Fermi National Accelerator Laboratory

FERMILAB-Conf-94/314-E
CDF

Evidence for $t\bar{t}$ Production at the Tevatron: Statistical Significance and Cross Section

Jacobo Konigsberg
The CDF Collaboration

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

Physics Department, Harvard University
42 Oxford St. Cambridge, MA. 02138

September 1994

Proceedings of the *Eighth Meeting of the Division of Particles and Fields of the Americal Physical Society,* Albuquerque, NM, August 2-6, 1994.

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CHO3000 with the United States Department of Energy

EVIDENCE FOR $t\bar{t}$ PRODUCTION AT THE TEVATRON: STATISTICAL SIGNIFICANCE AND CROSS-SECTION

JACOBO KONIGSBERG*
Physics Department, Harvard University, 42 Oxford St., Cambridge, Mass. 02138

ABSTRACT

We summarize here the results of the "counting experiments" by the CDF Collaboration in the search for $t\bar{t}$ production in $p\bar{p}$ collisions at $\sqrt{s} = 1800\ TeV$ at the Tevatron.[1] We analyze their statistical significance by calculating the probability that the observed excess is a fluctuation of the expected backgrounds and, assuming the excess is from top events, extract a measurement of the $t\bar{t}$ production cross-section.

1. Statistical Significance of the Counting Experiments

The counting experiments that search for the top quark in the CDF experiment all yield an excess of events over the estimated background. The reader is referred to the talks in these proceedings that summarize the searches.[1] In this Section we test the "Null Hypothesis"; we find the probability that the number of observed events is consistent with a fluctuation of the estimated background. We first do this for each of the experiments separately and then we combine the experiments. The detailed account of these derivations can be found in the PRD article published by the CDF collaboration[2]

1.1. Significance of the Individual Analyses

If n_B^i is the estimated background for each of the analyses and σ_B^i is its uncertainty on n_B^i, then the probability $P_i(n)$ that we observe n events is formed by smearing a Poisson distribution with mean n_B^i by a Gaussian distribution with mean n_B^i and width σ_B^i.

$$P_i(n) \sim Gaussian(\mathbf{n_B^i}, \sigma_\mathbf{B}^\mathbf{i}) \times Poisson(\mathbf{n_B^i}; n)$$

From this distribution we find the probability for having $n \geq n_{obs}$. Table 1 summarizes the results for the individual experiments. We notice that although the individual probabilities are small they are not negligible. The SVX tag refers to tagging heavy flavour with the CDF silicon vertex detector. The SLT algorithm tags heavy flavor by looking for soft leptons.

---

*Representing the CDF Collaboration

Analysis	n_B^i	σ_B^i	n_{obs}	$P_i(n \geq n_{obs})$
Dileptons	0.56	$^{+0.25}_{-0.13}$	2	12%
Lepton + jets + SVX tag	2.3	± 0.3	6	3.2%
Lepton + jets + SLT tag	3.1	± 0.3	7	3.8%

Table 1. Probability of a background fluctuation for the individual experiments

1.2. Significance of the Combined Analyses

1.2.1. Significance by counting events

In the sets of 6 and 7 events found by the independent SVX and SLT analyses respectively, 3 events are in common. Thus 3 events are double-tagged (the two tags are not necessarily in the same jet). If we combine the analyses by simply counting events that are tagged we need to subtract the overlap between the SLT and the SVX analyses when we add the number of events found and the background expected. We have therefore 10 events in the lepton + jets channels and 2 events in the dilepton channel for a total of 12. The total background is found by adding the first column in Table 1 and subtracting the expected overlap in background between the SLT and SVX analyses which amounts to 0.26 events.[2] This results in $5.7^{+0.47}_{-0.44}$ background events.

The probability that 5.7 events fluctuate to 12 or more is again found by smearing a Poisson distribution with a mean of 5.7 with a Gaussian distribution with the same mean and a sigma of 0.47. We find this probability to be 1.6%.

In combining the experiments this way we've ignored the fact that 3 of our lepton + jets events are double-tagged.

1.2.2. Significance including double-tags

Given that a tag in an event tagged by both SVX and SLT is approximately six times more likely to be from heavy flavor than from a mistag,[2] we combine the SVX and SLT analyses by simply counting the number of tags (both from SLT and SVX) in the lepton + jets sample. If an event is double-tagged (either in the same or in a different jet) we count twice. We therefore have 13 "counts" from the lepton + jets analyses and 2 from the dilepton analysis, for a total of 15 "counts". Note that for dileptons we do not count the b-tags in the events given that a-priori we did not request a b-tag for the selection. The acceptance for top events would be too small. However one of the dilepton candidate events is double-tagged.

The total background, without taking into account correlations between the tagging methods, is $5.96^{+0.49}_{-0.44}$, found by adding the first column in Table 1. The probability for 5.96 counts to fluctuate to 15 or more is 0.16%.

In order to properly take into account the background correlations between SVX and SLT taggers we estimate the combined probability of a background fluctuation using a Monte Carlo method. The correlations are studied in a sample of generic QCD jets. We find that the number of expected SLT tags in those jets with a negative L_{xy} is

17 ± 2 which is consistent with the 22 observed. (L_{xy} is the secondary vertex distance in the transverse plane. Negative values are obtained due to tracking errors and resolution effects). The probability for these mistags is found to be twice that of the independent probabilities. This correlation is due to the preference by both algorithms to tag jets with many tracks. For jets with positive L_{xy}, after subtracting the mistag content from resolution effects and leaving only heavy flavor, we expect 77 ± 9 SLT tags, consistent with the 66 observed. The probabilities for SLT and SVX to tag heavy flavor are found to be uncorrelated.

In the Monte Carlo method used for estimating the combined probability of a background fluctuation we take the estimates for the mean number of events and the uncertainty for each background type in the sample of 52 $W+ \geq 3$ jets. This includes $Wb\bar{b}$, $Wc\bar{c}$, Wc, $W+$no heavy flavor (which are estimated from Monte Carlo but scaled up to the amounts indicated by background method I[1,2]), and $b\bar{b}$, WW and $Z \rightarrow \tau\tau$. We perform a large number of Monte Carlo "background experiments". In each we sample from these populations with the constraint that $\sum n_i = 52$. We then "apply" the tagging algorithms including the correlations for mistags. We add the dilepton background independently.

We find that the probability for the background to fluctuate to 15 or more counts is:

$$P(n \geq 15) = 0.26\%$$

Had we used the method II background estimates[1,2] we would have obtained that $P(n \geq 15) = 0.036\%$.

We've tested the robustness of the excess by changing the jet E_T thresholds by $+5\ GeV(-5\ GeV)$ in the event selection. In this case we expect 3.6 (12.7) counts in background events, 5.8 (7.9) from top events (for $m_{top} = 160\ GeV/c^2$) and we observe 11 (19). We've also doubled the correlations between the tagging algorithms in the Monte Carlo method described above and we obtain $P(n \geq 15) = 0.37\%$.

We conclude that the excess has a small probability of coming from a background fluctuation and seems robust. However the limited statistics do not allow us to firmly establish the existence of the top quark. If we interpret this excess as due to $t\bar{t}$ production we can measure the cross-section, $\sigma_{t\bar{t}}$.

2. Cross-Section Measurement for $p\bar{p} \rightarrow t\bar{t} + X$

In order to calculate the top cross-section we need to re-estimate the amount of background in the 52 $W+ \geq 3$ *jets* event sample. (The estimates in the previous Section were done assuming the "Null Hypothesis"). We do this by an iterative method[2] and obtain that for the SVX (SLT) analysis we expect 1.6 ± 0.7 (1.5 ± 0.7) counts from backgrounds.

The cross sections are calculated by maximizing the following likelihood function:

$$L = e^{-\frac{(\int \mathcal{L}dt - \overline{\int \mathcal{L}dt})^2}{2\sigma_{\mathcal{L}}^2}} L_{DIL} \cdot L_{SVX} \cdot L_{SLT}$$

where each of the individual likelihoods is of the form

$$L_i = G(\epsilon_i, \overline{\epsilon_\mathbf{i}}, \sigma_{\epsilon_\mathbf{i}}) G(b_i, \overline{\mathbf{b_i}}, \sigma_{\mathbf{b_i}}) P(\{\epsilon_i \cdot \sigma_{t\bar{t}} \cdot \int \mathcal{L}dt + b_i\}, \mathbf{n_i}).$$

Here $G(x, \overline{x}, \sigma)$ is a Gaussian in x, with mean $\overline{x}$ and variance σ^2, and $P(\mu, n)$ is a Poisson probability for n with mean μ. In each of these likelihoods, $\overline{\epsilon_i}$ and σ_{ϵ_i} are the total acceptance and its uncertainty, $\overline{b_i}$ and σ_{b_i} are the expected background and its uncertainty, n_i is the number of observed candidate events, $\sigma_{t\bar{t}}$ is the $t\bar{t}$ production cross section, and $\overline{\int \mathcal{L} dt} = 19.3\ pb^{-1}$ is the integrated luminosity, with $\sigma_{\mathcal{L}}$ its 3.6% uncertainty.

To calculate the cross section from an individual analysis, the individual likelihood functions are used, in which case the maximum likelihood solution for $\sigma_{t\bar{t}}$ is just

$$\sigma_{t\bar{t}} = \frac{n - \overline{b}}{\overline{\epsilon} \cdot \overline{\int \mathcal{L} dt}}$$

The uncertainties on the measured cross section values are calculated as the $\Delta \log L = \frac{1}{2}$ points of the likelihood function.

The calculated cross sections from the individual SVX, SLT and dilepton results, as well as the combined result (labeled $\sigma_{t\bar{t}}^{ALL}$) are shown in Table 2. In Figure 1 we plot the calculated combined cross section as a function of mass, and the theoretical expectation from Reference.[3] Because the acceptance depends on m_{top}, four points are shown corresponding to measured values of the acceptance. Had we chosen to use the method II background estimate for SVX, and an equivalent estimate for SLT, the $t\bar{t}$ cross section measurement would have shifted upward by 11%. We also note that an alternate method of calculating the cross section, based on the total number of observed events, gives a result approximately 12% lower with comparable uncertainties. From a dataset of $lepton+ \geq 4jets$ we have estimated[4] the top mass to be $m_{top} = 174 \pm 16$ GeV/c^2. This mass yields:

$$\sigma_{t\bar{t}} = 13.9^{+6.1}_{-4.8}\ pb$$

This cross section is somewhat higher than the theoretical calculation[3] for the same mass. We address the mutual compatibility with a χ^2 analysis on our measured mass, our cross section as a function of mass, the theoretical cross section versus mass, and their respective uncertainties. We find that the three results are compatible at a confidence level of 13%. We note, however, that the QCD uncertainties on the top cross sections can be larger[5] than those reported in Reference.[3]

3. Conclusions

We have calculated the probability that the excess of events in our search for the top quark be due to a fluctuation in the estimated backgrounds. We find this probability to be 0.26%. The limited statistics do not allow us to definitely establish the existence of the top quark. Under the assumption of $t\bar{t}$ production we measure the mass to be $m_{top} = 174 \pm 16$ GeV/c^2 and the corresponding cross-section for $p\bar{p} \rightarrow t\bar{t} + X$ to be $\sigma_{t\bar{t}} = 13.9^{+6.1}_{-4.8}\ pb$.

M_{top}	120 GeV/c^2	140 GeV/c^2	160 GeV/c^2	180 GeV/c^2
$\sigma_{t\bar{t}}^{SVX}$(pb)	$21.5^{+17.2}_{-11.3}$	$14.9^{+12.2}_{-7.9}$	$13.0^{+10.6}_{-6.9}$	$12.4^{+10.0}_{-6.5}$
$\sigma_{t\bar{t}}^{SLT}$(pb)	$28.9^{+18.5}_{-13.3}$	$22.7^{+14.3}_{-10.4}$	$20.4^{+12.7}_{-9.3}$	$18.8^{+11.7}_{-8.6}$
$\sigma_{t\bar{t}}^{DIL}$(pb)	$15.2^{+19.5}_{-12.2}$	$11.3^{+14.2}_{-9.0}$	$9.6^{+12.0}_{-7.6}$	$8.8^{+11.0}_{-7.0}$
$\sigma_{t\bar{t}}^{ALL}$(pb)	$22.7^{+10.0}_{-7.9}$	$16.8^{+7.4}_{-5.9}$	$14.7^{+6.5}_{-5.1}$	$13.7^{+6.0}_{-4.7}$

Table 2. $t\bar{t}$ production cross sections calculated from the individual analyses and from the combination of the three analyses.

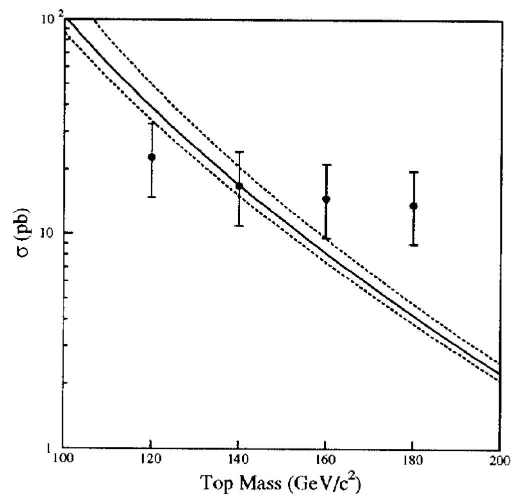

Fig. 1. Combined $t\bar{t}$ production cross section vs. M_{top} from data (points) and theory.[3] The dashed lines are estimates of the theoretical uncertainty quoted in Reference.[3]

References

1. These analyses are described in the talks by J. Benlloch and G. Watts in these proceedings.
2. F. Abe *el al*, Fermilab-Pub-94/097-E, submitted to *Phys. Rev.* **D**.
3. E. Laenen, J. Smith, and W. L. van Neerven, Nucl. Phys. **B369**, 543 (1992).
4. Talk by B. Harral in these proceedings.
5. R.K. Ellis, Phys. Letters, B259, 492(1991).