

IL NUOVO CIMENTO

Vol. ?, N. ?

?

CDF/DOC/TOP/PUBLIC/9371  
August 13, 2008; v3.0

## Tools for Top Physics at CDF

E. PALENCIA<sup>(1)</sup> <sup>(2)</sup>

<sup>(1)</sup> *Fermi National Accelerator Laboratory - Batavia, IL, USA*

<sup>(2)</sup> *On behalf of the CDF Collaboration*

**Summary.** — We describe here the different tools used for top physics analysis in the CDF Collaboration. In particular, we discuss how the jet energy scale, lepton identification, b tagging algorithms and the neural networks help to improve the signal to background ratio of the top sample in some cases and to reduce the dominant uncertainties in other. Results using each one of these tools are also presented.

PACS 14.65.Ha, 13.85.Ni, 13.85.Qk – Conference Proceedings.

### 1. – Jet Energy Scale (JES)

The jet energy scale is the largest source of systematic uncertainty in the top mass analysis due to the differences between data and Monte Carlo. Therefore, a good understanding of the jet energy corections is desirable.

Due to different processes, the properties of the jet measured at calorimeter level are different to those at production level. First of all, we have very complex detector properties (mainly due to instrumental effects) such as non linear response to hadrons, different response to electrons and muons and un-instrumented regions. We also have jet reconstruction algorithms (cone, cone-midpoint,...) with very complex behavior. And, finally, there are very complex underlying physics processes such as fragmentation, hadronization and initial and final state radiations.

So, in order to know the true properties of the jet, we need to correct the measured jet for these effects. These corrections are done in different levels at CDF. All details about the jet energy corrections used can be found in [1].

**1.1. *JES in situ Calibration.*** – In order to reduce the uncertainty due to JES, we can use (when possible) the hadronic decay of the  $W$  boson to measure, in situ, the JES. In other words, we can constraint the invariant mass of the non- $b$  tagged jets to be the mass of the  $W$  boson.

The method starts building templates using the invariant mass of all the non-tagged jet pairs in the event. And then, instead of assuming a given value for JES and measure

the mass of the  $W$ , we assume the mass of the  $W$  and measure JES. Figure 1 shows the parameterization of the  $W$  mass for different values of JES.

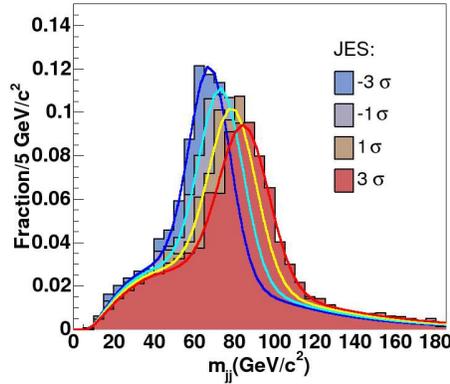


Fig. 1. – . Templates of the mass of the  $W$  boson for different values of JES.

This method gives around 40% improvement in the uncertainty of the top mass measurements. Another advantage is that, since the JES uncertainty scales directly with statistics, the dominant systematic uncertainty scales as  $1/\sqrt{N}$ .

Recent measurements in the lepton+jets and all hadronic channels have been performed using the JES in situ calibrations. The top mass measured are  $171.4 \pm 1.5$  (stat+JES)  $\pm 1.0$  (syst) and  $177.0 \pm 3.7$  (stat+JES)  $\pm 1.6$ (syst) GeV, respectively. Figure 2 shows the final simultaneous fit of the top mass and JES in the lepton+jets (left) and all hadronic (right) channel.

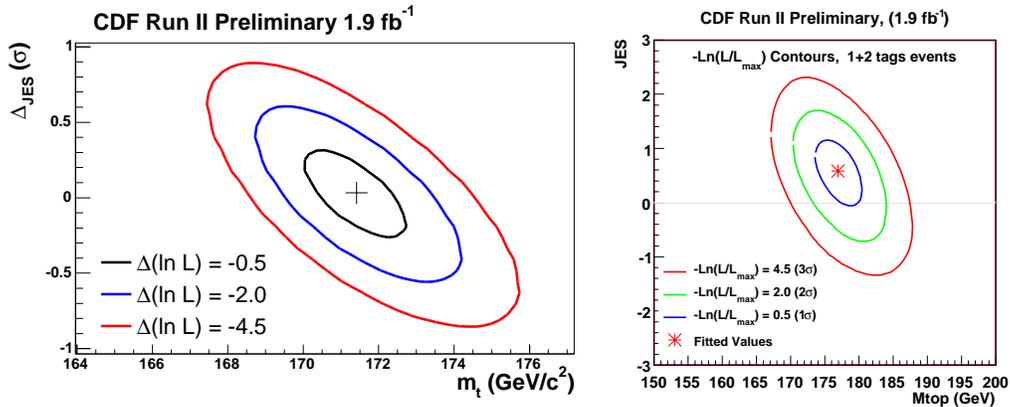


Fig. 2. – Simultaneous fit of the top mass and JES in the lepton+jets (left) and all hadronic (right) channel.

## 2. – Lepton Identification

The presence in the final state of a lepton from the leptonic decay of the  $W$  boson helps tremendously in the triggering of the top sample. Leptons are relatively easy to identify so requiring a lepton at trigger level helps to reduce the amount of background.

**2.1. Standard Leptons.** – Most of the top analysis at CDF to date use as standard leptons those electrons or muons from a high transverse momentum lepton trigger.

In order to select electrons, we basically require the presence of a charged particle track in the tracking system that leaves almost all of its energy in the electromagnetic calorimeter. Since we do not want electrons coming from heavy flavor jets, we also require no other nearby tracks.

And to select muons, we also require the presence of a charged particle track in the tracking system and no other nearby tracks. But in this case, there is very little energy deposited in the calorimeter and muons above a certain energy are naturally detected in the muon chambers (all other particles except neutrinos fail to make it to the muon system). Then, we require the presence of stubs in the muon chambers. The occupancy in the  $\phi$ - $\eta$  plane for these muons coming from a high transverse momentum trigger is shown in Fig. 3 (left).

**2.2. Extended Muon Coverage.** – It is clear that, due to the geometry of the CDF muon chambers, there are many gaps. In other words, we are missing many events due to detector inefficiency. But, since muons in these kind of events are undetected, they contribute to the missing transverse energy in the event, MET. So, in order to recover these events, and increase the lepton acceptance, we use complementary triggers (MET + 2 jets). The occupancy in the  $\phi$ - $\eta$  plane adding the muons coming from this trigger is shown in Fig. 3 (right).

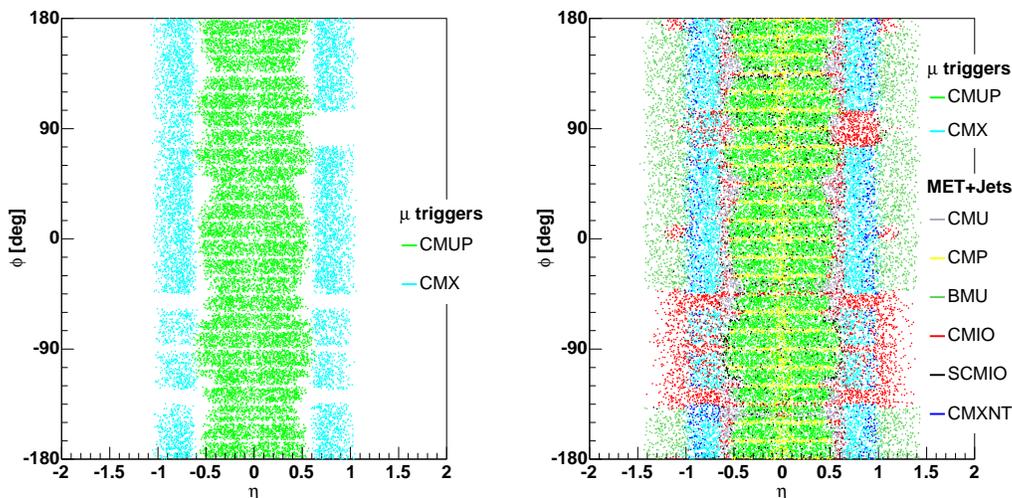


Fig. 3. – Occupancy in the  $\phi$ - $\eta$  plane for muons coming from a high transverse momentum trigger (left) and adding muons coming from the MET + jets trigger (right).

As expected, the muons coming from the MET + jets trigger fill the gaps left by the muons coming from the high transverse momentum trigger, giving around 30% gain in

muon acceptance. These muons have been used in the single top search at CDF giving an improvement in the sensitivity of around 7%.

### 3. – $B$ Tagging

Since, within the standard model (SM), the top quark decays most of the time to a  $W$  boson and a  $b$  quark, the final state is characterized by the presence of, at least, a  $b$  quark jet. On the contrary, most of the backgrounds contributing in the top sample do not have bottom quark jets in the final state. Therefore, the identification of the jets coming from  $b$  quarks is a great tool to increase the signal over background ratio in the top sample. Another advantage of  $b$  tagging is that it helps reducing the combinatorics in some top mass analysis.

Any time a  $b$  tagging algorithm is used, we need to know how many times the algorithm is able to identify a real  $b$  jet ( $b$  tag efficiency) and how often the algorithm misidentifies a light flavor jet as a heavy flavor one (mistag rate).

**3.1. Efficiency Measurement.** – Understanding the tagging efficiency in Monte Carlo is easy since, as we know which jets come from  $b$  quarks, we only need to count how many of those are tagged. What one really needs is the efficiency of tagging  $b$  jets in data. This measurement is done using a 8 GeV lepton data sample that is rich in real heavy flavor. We require the presence of a jet with a lepton inside (indicative of semileptonic  $b$  decay). Furthermore, in order to enhance the  $b$  fraction of the sample, we also require the presence of another jet (away jet) identified as  $b$  jet.

The efficiency measurement uses the high relative transverse momentum ( $p_T^{rel}$ ) of the muons from  $b$  decays. We divide the muon jets sample into two subsamples: tagged and untagged. Then, we generate  $p_T^{rel}$  templates for  $b$  (using MC) and non- $b$  (using MC and data) jets. We have four non- $b$  templates, two of them use data and the other two Monte Carlo, and are shown in Fig. 4 (left). The  $b$  templates for the tagged and untagged samples are shown in Fig. 4 (right).

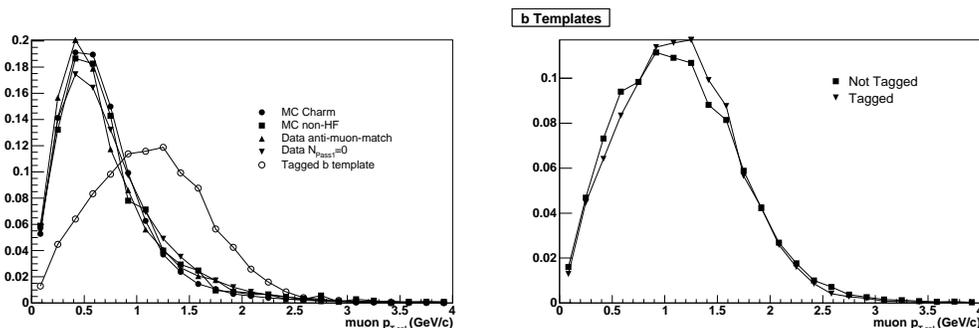


Fig. 4. – Left: the four non- $b$  templates (full markers) and the tagged  $b$  template (empty circles). Right: tagged (triangles) and untagged (squares)  $b$  templates.

As shown, the  $b$  and non- $b$  templates have different distribution as a function of  $p_T^{rel}$ . Also, as expected, the four non- $b$  templates are very similar, and the  $b$  templates are similar for tagged and untagged  $b$ -jets.

Finally, we fit the  $p_T^{rel}$  spectrum from data to find the fraction of  $b$  and non- $b$  jets in the tagged and untagged samples. Once we know the fractions, the efficiency is directly

measured as

$$\epsilon_{data} \equiv \frac{b's \text{ in tagged sample}}{b's \text{ in tagged sample} + b's \text{ in untagged sample}} = \frac{F_b^{tag} \times N_{jets}^{tag}}{F_b^{tag} \times N_{jets}^{tag} + F_b^{untag} \times N_{jets}^{untag}}.$$

**3.2. Mistag Rate Measurement.** – The mistag rate is calculated using inclusive jet data samples. We parameterized it as a six dimensional look-up table (so called mistag matrix) of the following variables: transverse energy of the jet, number of tracks in the jet, sum of the transverse energy of the jets in the event, pseudorapidity of the jet,  $z$  vertex position of the event and number of  $z$  vertices in the event. For each one of the bins in the table, we count the number of taggable jets and the number of negatively tag jets. Finally, the mistag rate is calculated as the ratio of these two numbers.

#### 4. – $B$ Tagging Algorithms

The  $B$  hadrons, due to their long lifetimes, can be exploited in order to identify jets consistent with originating from  $b$  quark production. Given their long lifetimes, and the large boost,  $B$  hadrons, product of the hadronization of the  $b$  quark, travel a macroscopic distance away from the primary interaction point before decaying into charge and neutral particles. Then, reconstruction of charge particle tracks allows us to look for trajectories inconsistent with originating from the initial interaction point. At CDF, there are two algorithms that make use of this feature and these are described in Sect. 4.1 and 4.2.

Other kind of taggers do not rely on the long lifetime of the  $B$  hadrons but in the semileptonic decay of the  $B$  hadrons:  $b \rightarrow l\nu c$ . These taggers basically look for an energetic lepton (electron or muon) inside a jet and are described in Sect. 4.3 and 4.4.

**4.1. Secondary Vertex.** – This algorithm is applied to each jet in the event, using only the tracks which are within an  $\eta$ - $\phi$  distance of  $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$  of the jet direction. Displaced tracks in jets are used for the reconstruction and are distinguished by a large impact parameter significance ( $S_{d_0} \equiv |d_0/\sigma_{d_0}|$ ) where  $d_0$  and  $\sigma_{d_0}$  are the impact parameter and the total uncertainty from tracking and beam position measurements. Secondary vertices are reconstructed with a two-pass approach which tests for high-quality vertices in the first pass and allows lower-quality vertices in the second pass. In pass 1, at least three tracks are required to pass loose selection criteria ( $p_T > 0.5$  GeV,  $|d_0/\sigma_{d_0}| > 2.0$ ), and a secondary vertex is fit from the selected tracks. One of the tracks used in the reconstruction is required to have  $p_T > 1.0$  GeV. If pass 1 fails, then a vertex is sought in pass 2 from at least two tracks satisfying a tight selection criteria ( $p_T > 1.0$  GeV,  $|d_0/\sigma_{d_0}| > 3.5$  and one of the pass 2 tracks must have  $p_T > 1.5$  GeV). If either pass is successful, the transverse distance ( $L_{xy}$ ) from the primary vertex of the event is calculated along with the associated uncertainty. This uncertainty  $\sigma_{L_{xy}}$  includes the uncertainty on the primary vertex position. Finally jets are tagged positively or negatively depending on the  $L_{xy}$  significance ( $L_{xy}/\sigma_{L_{xy}}$ ). More details about the Secondary Vertex tagger can be found in [2].

The Secondary Vertex tagging algorithm is, by far, the most common in the top physics analysis at CDF. Most of the cross section, top mass, top properties, single top analysis that apply  $b$  tagging use this algorithm.

**4.2. Jet Probability.** – This algorithm uses tracks associated with a jet to determine the probability for these to come from the primary vertex of the interaction. The calculation of the probability is based on the impact parameter of the tracks in the jet and their

uncertainties. The  $S_{d_0}$  distribution, see Fig. 5 (left), peaks at zero and falls quickly with increasing absolute value of  $|S_{d_0}|$ , but the tails are rather long. Using this distribution, one can build the probability that the impact parameter significance of a given track is due to detector resolution. And, finally, using the probabilities off all the tracks in the jet, one can build the probability that the jet is consistent with a zero lifetime hypothesis. The Jet Probability distribution for different Monte Carlo samples is shown in Fig. 5 (right).

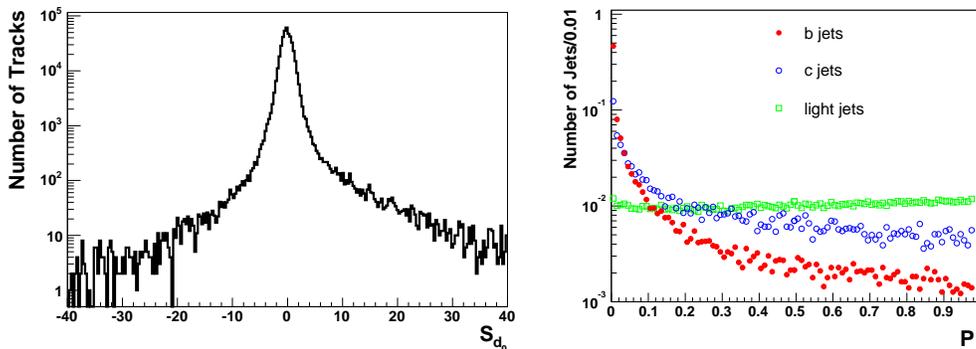


Fig. 5. – Left: Distribution of the impact parameter significance for tracks in an inclusive jet sample with at least 5 good SVX hits,  $p_T > 5$  GeV, and  $|\eta| < 0.6$ . Right: Jet probability distributions for jets matched to  $b$  (full circles),  $c$  (empty circles) and light (empty squares) quarks in Monte Carlo simulated events.

As shown, the probability for tracks originating from the primary vertex is, by construction, uniformly distributed from 0 to 1. For a jet coming from heavy flavor hadronization, the distribution peaks at 0, due to tracks from long lived particles that have a large impact parameter with respect to the primary vertex. A feature of this tagger is that the  $b$  tagging is performed using a continuous variable rather than a discrete object. More details about Jet Probability can be found in [3].

This tagging algorithm has been used, for example, to measure the top-antitop production cross section and in the s-channel single top search.

**4.3. Soft Muon Tagger.** – This tagger tries to identify muons inside a jet. This identification is made based on a global  $\chi^2$ ,  $L$ , that combines information from muon matching variables (there is no calorimetry or isolation requirements). It considers tracks with  $\Delta R_{slt-jet} < 0.6$ ,  $|\Delta Z_{slt-Zvtx}| < 5$  and  $p_T > 3$  GeV and defines a jet tagged if there is a  $SLT\mu$  with  $L < 3.5$ . More details can be found in [4].

This tagging algorithm has recently been used to measure the top-antitop production cross section with  $2 \text{ fb}^{-1}$  of data and the resulting measured value is  $8.7 \pm 1.1$  (stat)  $^{+0.9}_{-0.8}$  (syst)  $\pm 0.6$  (lumi) pb. The number of expected  $SLT\mu$  tagged events as a function of the number of jets in the event is shown in Fig. 6 (left).

**4.4. Soft Electron Tagger.** – If, instead of muons, we consider electrons, we have a much more busy environment and, therefore, a more complicated tagger. This algorithm begins by extrapolating well measured tracks ( $p_T > 2$  GeV) to the calorimeter. Before consideration in a likelihood, electrons must pass some basic identification cuts. It uses

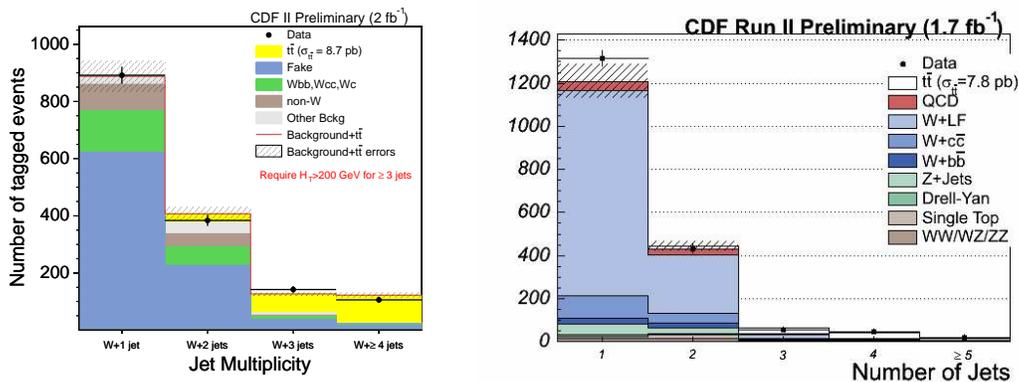


Fig. 6. – Left (right): Number of expected signal and background events as a function of the jet multiplicity when applying the soft muon (electron) tagger.

data to parameterize the tagging rate for electrons and fake electrons in MC (as a function of  $p_T$ ,  $\eta$  and isolation): conversion electrons for the electron tagging rate and tracks in generic jets for the fake electron tagging rate. Finally, it rejects electrons from conversions by looking for the conversion partner or by finding missing silicon layers.

This tagging algorithm has been used for the first time in the Run II at CDF to measure the top-antitop production cross section with  $1.7 \text{ fb}^{-1}$  of data and the resulting measured value is  $7.8 \pm 2.4$  (stat)  $\pm 1.5$  (syst)  $\pm 0.5$  (lumi) pb. The number of expected SLTe tagged events as a function of the jet multiplicity is shown in Fig. 6 (right).

## 5. – Neural Networks

As in many other analysis nowadays, neural networks (NN) are also widely used in top quark measurements in CDF.

One NN used is based in the discrimination available from kinematical and topological variables to distinguish top pairs from background events. Since top pair events are central, spherical and energetic, this NN uses seven input variables related with this properties ( $H_T$ , event aplanarity,  $\eta_{max}$ ,  $\min(M_{jj})$ ,  $E_T(j1) + E_T(j2) + E_T(j3)$ , ...) to distinguish top from background events in a not tagged sample, where  $S/B \sim 1:5$  ( $1:1.5$ ) in  $W + \geq 3$  ( $4$ ) pretag jets samples. The output of this NN for signal and background events is shown in Fig. 7 (left). It is clear that signal events have higher values of the NN output than background events. This discrimination is used to measure the  $t\bar{t}$  production cross section and the resulting measured value is  $6.0 \pm 0.6$  (stat)  $\pm 0.9$  (syst) pb.

There is another analysis that applies a NN over the sample tagged by the Secondary Vertex tagger. This sample, even if is rich in  $b$  jets, has some contamination of bottom and light jets. Two NN are trained with  $t\bar{t}$  Monte Carlo sample: one to distinguish bottom from light jets and another one to distinguish bottom from charm. These NNs use 16 input variables (vertex mass, decay length, track multiplicity...). Events with NN outputs higher than given values are selected such as 90% of the Secondary Vertex tagged  $b$  jets are kept and 65% of light jets and 51% of charm jets are rejected. The  $t\bar{t}$  production cross section measured in this case is  $8.2 \pm 0.6$  (stat)  $\pm 1.0$  (syst) pb.

Finally, NNs are also used in the single top search. Since about 50% of the background

in the W+2 jets sample do not contain  $b$  quarks (even though a secondary vertex is required), a NN to separate jets from  $b$  quark from  $c$  and light quark jets is used. This NN uses jet and track variables (vertex mass, decay length, track multiplicity...) as input. The output of this NN for different samples is shown in Fig. 7 (right). The use of this NN provides a significant improvement to the single top search sensitivity of about 10-20%. Fig. 7 (right).

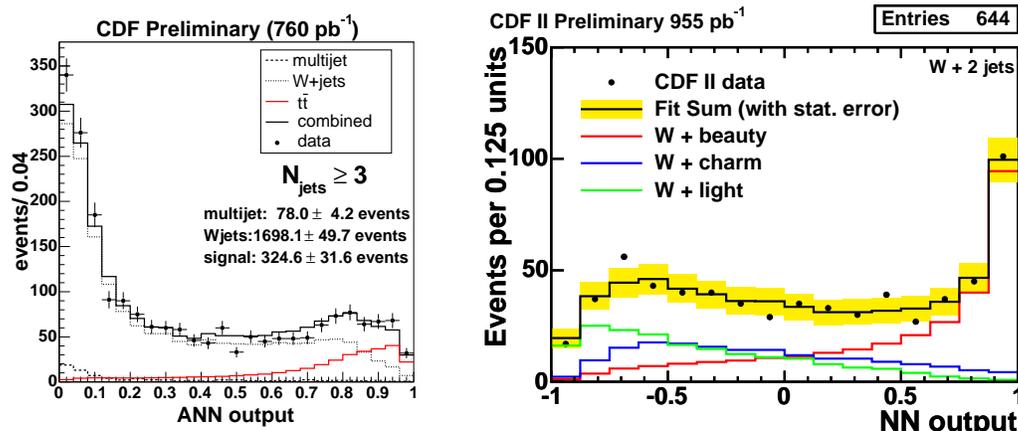


Fig. 7. – Neural network outputs for different samples used in the  $t\bar{t}$  cross section analysis (left) and in the single top search (right).

\* \* \*

The results shown here represent the work of many people. We thank my CDF colleagues for their efforts to carry out these challenging physics analyses. We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Educación y Ciencia and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland. We also thank the conference organizers for a very nice week of physics.

## REFERENCES

- [1] BHATTI A. *et al.*, *Nucl. Instr. Meth. A*, **566** (2006) 375-412.
- [2] ACOSTA D. *et al.* (CDF COLLABORATION), *Phys. Rev. D*, **71** (2005) 052003.
- [3] ABULENCIA A. *et al.* (CDF COLLABORATION), *Phys. Rev. D*, **74** (2006) 072006.
- [4] ACOSTA D. *et al.* (CDF COLLABORATION), *Phys. Rev. D*, **72** (2005) 032002.