CDF note 9841

$t\bar{t}$ production cross section measurement in the all-hadronic channel ($\mathbf{2.9}\, fb^{-1}$)

The CDF Collaboration

URL http://www-cdf.fnal.gov

(Dated: August 6, 2009)

We present here the measurement of the $t\bar{t}$ production cross section obtained from the results of the Template Method analysis (TMT2D) used to measure, in the all-hadronic channel, the top quark mass with simultaneous *in situ* measurement of the Jet Energy Scale (JES). The measurement discussed here is performed using about 2.9 fb^{-1} of $p\bar{p}$ collisions collected with a multijet trigger and selected using a dedicated neural network. We run pseudo-experiments on simulated samples to evaluate the need for calibration and to study the main systematic effects and the respective relative uncertainties on the measurement.

The measured $t\bar{t}$ production cross section amounts to $7.2 \pm 0.5\,(stat) \pm 1.0\,(syst) \pm 0.4\,(lumi)$ pb for a top mass of $174.8^{+2.7}_{-2.8}$ GeV/c^2 and ΔJES $= -0.3 \pm 0.6$ (i.e. the all-hadronic value) and $7.2 \pm 0.5\,(stat) \pm 1.1\,(syst) \pm 0.4\,(lumi)$ pb for a top mass of 172.5 GeV/c^2 and ΔJES $= 0$ (for the CDF average).

Preliminary Results for Summer 2009 Conferences

I. INTRODUCTION

The simultaneous measurement of the top quark mass, M_{top}, and of the jet energy scale, JES, by the template method (TMT2D) [1] consisted in finding the values of M_{top}, JES and the number of signal (n_s) and background (n_b) events, for events with 1 and ≥ 2 tagged jets, which best reproduced the observed mass distributions of top candidates, m_t^{rec}, and of W candidates, m_W^{rec}, as reconstructed in the selected data samples, given the p.d.f. expected for signal and background. This was done by performing a fit where a mass likelihood function was maximized, returning not only the best values of M_{top} and of the scale shift w.r.t. the default, ΔJES (expressed in units of the uncertainty on the energy correction, σ_c), but also the signal yields (n_s^{1tag}, $n_s^{\geq 2tags}$) in the two tagging classes. These yields can be used here to measure of the cross section, using the efficiencies for selecting signal events in the two classes.

Given the signal yields coming from the mass fit, the value of the cross section depends on the mass at which we evaluate the efficiency, so we need to consider two cases:

1. an all-hadronic stand-alone measurement, i.e. referring to the top quark mass and ΔJES as measured by TMT2D: $M_{top} \approx 175$ GeV/c^2 and $\Delta\text{JES} = -0.3$ (AH-TMT), corresponding to our measurement $M_{top} = 174.8^{+2.7}_{-2.6}$ GeV/c^2 and $\Delta\text{JES} = -0.3^{+0.57}_{-0.59}$ in this same all-hadronic channel;

2. a measurement for the inclusion in a CDF cross section average, i.e. referring to the fixed value of 172.5 GeV/c^2 and $\Delta\text{JES} = 0$ (CDF-COMB). For reference we will provide also measurements for 170 and 175 GeV/c^2 ($\Delta\text{JES} = 0$).

Before applying the method to real data we run simulated experiments (*pseudo-experiments*) to check for possible biases and to extract a calibration factor.

II. SUMMARY OF THE MASS RESULTS

We summarize in Table I all the relevant numbers from [1] relative to an integrated luminosity $L = 2.9$ fb^{-1}. The uncertainties quoted on the signal yields contain both the statistical uncertainty and the uncertainty due to the simultaneous fit of M_{top}, ΔJES and the background.

CDF preliminary, L = 2.9 fb^{-1}			
	Symbol	1 tag	≥ 2 tags
Top mass (GeV/c^2)	M_{top}	$174.8 \pm 2.4^{+1.2}_{-1.0}$	
JES displacement	ΔJES	$-0.3 \pm 0.47^{+0.34}_{-0.37}$	
Observed candidates	n_{obs}	3452	441
Expected background	$n_{(b,exp)}$	2785 ± 83	201 ± 29
Fitted background	$n_{(b,fit)}$	2802 ± 70	220 ± 21
Signal from fit	$n_{(s,fit)}$	643 ± 80	216 ± 25

TABLE I: Top quark mass, displacement of JES from the default, observed candidates, signal and background yields.

III. CROSS SECTION LIKELIHOOD

The measurement of the $t\bar{t}$ production cross section is done by performing a fit where a likelihood function is maximized. This function is divided into 2 parts corresponding to the two tagging classes:

$$\mathcal{L} = \mathcal{L}_{1\,tag} \times \mathcal{L}_{\geq 2\,tags} \times \mathcal{L}_L$$

The $\mathcal{L}_{1,\geq 2\,tags}$ terms further consist of other factors:

$$\mathcal{L}_{1,\geq 2\,tags} = \mathcal{L}_{n_s} \times \mathcal{L}_{\epsilon}$$

which correspond, respectively, to the signal yield, n_s, the signal efficiency, ϵ, and the integrated luminosity, L.

The likelihood terms for the integrated luminosity, $\mathcal{L}_L$, and for the efficiencies, $\mathcal{L}_\epsilon$, assume the usual Gaussian form:

$$\mathcal{L} = e^{-(x-\overline{x})^2/2\sigma_x^2}$$

where x represents the variable, $\overline{x}$ its expectation, and σ_x the uncertainty with which the variable is known.

The term corresponding to the signal yield, $\mathcal{L}_{n_s}$, is also Gaussian, but can be written in terms of the cross section as:

$$\mathcal{L}_{n_s} \Rightarrow \mathcal{L}(\sigma_{t\bar{t}}) = e^{-(\sigma_{t\bar{t}}\cdot\epsilon\cdot L - n_s)^2/2\sigma^2_{n_s}}$$

with $\sigma_{t\bar{t}}$ being a free parameter together with ϵ and L (the index referring to the tagging classes has been omitted).

In order to facilitate the computation, we minimize the negative logarithm of the likelihood using MINUIT, instead of maximizing the likelihood itself. The uncertainties on the parameters are given by MINOS, taking positive and negative statistical error as the difference between the observable (O) central value and the values O^+ and O^- corresponding to $-lnL(O^{\pm}) + lnL(O) = -1/2$. We then take as unique, symmetric uncertainty, the average between O^+ and O^- for each parameter.

IV. CALIBRATION OF THE METHOD

We want to investigate the possible presence of biases in the cross section measurement which can be introduced by our method. To do so, we use the realistic pseudo-experiments (PEs) produced for the mass measurement, where "pseudo-data" are extracted from simulated signal and data-driven background templates corresponding to known values of M_{top} and ΔJES (M^{in}_{top}, $\Delta\mathrm{JES}^{in}$). Only a small set of new PEs have been produced specifically for this cross section measurement. A full description on how these pseudoexperiments have been prepared can be found in [1]. The corresponding signal yields are then used as inputs to the cross section likelihood fit, which returns a value for the cross section.

Two sets of 3000 PEs have been performed for the two cases define above, i.e. for the two points $(M^{in}_{top}, \Delta\mathrm{JES}^{in}) = (175, -0.3)$ and $(172.5, 0)$. From the signal yields obtained from the mass likelihood fit, see Table I, and considering the signal efficiencies for the two cases we obtain a nominal cross section value, $\sigma^{in}_{t\bar{t}}$, for the two points $(175, -0.3)$ and $(172.5, 0)$, which will be used as input for the two sets of PEs.

From the application of the mass likelihood fit to the PEs we obtain signal yields from which we derive fitted cross section values for the two sets of PEs, $\sigma^{fit}_{t\bar{t}}$. We then consider the ratio $k_\sigma = \sigma^{in}_{t\bar{t}}/\sigma^{fit}_{t\bar{t}}$ which represents the calibration constant we need to multiply our fit result by, in order to obtain an unbiased cross section measurement. This ratio turns out to be $k_\sigma = 0.982 \pm 0.008$ and will be the factor we multiply the fitted cross sections by in order to obtain an unbiased measurement.

V. SYSTEMATIC UNCERTAINTIES ON THE CROSS SECTION MEASUREMENT

Various sources of systematic uncertainties might affect the cross section measurement, and the procedure to evaluate the corresponding uncertainties is the same used in [1]. In this case, however, we do not consider the effect on M_{top} or JES, but the effect on the signal yields n_s^{1tag} and $n_s^{\geq 2tags}$. We need also to consider how the efficiencies ϵ_{1tag} and $\epsilon_{\geq 2tags}$ are affected, and finally we derive the effect on the cross section itself.

To apply such a procedure, PEs are performed extracting pseudo-data from mass templates built using signal/background samples where the possible systematic variations have been considered and included.

The results from these PEs are then usually compared to the ones obtained by using default mass templates considering also the change in efficiencies, and the shifts in the average cross section values are considered to calculate the systematic uncertainty, following the prescriptions described in [1].

The AH-TMT and CDF-COMB measurements have most of the systematic uncertainties in common, with no strong dependence on the top quark mass or JES. In some cases, however, the systematic uncertainty have been evaluated separately.

In the following we summarize the major contributions to the systematic uncertainty.

1. **Cross section calibration.** The uncertainty on the calibration factor, k_σ, introduces a small uncertainty on the calibrated cross section. The corresponding relative uncertainty is $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (σ calib.) $= \frac{\Delta k_\sigma}{k_\sigma} = 0.8\%$.

2. **Generator.** Many sources of systematic effects arise from uncertainty in the Monte Carlo modeling of the hard interaction. PYTHIA and HERWIG generators differ in their hadronization schemes and in their description of the underlying event and multiple interactions. Comparison of results obtained running PEs where Monte Carlo samples generated using the two different hadronization models leads to $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (Generator) $= 4.2$ %

3. **BR($W \rightarrow hadrons$).** The PDG uncertainty on the branching ratio of the W boson intro hadrons introduces a 0.8% relative uncertainty on the signal efficiency and hence on the cross section.

4. **Initial and final state radiation (ISR / FSR).**

   Additional jets coming from possible emission of hard gluons might fall among the six leading jets and populate the tails in the top quark invariant mass distribution. The amount of radiation from partons in the initial (ISR) or final (FSR) state is set by parameters of the PYTHIA generator used to simulate signal events. To study these effects, mass templates are built using samples where the values of the parameters have been changed with respect to the default, to increase or to decrease the amount of radiation. The current prescription for the evaluation of this systematic uncertainty is to use "IFSR" samples where ISR and FSR are increased or decreased simultaneously. These samples are used also to evaluate the efficiencies ϵ_{1tag} and $\epsilon_{\geq 2tags}$.

   As a relative systematic uncertainty we consider $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (IFSR) = 0.6 %.

5. ***b*-jets energy scale.**

   Since the jet energy corrections are derived on data samples deprived of heavy flavors, an additional uncertainty comes from considering the different properties of b quarks. Three different sources of systematic uncertainties have been considered, that is: the uncertainties on branching ratios of semi-leptonic decays of b and c-quarks; the uncertainties on b-quarks fragmentation parameters; the uncertainty on the calorimeter response to b and c hadrons.

   In this case, the quadrature sum of the three relative uncertainties gives: $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ ($b-$JES) = 2.8 %.

6. ***b*-tagging efficiency.**

   The different efficiency of the b-tagging algorithm on real data and Monte Carlo simulated events is usually considered as a constant scale factor (b-tag SF). However, it might have a dependence on the transverse energy of jets, leading to possible variations in the shapes of m_t^{rec} and m_W^{rec} templates. The possible statistical positive and negative variations of the slope of the SF as a function of the jet E_T have been considered along with the absolute uncertainty on the SF itself.

   Summing the two contributions in quadrature we obtain: $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ ($b-$tag SF) = 5.4 %.

7. **Parton distribution functions.** The choice of parton distribution functions (PDF) inside the proton can affect the kinematics of $t\bar{t}$ events and thus the cross section measurement. We estimate the uncertainty resulting from the possible PDF models by using the default Monte Carlo sample, and weigh the events by their probability to occur according to many different PDF's (CTEQ5L, MRST72, MRST75, CTEQ6M [2, 3]). We considered also the effect derived from the experimental uncertainties on the CTEQ6M PDF's, encoded by 20 pairs of values. Considering the quadrature sum of all these contributions we obtain a relative uncertainty of $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (PDF) = 2.4 %.

8. **Pileup.** The effects of possible discrepancies between real data and Monte Carlo due to pileup of events and related to changes in the istantaneous luminosity have been studied by modifying the mass templates and the signal efficiencies, taking into account the distribution of events as a function of the number of reconstructed primary interaction vertices as observed in the data. We thus estimate a relative uncertainty $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (Pileup) = 2.5 %.

9. **Color Reconnections.** Uncertainties from modeling of color reconnections effects are estimated by comparing the results of PEs performed using Monte Carlo samples generated including an explicit color reconnections model [4–6]. From the comparison with the default generation we derive $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (Color Reconn.) = 0.8%.

10. **Templates statistics.** The shapes of signal and background templates are affected by uncertainties due to the limited statistics of the Monte Carlo (for the signal) and data (for the background) samples used to build them. These uncertainties might affect the measurement of the signal yields which is performed by an unbinned likelihood where parametrized p.d.f.'s, fitted to default histograms shapes, are evaluated. Performing different PEs with fluctuated templates, while using the same value for the efficiencies, we estimate $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (Templ. Stat.) = 0.8%.

11. **Background shape.** Beside the purely statistical effects quoted above, the shape of background mass templates has also uncertainties due to the systematic error on the background normalization, which translates into a relative uncertainty $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (Bkg Shape) = 0.3%.

12. **Background normalization.** The expected numbers of events due to the background in our data sample are evaluated *a priori* by the method explained in [1], with uncertainties estimated by considering the discrepancies with the observed number of events in the signal region. The effects of the uncertainties on the background normalization correspond to $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (Bkg Norm) = 8.2%.

13. **JES dependence.** The signal efficiency depend strongly on the value used for the JES, and the uncertainty on it becomes a systematic uncertainty on the cross section measurement.

    In the AH-TMT case we measured the ΔJES to be $-0.3^{+0.58}_{-0.60}$, so the systematic uncertainty on the cross section measurement refers to a $-0.9\sigma_c$, $+0.3\sigma_c$ variation in ΔJES. For the CDF-COMB measurement we need instead to move the JES by $\pm 1\sigma_c$. The systematic effect on the cross section amounts to $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}\,(\mathrm{JES,\ AH-TMT}) = 6.1\%$ in the first case and $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}\,(\mathrm{JES,\ CDF-COMB}) = 9.2\%$ in the second case.

14. **Residual JES dependence.** The uncertainty on the jet energy scale is the result of many independent effects with different behavior with respect to properties of jets like E_T and η and represents therefore a leading order correction. So, second order effects can arise from uncertainties on single levels of correction of the jet energies and affect both the signal yield and the signal efficiencies.

    To evaluate these possible effects, we built mass templates for the signal varying by $\pm 1\,\sigma$ the individual levels of correction and generate different sets of PEs from which we derive a residual JES contribuition, not accounted for already in the previous item, of $(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}\,(\mathrm{Res.\,JES - efficiency}) = 2.1\ \%$.

15. **Primary vertex.** The distribution of the z_0 coordinate in the Monte Carlo simulation is slightly smaller than what seen on the data. We account for this multiplying by a scale factor which is known with a 0.2% uncertainty. The same amount of uncertainty is expected on the cross section.

16. **Trigger.** Studies on the trigger simulation indicate the need for rescaling the trigger efficiency by -1.8%. We then consider a 1.8% relative uncertainty on the cross section.

Table II shows a summary of all the relative systematic uncertainties corresponding to the two cases described above. The quadrature sum of all contributions, gives the total relative uncertainty of ± 13.7 % for the AH-TMT case, and ± 15.3 % for the CDF-COMB case.

Case	AH-TMT	CDF-COMB
Source	$(\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}})^{syst}$ (%)	
Cross section calibration	0.8	
Generator	4.2	
BR($W \to hadrons$)	0.8	
ISR/FSR	0.6	
b-jets energy scale	2.8	
b-tagging effficiency	5.4	
PDF	3.4	
Pileup	2.5	
Color Reconnections	0.8	
Templates Statistics	0.8	
Background Shape	0.3	
Background Normalization	8.2	
JES (efficiency)	6.1	9.2
Residual JES	2.1	
Primary vertex	0.2	
Trigger	1.8	
Total	13.7	15.3

TABLE II: Breakdown of *expected* relative systematic uncertainties from different sources and their respective amount. The total relative uncertainty is obtained by the quadrature sum of all individual contributions.

VI. CROSS SECTION MEASUREMENT

After the neural network based kinematic selection with a cut on the output $N_{out} \geq 0.90$ ($N_{out} \geq 0.88$), followed by the cut on the reconstructed top mass fit $\chi^2 \leq 6$ ($\chi^2 \leq 5$) for events with 1-tag ($\geq$ 2-tags), we are left with 3452 events with 1 b-tag and 441 events with ≥ 2 b-tags, as defined by the identification of displaced vertices. The expected background, corrected for the contribution from $t\bar{t}$ events to the data-driven estimate, amounts to 2785 ± 83 and 201 ± 29 events respectively. For these events the mass likelihood fit described in [1] returned the signal yields of 643 ± 80 and 216 ± 25 respectively. These yields are turned into a cross section measurement once we account for the signal efficiencies. These efficiencies have been evaluated specifically in two cases, one corresponding to $M_{top} = 175$ GeV/c^2 and ΔJES $= -0.3$ (i.e. the values measured in the all-hadronic channel), and the other corresponding to $M_{top} = 172.5$ GeV/c^2 and ΔJES $= 0$ (i.e. the values to be used for the CDF combination), and are summarized in Table III, along with the cross section values from the likelihood fit. To be accurate we use the precise value of the integrated luminosity, i.e. $L = 2.874$ fb^{-1}.

To evaluate the statistical uncertainty we perform the fit setting to zero the uncertainties on the integrated luminosity and efficiencies, while considering a purely statistical uncertainty of ± 59 and ± 21 on the signal yields.

CDF preliminary		
Case	AH-TMT	CDF-COMB
Integrated luminosity	2874 pb	
Signal yield (1 tag)	643 ± 80 (stat-only=59)	
Signal yield (≥ 2 tags)	216 ± 25 (stat-only=21)	
Signal efficiency (1 tag)	3.29%	3.27%
Signal efficiency (≥ 2 tags)	0.93%	0.95%
Cross section from the fit	$7.37 \pm 0.49(stat)$ pb	$7.34 \pm 0.51(stat)$ pb
Calibrated cross section	$7.24 \pm 0.48(stat)$ pb	$7.21 \pm 0.50(stat)$ pb

TABLE III: Signal yields including their statistical uncertainty, signal efficiencies, cross section from the fit and calibrated cross section (including statistical uncertainties). Values referring to the two cases.

Once we multiply by the calibration constant and account for the relative systematic uncertainties we obtain the following measurements:

$$\sigma_{t\bar{t}} = 7.24 \pm 0.48\,(stat) \pm 0.99\,(syst) \pm 0.42\,(lumi)\ \text{pb} \quad (\text{AH} - \text{TMT})$$

for the AH-TMT case and

$$\sigma_{t\bar{t}} = 7.21 \pm 0.50\,(stat) \pm 1.10\,(syst) \pm 0.42\,(lumi)\ \text{pb} \quad (\text{CDF} - \text{COMB})$$

for the CDF-COMB case.
The last one of the uncertainties corresponds to the 6% uncertainty expected on the measured integrated luminosity.

To evidentiate the dependence on the top quark mass we quote also the central values of the cross sections for $M_{top} = 170$ and 175 GeV/c^2 and ΔJES=0: $\sigma_{t\bar{t}}(170) = 7.29$ pb and $\sigma_{t\bar{t}}(175) = 7.00$ pb.

These measurements represent an improvement with respect to the previous all-hadronic measurement [7] of

$$\sigma_{t\bar{t}} = 8.3 \pm 1.0\,(stat)^{+2.0}_{-1.5}\,(syst) \pm 0.5\,(lumi)\ \text{pb} \quad (\text{AH},\ 1\ \text{fb}^{-1})$$

We notice that the statistical uncertainty improved more than the expected factor, $1.0/\sqrt{3} = 0.6$, because of the improvement in the selection. As for the systematic uncertainty, we have improved on some contribution (the JES for instance) but added additional terms. All in all the relative systematic uncertainty improves from 21% to 15%.

We would like to remark that the signal candidates used here do not represent a "simple" excess w.r.t. the background estimate, but an "excess" with $t\bar{t}$ identity, as proven by the reconstruced top mass, which behaves as expected for $t\bar{t}$ events. It is true, however, that a counting experiment would have had, in principle, less sources of systematic uncertainty.

VII. CONCLUSIONS

We have described in this note the combination of the signal yields from the Template Method Technique top quark mass measurement [1] with the evaluation of signal efficiencies, to obtain a cross section measurement. Due to the strong dependence of the signal efficiencies on the assumed top quark mass and the assumed JES, two different choices have been made and two corresponding cross section measurements have been derived, one for a stand-alone all-hadronic measurement, and one to be used in a CDF combination.

[1] The CDF Collaboration, *Measurement of the top mass with in-situ jet energy scale calibration in the all-hadronic channel using the Template Method with 2.9 fb^{-1}*, CDF public Note 9694, February 2009, and http://www-cdf.fnal.gov/physics/new/top/2009/mass/TMT_AH_P19_public/TMT_AH_3fb.html.

[2] H. L. Lai *et al.*, Eur. Phys. J. **C12**, 375 (2000).

[3] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S Thorne, Eur. Phys. J. C **4**, 463 (1998).

[4] P. Skands, D. Wicke, *Non-perturbative QCD effects and the Top Mass at the Tevatron*, arXiv:0807.3248 [hep-ph].

[5] P. Skands, *Color Reconnections, Multijets*, Presentation at "MPI@LHC' 08, First International Workshop on Multiple Partonic Interactions at the LHC", Perugia, Italy, 27–31 October 2008.

[6] H. Hoeth, *Tuning and Generator comparison*, Presentation at "MPI@LHC' 08, First International Workshop on Multiple Partonic Interactions at the LHC", Perugia, Italy, 27–31 October 2008.

[7] T. Aaltonen et al., Phys. Rev. D 76, 072009, 2007.