Fermi National Accelerator Laboratory

FERMILAB-Conf-94/152-E

CDF

# Multi-Jet Analysis for the Top Search at the Fermilab Collider

The CDF Collaboration

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

June 1994

Submitted to the *27th International Conference on High Energy Physics*, Glasgow, Scotland, July 20-27, 1994

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

## Disclaimer

Contribution code: gls0430
CDF/PUB/TOP/PUBLIC/2593
FERMILAB-CONF-94/152-E

# Multi–jet Analysis for the Top Search at the Fermilab Collider

The CDF collaboration[1]

**Abstract**

The search for a top signal in the multi-jet sample is presented. The $t\bar{t}$ decays considered here are the Standard Model ones in which both $W$'s decay into a quark pair, thus generating a nominal *6-jets topology*. Despite the enormous initial background rate, a reasonable signal-to-background ratio has been achieved once the SVX detector is used for $b$–tagging the events. Application of the method to the existing data set is presented, along with expectations for future runs.

## 1 Introduction

The Standard Model requires the top quark as a weak isospin partner of the $b$ quark. The top mass is not predicted by the theory, however global fits to electroweak measurements[2] favor a top mass of $m_t = 174^{+11+17}_{-12-19}$ $GeV/c^2$ while CDF has presented evidence[3] for a top quark with mass $174 \pm 10^{+13}_{-12}$ $GeV/c^2$ and a production cross section of $13.9^{+6.1}_{-4.8}$ $pb$.

In $p\bar{p}$ collisions top quarks are expected to be produced in pairs by both gluon fusion and $q\bar{q}$ annihilation, the latter being dominant if $m_t > 100$ $GeV/c^2$. Top quarks then decay almost exclusively to a $W$ boson and a $b$ quark. The final state topology depends on the decays of the two $W$'s, and we distinguish the following channels: *di-lepton* where both $W$'s decay leptonically

[1] Contact person: A. Castro, FNALD::CASTRO

[2] B. Pietrzyk for the LEP Collaborations and the LEP Electroweak Working Group; XXIXth Rencontres de Moriond, Méribel, Savoie, France, March 12-19, 1994.

[3] F. Abe *et al.*, "Evidence for Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ $TeV$." submitted to Phys. Rev. D and FERMILAB-PUB-94-097-E.

Submitted 27th International Conference on High Energy Physics, University of Glasgow, Glasgow, Scotland, July 20-27, 1994

to an electron or a muon (branching ratio $\simeq 5\%$), *single* lepton where only one $W$ decays to $e$ or $\mu$ ($BR \simeq 30\%$) and the *fully hadronic* channel where both $W$'s decay to a quark pair ($BR \simeq 44\%$). The top search in the *fully hadronic* channel is the topic of this paper.

In the multi-jet sample $t\bar{t}$ events derive mainly from the decays of both $W$'s to a $q\bar{q}'$ pair, with at least 6 hard partons in the final topology. The search for top in this sample is quite a challenging task, since the background rate due to the normal QCD processes is several orders of magnitude higher than the signal rate. For the 1992/93 Fermilab Collider run (Run Ia) the multi-jet ($\geq 4$ jets) yield from the CDF trigger turned out to be $\approx 2 \times 10^5$ events, while the total cross section for $t\bar{t}$ production at 1.8 $TeV$ is expected to be quite low[4], going from $\approx 40$ $pb$ when $m_t = 120$ $GeV/c^2$ to $\approx 4$ $pb$ when $m_t = 180$ $GeV/c^2$, corresponding to a signal–to–background ($S/B$) ratio at trigger level of $\approx 1/500$ and $\approx 1/3500$ respectively. The tiny $S/B$ ratio can be drastically enhanced by requiring a *6-jets topology* and with an adequate kinematical selection. The application of a secondary vertex finding algorithm in order to tag the presence of heavy flavors in the event allows eventually to reach a reasonable $S/B$ ratio.

## 2 Kinematical selection

The data set used for this analysis is based on a multi-jet trigger and amounts to $\int \mathcal{L} = 17.0 \pm 0.6$ $pb^{-1}$, corresponding to $\approx 90\%$ of the total integrated luminosity collected with the Collider Detector at Fermilab (CDF) between August of 1992 and May of 1993. The trigger required at least 4 jets with $E_T > 10$ $GeV$ and its cross section amounted to $\sigma \approx 10$ $nb$. The efficiency of such a trigger is quite high for $t\bar{t}$ events with both W's decaying hadronically ($\epsilon_{trigger} \geq 87\%$ for $m_t \geq 160$ $GeV/c^2$).

The purpose of the kinematical selection is to drastically reduce the data sample, which consists mainly of background events, while keeping an efficiency as high as possible for $t\bar{t}$ detection. We want to have a reasonable efficiency over a wide range of top masses ($120 - 200$ $GeV/c^2$) and in the following we will use the average mass value of 160 $GeV/c^2$ as a guideline in tuning our selection.

The first stage (*loose cuts*) simply reflects the topology we are looking for, with a requirement of $\geq 6$ and $\leq 9$ jets (clustering cone R=0.4, $E_T \geq 15$ $GeV$, $|\eta| \leq 2.4$), and selects 3575 events in the *multi-jet* sample. The efficiency of such a cut is $\epsilon_{loose} = 40 - 50\%$ for $m_t = 160 - 200$ $GeV/c^2$ (where both W's decay hadronically, i.e. *exclusive* sample).

To increase the $S/B$ ratio while maintaining a high efficiency for $t\bar{t}$ we consider the following

[4] E. Laenen *et al.*, Phys. Lett 312B, 254 (1994)

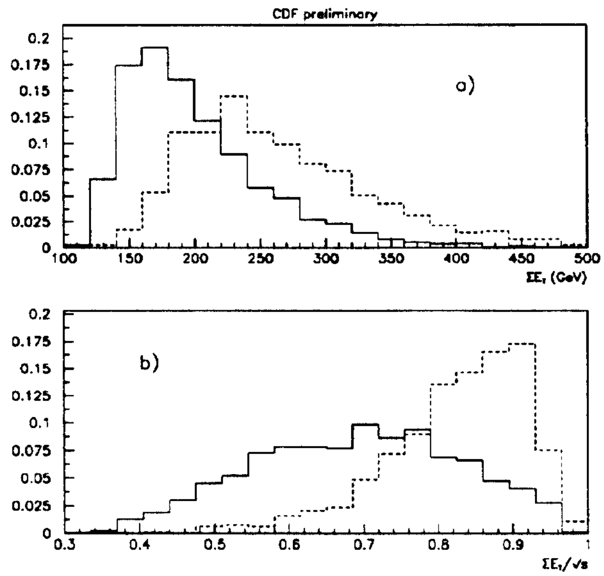


Figure 1: (*a*) Spectra (normalized to equal area) for $\sum E_T$ at *loose* level and (*b*) $\sum E_T/\sqrt{\hat{s}}$ after the $\sum E_T$ cut. Solid line for $multi-jet$ data and dashed line for Isajet $t\bar{t}$ ($m_t = 160\ GeV/c^2$) events.

kinematical variables calculated over all jets with $E_T \geq 15\ GeV$ and $|\eta| \leq 2.4$:

- $\sum E_T = \sum_{j=1}^{N_{jet}} E_T^j$
- $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$ where $x_{1,2} = (\sum E \pm \sum p_z)/\sqrt{s}$
- the aplanarity $A$, calculated in the center of mass system of all jets

For QCD events the sum of all the jet tranverse energies has a steeply falling distribution while for $t\bar{t}$ events this distribution is supposed to concentrate around a value which is between 1 and 2 times the top mass. Fig. 1 *a* shows the comparison between the $\sum E_T$ distributions (normalized to 1) for the *multi-jet* sample and Isajet $t\bar{t}$ ($m_t = 160\ GeV/c^2$) events, the rising edge at 140 $GeV$ in the *multi-jet* data being due to the event selection criteria. An additional characteristics of $t\bar{t}$ decays is that a large fraction of the energy available to the hard scattering is expected to be deposited centrally in the detector. This behavior is reflected in the variable *transverse energy fraction*, $\sum E_T/\sqrt{\hat{s}}$, which approaches 1 for $t\bar{t}$ decays. Fig. 1 *b* shows a comparison between normalized $\sum E_T/\sqrt{\hat{s}}$ distributions for *multi-jet* and $t\bar{t}$ Monte Carlo events: the discriminating power of this variable is apparent, since an unbiased QCD sample would populate mostly the forward region ($\sum E_T/\sqrt{\hat{s}} \ll 1$).

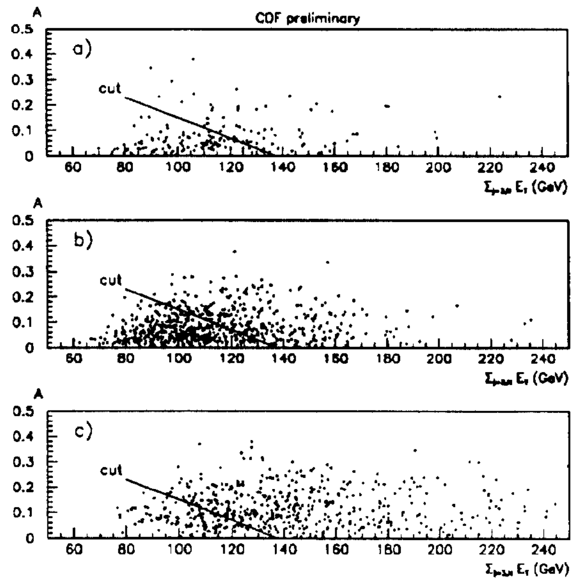


Figure 2: Aplanarity *vs* $\sum_3^N E_T$ scatter plot for $(a)$ Herwig QCD, $(b)$ *multi-jet* and $(c)$ Isajet $t\bar{t}$ ($m_t = 160\ GeV/c^2$) events after *tight* cuts have been applied.

The chosen cuts on these two variables are: $\sum E_T \geq (100 + 15 \times N_{jet})\ GeV$ and $\sum E_T/\sqrt{\hat{s}} \geq 0.7$, obtained by optimizing the signal significance $S/\sqrt{B}$ over a wide range of top masses, where as background we assume the *multi-jet* events themselves.

After the application of these so–called *tight* cuts, 754 events remain. This new set of cuts has an efficiency $\epsilon_{tight} = 74 - 82\%$ for $m_t = 160 - 200\ GeV/c^2$, with respect to the *loose* sample.

The main background at this stage comes from *top*–less QCD $2 \rightarrow 2$ processes with additional gluon radiation. These events have essentially two strong back-to-back jets in the CM system, making the whole event quite planar ($A \approx 0$) and with a large fraction of the total transverse energy going into the 2 leading jets. A strong contribution from gluon radiation is what gives the $2 \rightarrow 2$ process the *6-jets topology*, making the event less planar and providing energy to the additional jets. Fig. 2 shows the scatter plots $A$ *vs* $\sum_{j=3}^{N_{jet}} E_T^j$ $(a)$ for a QCD Monte Carlo sample, $(b)$ for the *multi–jet* events, and $(c)$ for an Isajet $t\bar{t}$ ($m_t = 160\ GeV/c^2$) sample. The $t\bar{t}$ events have the tendency to populate the region at large aplanarity and high $\sum_{j=3}^{N_{jet}} E_T^j$. Cutting on such a scatter plot provides us with additional discrimination. We reject all events falling below the line indicate on fig. 2: $A \geq (-0.004 \times \sum_{j=3}^{N_{jet}} E_T^j + 0.55)$. The number of *multi–jet* events surviving this *final* selection is

291. The efficiency of this so-called *final* set of cuts is $\epsilon_{final} = 78 - 84\%$ for $m_t = 160 - 200\ GeV/c^2$ (*exclusive* sample), with respect to the *tight* sample.

We have looked at distributions for various kinematical variables such as $E_T^{jet}$, $< |\eta^{jet}| >$, $\sum E_T$, invariant mass of all jets, sphericity, and we have found that after the *final* selection the distributions for *multi-jet* and for $t\bar{t}$ Monte Carlo events at various top masses have become strongly similar. No further discrimination seems possible, over a wide top mass range, on the basis of kinematical quantities alone.

The global efficiency for $t\bar{t}$ events when we select only the hadronic decays for both W's, not including the $BR(W^+ \to q\bar{q}',\ W^- \to q\bar{q}') = 4/9$, is $\epsilon_{exclusive} = 20 - 32\%$ for $m_t = 160 - 200\ GeV/c^2$. However, if we consider all possible W decays, then the $W^+ \to q\bar{q}', W^- \to q\bar{q}'$ topology accounts for $\approx 80\%$ of the *final* events. In order to have the largest possible acceptance for $t\bar{t}$ events we will consider from now on all possible W decays (*inclusive* sample). Table 1 gives the global efficiency along with the number of expected $t\bar{t}$ events in 17.0 $pb^{-1}$ and the expected $S/B$ ratio.

Table 1: *Final* cuts (*standard*)

| Isajet *inclusive* | | | | | |
|---|---|---|---|---|---|
| $m_t$ $(GeV/c^2)$ | 120 | 140 | 160 | 180 | 200 |
| *Global efficiency* | 2.8% | 6.3% | 11.5% | 15.6% | 18.9% |
| $t\bar{t}$ *events* | 18.3 | 18.1 | 16.0 | 11.2 | 7.2 |
| $S/B$ | 1/15 | 1/15 | 1/18 | 1/26 | 1/40 |

The residual backgrounds are believed to come mainly from QCD $2 \to 2$ processes with additional gluon radiation and from $W + jets$ production. A Monte Carlo study shows that QCD events represent $\approx 99\%$ of the total expected background.

The kinematical selection described so far has been devised to be effective over a wide range of top masses ($\geq 120\ GeV/c^2$) but we can tune it to be the most effective for high top masses ($m_t \simeq 180\ GeV/c^2$). This is done by choosing the cuts on the kinematical variables which maximize the signal significance $S/\sqrt{B}$ for Isajet $t\bar{t}$ events with $m_t = 180\ GeV/c^2$. The values chosen for these cuts are: $\sum E_T > 250\ GeV$, $\sum E_T/\sqrt{\hat{s}} > 0.8$ and $A \geq (-0.004 \times \sum_{j=3}^{N_{jet}} E_T^j + 0.60)$. The effect of this new set of cuts is impressive: the *multi-jet* sample is reduced to 76 events (instead of 291) with an efficiency $\epsilon_{rel} = (53.0 \pm 2.6)\%$ for Isajet Monte Carlo *inclusive* $t\bar{t}$ events, representing a $S/B$ gain of about 2. We will refer to this particular kinematical selection as the *high mass* kinematical selection as opposed to the *standard* one.

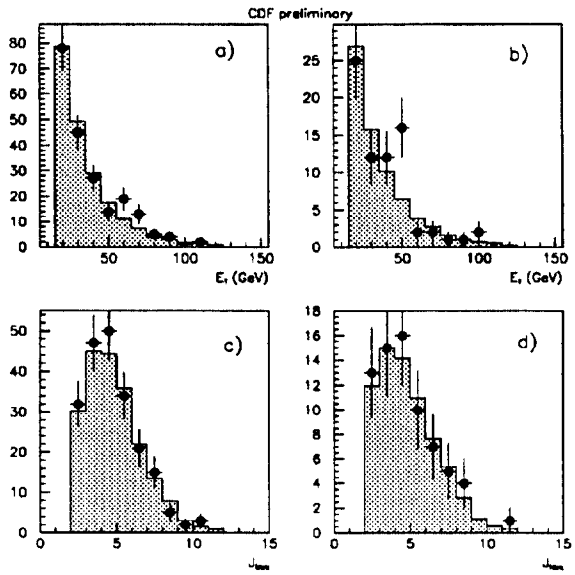


Figure 3: Jet-$E_T$ distribution for tagged jets with positive (*a*) and negative decay length (*b*) at *loose* level. Jet track multiplicity distribution for tagged jets with positive (*c*) and negative decay length (*d*) at *loose* level. Histogram=estimated tags, dots=observed tags. (Bars represent statistical errors).

# 3 $b$–tagging

Tagging the presence of $b$–hadrons in the event is a powerful way to improve the $S/B$ ratio taking advantage of the long lifetime of $b$–hadrons ($c\tau \simeq 400\ \mu m$) that will translate into a secondary vertex significantly displaced with respect to the interaction point (primary vertex). This secondary vertex may be found with the help of a microstrip vertex detector (SVX) in CDF. The algorithm is based on the search of at least one secondary vertex using the tracks belonging to a jet and well reconstructed inside the SVX. These tracks must be of good quality, have $p_T > 2.0\ GeV/c$ and an impact parameter $d$ with respect to the primary vertex larger than 3 times the quadrature sum of the impact parameter resolution and the primary vertex resolution. At least 2 of such tracks must lie inside a cone of $\cos\theta \geq 0.8$ centered on the jet axis and the reconstructed vertex has to be separated from the primary vertex by a distance (in the $R\phi$ plane) $|L_{xy}| \geq 3\sigma_{L_{xy}}$ where $\sigma_{L_{xy}}$ is the quadrature sum of primary and secondary vertex resolutions. This distance is defined positive (negative) when its projection along the jet axis is positive (negative).

The tagging efficiency is almost independent of the top mass and is $\epsilon_{b-tag} = 22 \pm 6\%$, however each tagged $t\bar{t}$ event has a probability $\approx 20\%$ of having more than one tagged jet.

In our kinematically selected sample the tags come, besides $t\bar{t}$ events, from real heavy flavor production (direct or through gluon splitting) or from tracking mismeasurements (*mis-tags*). The tags coming from a heavy flavor populate mainly the region $L_{xy} > 0$ with only a small presence at $L_{xy} < 0$ due to resolution effects, while *mis-tags* are symmetric around $L_{xy} = 0$. The tagging rate in inclusive jet events has the contributions from both sources so it can be used as a way to estimate the background. This tagging rate, defined to be the number of tagged jets divided by the number of jets having two or more good associated SVX tracks (so-called *taggable* jets), has been parametrised as a function of jet-$E_T$, the number of SVX tracks associated with the jet ($J_{Ntrk}$) and the jet pseudorapidity, and has been calculated using a sample of $\approx 210,000$ *taggable* jets coming out of the *multi-jet* trigger, that is the same trigger used for our search. This tagging rate assumes values $1.5 - 3.0\%$ ($0.5 - 1.0\%$) for positive (negative) tags, with a slight dependence on the jet-$E_T$ and a stronger one on the jet track multiplicity.

Table 2 summarizes, at each level of kinematical selection, the comparison between the observed number of tagged jets ($N_{obs}$) and the expected one ($N_{exp}$) calculated using the estimated tagging rate in generic jets.

Table 2: $b$-tags (*standard*)

| Selection | $N_{obs}$ + | $N_{exp}$ + | $N_{obs}$ - | $N_{exp}$ - |
|---|---|---|---|---|
| *loose* | 209 | 204 | 73 | 70 |
| *tight* | 69 | 55.4 | 26 | 19.1 |
| *final* | 31 | 22.5 | 9 | 7.7 |

The statistical error in the expected number of tags at *loose* level is $\approx 0.2\%$. A preliminary estimate of the systematic error has been obtained calculating the tagging rate using a different control sample of 67,000 events passing a 50 *GeV* trigger (JET50). The difference between the number of expected tags (i.e. background estimation) using the two sample is 15% for positive rates and 20% for negative ones.

In the *loose* sample we expect only a small fraction of $t\bar{t}$ events so we compare, in fig. 3 *a* and *b*, the distribution for observed and estimated tagged jets as a function of the jet $E_T$ for positive and negative decay lengths. Fig. 3 *c* and *d* show them instead as a function of the jet track multiplicity. There is a very good agreement not only in the absolute predictions but also in the differential ones. This indicates that our rate estimate is indeed a good way to calculate the background. The agreement is present, within errors, also for the negative tags, which however are used in this analysis only for a consistency check. Fig. 4 shows $E_T$ and $J_{Ntrk}$ distributions for tagged jets in the *final*

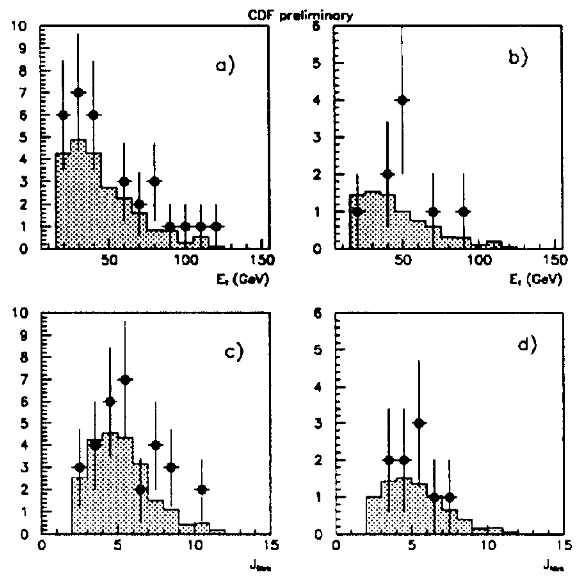


Figure 4: Jet–$E_T$ distribution for tagged jets with positive (*a*) and negative decay length (*b*) at *final* level. Jet track multiplicity distribution for tagged jets with positive (*c*) and negative decay length (*d*) at *final* level. Histogram=estimated tags, dots=observed tags. (Bars represent statistical errors).

sample.

Table 3 shows the expected number of $t\bar{t}$ tags and events in the *final* sample, after $b$–tagging, considering all possible W decays, along with the expected $S/B$ .

Table 3: *Final* cuts + $b$–tagging (*standard*)

| Isajet *inclusive* | | | | | |
|---|---|---|---|---|---|
| $m_t$ $(GeV/c^2)$ | 120 | 140 | 160 | 180 | 200 |
| $t\bar{t}$ *tags* | $6.2 \pm 2.1$ | $5.6 \pm 1.7$ | $4.2 \pm 1.2$ | $2.9 \pm 0.8$ | $2.1 \pm 0.6$ |
| $t\bar{t}$ *events* | $4.6 \pm 1.6$ | $4.5 \pm 1.4$ | $3.5 \pm 1.0$ | $2.4 \pm 0.7$ | $1.7 \pm 0.5$ |
| $S/B$ (*tags*) | 1/4.0 | 1/4.5 | 1/6.4 | 1/9.7 | 1/13.8 |
| $S/B$ (*events*) | 1/4.9 | 1/5.0 | 1/6.7 | 1/10.3 | 1/14.5 |

There is indeed a small difference between $N_{obs}$ and $N_{exp}$ (Table 2). This excess, $N_{obs} - N_{exp} = 8.5 \pm 5.6$ (*stat*) $\pm$ 3.4 (*syst*), is not statistically significant however its size is comparable with what is expected from the results of the *single* and *di-lepton* CDF analyses[3]: using the measured cross section and the efficiency extrapolated for $m_t = 174$ $GeV/c^2$, we would expect to see $8.9 \pm 2.4$

positive $t\bar{t}$ tags.

Table 4 shows the result of the application of the $b$–tagging on the *high mass* selected sample. In this case the number of expected positive $t\bar{t}$ tags is $4.5 \pm 1.4$ to be compared with $N_{obs} - N_{exp} = 4.8 \pm 3.3$ (*stat*) $\pm$ 0.9 (*syst*). Again the excess is not statistically significant and systematic uncertainties are quite important here. These effects have not been studied in detail yet, but are the object of our current studies.

Table 4: $b$–tags (*high mass*)

| Selection | $N_{obs}$ + | $N_{exp}$ + | $N_{obs}$ - | $N_{exp}$ - |
|---|---|---|---|---|
| *loose* | 209 | 204 | 73 | 70 |
| *tight'* | 17 | 12.3 | 4 | 4.3 |
| *final'* | 11 | 6.2 | 3 | 2.2 |

# 4 Conclusions

We have performed an extensive study on the possibility of extracting a statistically significant sample of $t\bar{t}$ events from the *multi–jet* decay channel. Despite the huge initial background, we finally achieve, for $m_t = 160\ GeV/c^2$, a $S/B$ ratio of the order of 1/20 after the kinematical selection, which can be increased to $\approx 1/6$ after a secondary vertex $b$–tagging algorithm is applied. If instead we assume for the top mass and the $t\bar{t}$ production cross section the values indicated by other CDF analyses[3], then the $S/B$ ratio after the $b$–tagging is $\approx 1/2$ and could be as high as $\approx 1/1$ using a *high mass* kinematical selection.

Fig. 5 summarizes the performance of this analysis in terms of S/B ratio at the various stages of the selection.

Assuming a $S/B \approx 1/6$ then in order to observe a $2\sigma$ effect, we would need at least 25 $t\bar{t}$ events. Supposing for the detector the same performance as in Run Ia, this number would be reachable if the total luminosity collected in the ongoing run (Run Ib) exceeds 80 $pb^{-1}$.

If however we tune the analysis for a high mass top and use the cross section from ref. 3, then the $S/B$ ratio is expected to be $\approx 1/1$. This means that 80 $pb^{-1}$ of integrated luminosity for Run Ib could provide a $3 - 4\sigma$ excess over the background.

The background rate is estimated from the data themselves. With this background we observe a small excess in the number of tags which is not statistically significant but whose size is comparable with what is expected for top production using the mass and cross section from ref. 3.

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital

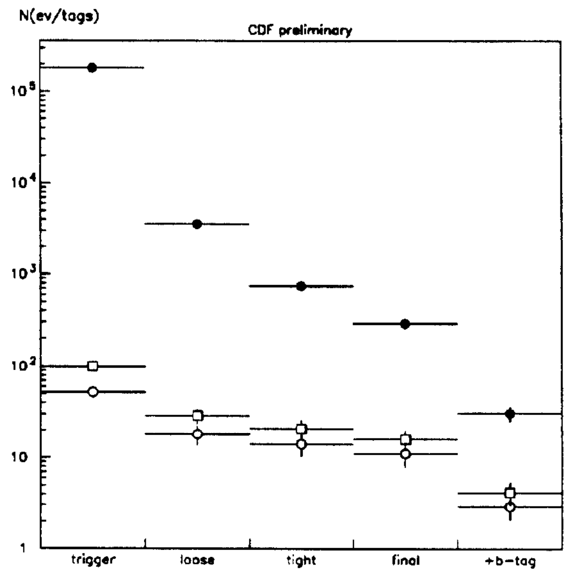


Figure 5: Number of events (or tags) observed [•] and expected for Isajet $t\bar{t}$ events ($m_t$ = 160 $GeV/c^2$ [□], $m_t$ = 180 $GeV/c^2$ [o], cross sections of ref. 4) at each stage of the *standard* analysis.

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, Science and Culture of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the A. P. Sloan Foundation; and the Alexander von Humboldt-Stiftung.