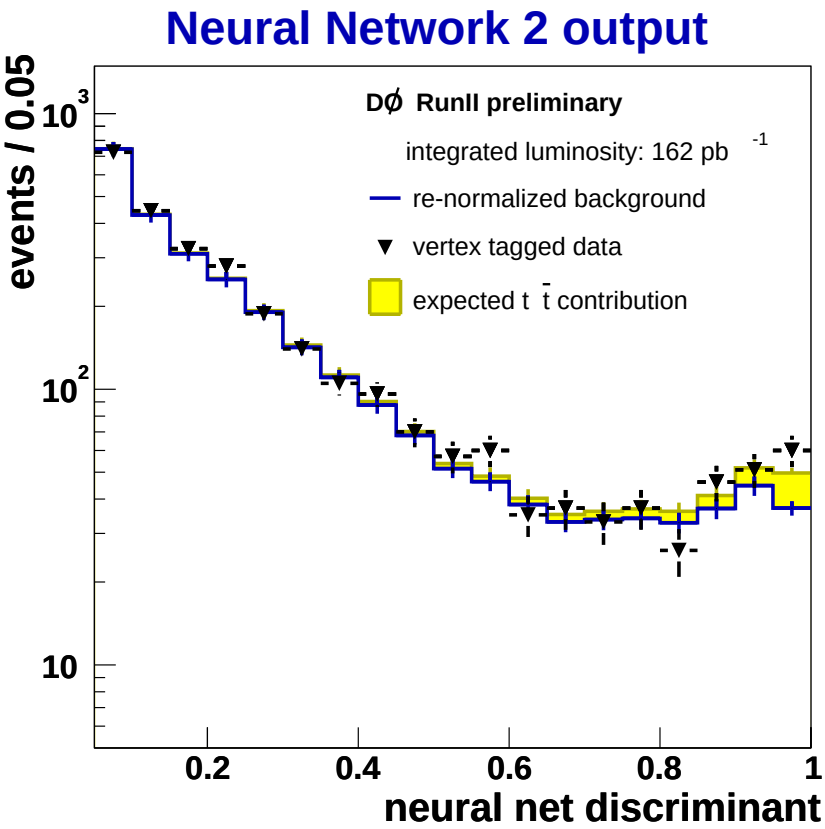

FIG. 4: *Neural Network 2 distribution for tagged data, expected background and expected signal+background.*

and

$$\epsilon_{\text{lepton+jets}} = 0.0018 \pm 0.0001(\text{stat}),$$

where the errors are due to limited Monte-Carlo statistics. Combining these two efficiencies according to the expected branching ratios (which are, respectively, 0.4619 for the all-jets and 0.4355 for all three lepton+jets channels combined) gives us an efficiency to measure the cross section for all $t\bar{t}$ production channels which produce an event topology of at least 6 jets and no isolated leptons. This gives us a final signal efficiency of

$$\epsilon \cdot \text{BR} = 0.0275 \pm 0.0003(\text{MC}, \text{stat}).$$

The main systematic uncertainties on $\epsilon \cdot \text{BR}$ are the used jet energy scale, and the top mass. Systematic uncertainties on the efficiency and background are are listed in tables IV and V.

We observe the following number of events above our cut of $NN2 > 0.75$:

$$N_{background}(\text{expected}) = \ 186 \ \pm \ 5(\text{stat}) \text{ events},$$

where the error comes from statistical fluctuations in our background distribution only. As the background events are determined using TRFs, this number includes a more than 186 events, hence the small statistical error. In the single-tagged sample, we observe the following:

$$N_{observed} = 220 \text{ events}$$

systematic uncertainties on signal efficiency (in %)		
Vertex reconstruction	-1.0	1.0
Jet Identification	-9.8	-
Jet Energy Scale $\pm\sigma$	-28.3	$+28.1$
Jet Resolution $\pm\sigma$	-0.6	$+0.2$
Top Mass $\pm 5\mathrm{GeV/c}^2$	-7.6	$+5.9$
Trigger Efficiency(trigsim)	-4.0	4.0
SVT parametrizations	-4.1	$+3.6$
total	-31.1	+ 31.2

TABLE IV: *Systematic uncertainties on the signal efficiency.*

systematic uncertainties on background estimation (in %)	
Statistical error TRFs	± 3.6
Background modeling (TRFs)	± 6.6
Total	± 7.5

TABLE V: *Systematic error on expected background calculation.*

The $t\bar{t}$ cross section is then measured to be:

$$\sigma(t\bar{t}) = 7.7^{+3.4}_{-3.3}(\mathrm{stat})^{+4.7}_{-3.8}(\mathrm{syst}) \pm 0.5(\mathrm{lumi})\mathrm{pb},$$

where we presume a 6.5% error on the integrated luminosity of our analysis sample. The likelihood curve as a function of the $t\bar{t}$ production cross section is shown in figure 5.

VIII. CONCLUSION

We have measured the top and anti-top quark ($t\bar{t}$) pair cross section in the top to 'all-jets' channel using an event sample corresponding to an integrated luminosity of 162 pb^{-1}, which was recorded by DØ in the period 2002/2003.

The analysis is restricted to events with at least six jets (and no isolated lepton); the expected signature for $t\bar{t}$ pairs decaying in a b quark and a W boson and the subsequent decay of each W boson into two quarks. Assuming a $t\bar{t}$ cross section of 7 pb, we expect 28 signal events.

The background from QCD events is reduced using several variables related to the topology and the underlying physical structure of the events. These variables are combined in a chain of two artificial neural networks, after which we apply a secondary vertex tag (SVT) to reduce background. The efficiency after a cut on the output on the second neural network is

$$\epsilon \cdot \mathrm{BR} = 0.0275 \pm 0.0003(\mathrm{stat})^{+0.0086}_{-0.0079}(\mathrm{syst}) \tag{1}$$

The remaining background is predicted using all events except events with a double tag, weighted with event weights predicted by an SVT tag rate function. We observe 220 events, where we predict $186 \pm 5(stat) \pm 14(syst)$ background events.

The measured cross section is

$$\sigma(t\bar{t}) = 7.7^{+3.4}_{-3.3}(\mathrm{stat})^{+4.7}_{-3.7}(\mathrm{syst}) \pm 0.5(\mathrm{lumi})\ \mathrm{pb}. \tag{2}$$

[1] ALPGEN, a generator for hard multiparton processes in hadronic collisions, M.L. Mangano et al., JHEP 0307:001, 2003, hep-ph/0206293.

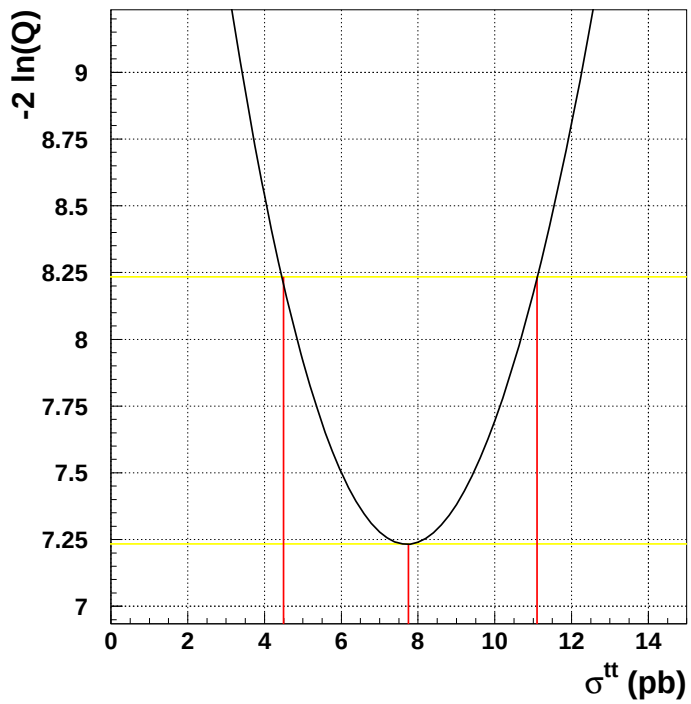

FIG. 5: *Likelihood as a function of the $t\bar{t}$ production cross section. The central value and the statistical errors are indicated as a vertical line.*

[2] New Generation of Parton Distributions with Uncertainties from Global QCD Analysis, CTEQ6, hep-ph/0201195.
[3] PYTHIA 6.2, Physics and Manual, T.Sjoestrand et al., LU TP 01-21, hep-ph/0108264.
[4] Measurement of the top quark pair production cross section in $p\bar{p}$ collisions using multijet final states,DØ Collaboration, Phys. Rev. D 60, 012001 (1999).
[5] JETNET 3.0: A Versatile artificial neural network package, C. Peterson, T. Rognvaldsson and L. Lonnblad, Comput. Phys. Commun. **81**, 185 (1994).