Top Quark Physics at D0
G. Gutierrez, for the D0 collaboration.

Fermilab, PO Box 500, Batavia, IL 60510

Abstract.
This note describes some of the most recent top quark measurements produced by the D0 collaboration. At the same time it tries to give the reader an understanding of why after more that 10 years since the top quark discovery only the top quark pair production cross section and the top quark mass have been measured with relevant accuracy. The rewards that an increase in statistics will bring in the quest to completely characterize the top quark are also discussed.

Keywords: Top Physics
PACS: 14.65.Ha

1. INTRODUCTION

The top quark with its very large mass and Yukawa couplings close to unity is expected to play a central role in the Standard Model (SM). The top quark mass together with the $W$ boson mass provide today’s best constraints on the Higgs boson mass. Any deviations from the top quark’s SM predictions will be a signal for new physics. But even after more than a decade since its discovery little is known today about the top quark. The decay width, spin and parity have not yet been measured, the top quark production via the weak force has only recently been observed. Measurements of the lifetime, branching fractions and charge are ongoing but are still far from testing the SM expectations [1, 2]. Only the top quark mass and the pair production cross section have been measured with relevant accuracy. Measurements of the top quark decay vertex $V – A$ structure will only achieve relevant accuracy as the Tevatron approaches the $8 \text{ fb}^{-1}$ of delivered luminosity.

The reason for all this in one word is: statistics. Three of the most recent and important D0 results with $1 \text{ fb}^{-1}$ of data and the $W$ boson helicity measurement with $0.37 \text{ fb}^{-1}$ of data will be described in this note and will be used to illustrate the importance of statistics in different measurements. The top quark mass statistical error is the smallest because three jet invariant masses are peaked. The statistical error in the cross section is next in line because the backgrounds are manageable. Single top quark production is hard to measure because the backgrounds are very large. And all measurements involving angles need very large statistics because angular distributions cover all the allowed range and are fairly flat. For detail information on all D0 top quark physics measurements see Reference [1].

2. TOP QUARK MEASUREMENTS

Top quarks at the Tevatron are produced either through the strong or weak interactions. The strong interaction produces top anti-top quark pairs when quarks and gluons inside
the protons and anti-protons collide with each other. The weak interaction produces single top quarks via two different channels. In the s-channel, a beam quark and a beam anti-quark annihilate into a virtual \( W \) boson which in turn decays into a \( b \) and a top quark. In the t-channel, a beam quark produces a real quark and a virtual \( W \) boson and a gluon from the opposite beam produces a \( b\bar{b} \) pair, with one real and one virtual \( b \) quark; the virtual \( b \) and \( W \) then interact to produce a top quark. Even though the (strong) \( t\bar{t} \) cross section is only about a factor of two larger than the (weak) single top quark cross section, evidence for the production of single top was reported for the first time only recently [3]. The reason being that single top quark production offers fewer constraints and since the final state has fewer jets the backgrounds are much larger.

In the SM the top quark decays almost 100% of the time to a \( b \) quark and a \( W \) boson. The \( W \) in turns decays equally to a pair of leptons or a pair of quarks (e.g. \( W^+ \rightarrow e\nu_e, \mu\nu_\mu, u\bar{d}, c\bar{s} \)). Since there are three colors for every quark the \( W \) boson decay to two jets is three times more probable than its decay to an electron or a muon. The \( t\bar{t} \) final states can then be classified according to the \( W \) boson decay into all jets, lepton-jets, or dileptons, which are produced in a ratio of 9:3:1. The most accurate measurements have been done in the lepton+jets channel because it has more statistics that the dilepton channel and fewer background than the all jets one.

In the \( W \) boson center of mass (CM) the decay products have a momentum of 40 GeV/c, and in the top quark CM the \( b \) and the \( W \) have a momentum of 66 GeV/c (for a top quark mass of 170 GeV/c\(^2 \)). So the main objects in top quark analysis are a 40 GeV/c electron, muon or neutrino, a 40 GeV/c light jet, and a 66 GeV/c b-jet. Lorentz boosts increase or decrease the particle’s momenta depending on the direction of the boost relative to the particles. This has the effect of widening the momentum distributions keeping the peaks at the CM momenta. Therefore most event selection criteria require that jets, leptons and missing (neutrino) transverse momentum be bigger than 15-20 GeV/c, that the pseudorapidity \( \eta \) of these objects be within the active part of the detector (typically \(|\eta|<2\)), and that the leptons are isolated from the jets to avoid confusion with leptons that can be generated from the decay of particles inside a jet.

2.0.1. \( t\bar{t} \) cross section

A cross section measurement is performed by counting events. In the simplest case there is a number \( N \) of data events that passes all cuts and a prediction \( B \) of background events. The number of signal events is

\[
S = N - B
\]

and the cross section \( \sigma \) is proportional to \( S \),

\[
\sigma = S / (\epsilon \cdot L \cdot Br) \quad \text{where} \quad \epsilon \text{ is the selection efficiency,} \quad L \text{ the integrated luminosity and} \quad Br \text{ is the decay branching ratio for the final state under study.}
\]

If the experiment is performed many times the distribution of the number of data events \( N \) will have an rms equal to \( \sqrt{N} \) (the rms of a Poisson distribution). The fluctuations in \( N \) will be due to both the signal and background events, but \( \sigma \) is only proportional to \( S \), therefore the relative error in the cross section will be \( \sqrt{N}/N \).

Cross section measurements are classified as b-tagged or un-tagged when they are performed with or without b-jet identification. In the un-tagged analysis a probability \( P_S \) (\( P_B \)) for an event being signal (background) is calculated for every event. These proba-
bilities are calculated multiplying properly normalized one dimensional histograms obtained from MC. The histograms are defined using variables like Aplanarity, Sphericity, Centrality, $H_T$, $\Delta \phi$ and $K_T^{\text{min}}$ \cite{4}. A discriminant $D = P_S/(P_S + P_B)$ is then used to separate signal from background. Figure 1 shows the value of the discriminant for $e+\text{jets}$, $\mu+\text{jets}$ and both channels combined. In each histogram the discriminant is shown for data, $t\bar{t}$ signal and for the two main backgrounds $W+\text{jets}$ and multijets. The signal is extracted by fitting the discriminant distributions allowing the number of signal and background events to float. The number of events resulting from the fit are shown in the upper right hand corner in the discriminant plots. The statistical error can be estimated in the following way. Two bins can be formed by splitting the combined discriminant histogram at $D \approx 0.3$. The lower bin contains mostly background, in the upper bin the signal and background are distributed in approximately two equal parts. The extraction of the signal ($S = N - B$) from the second bin will be subject to fluctuations $\sqrt{S + B}/2$ in the number of events $N$ and $\sqrt{B}/2$ in the background prediction $B$. This last fluctuation comes from using the first bin to determine the number of background events. Adding the two errors in quadratures gives a total error for $S$ of $\sqrt{S + B}$ which implies a relative error in the cross section of $\sqrt{S + B}/S = \sqrt{1/224} = 12\%$. This error agrees very well with the 13% statistical error quoted in the final measurement given in Eq. 1 below. A relative error of 12% corresponds to 70 pure signal events as opposed to the 224 $t\bar{t}$ events measured in the un-tagged sample. This means that the presence of background produces most of the fluctuations observed in the statistical error and that this error could be reduced by reducing the background even at the expense of cutting a substantial amount of signal.

One way to reduce the background is by looking for the presence of $B$ mesons in $b$-jets. Due to their long lifetime the decay of $B$ mesons is observed as a vertex separated from the primary vertex of the interaction. If a secondary vertex is found inside a jet that jet is “$b$-tagged” and it is considered as a jet produced by a $b$ quark. Given that there are two $b$-jets in each $t\bar{t}$ event and that only a small part of the $W+\text{jets}$ background contains $b$-jets, using $b$-tagging keeps most of the signal while greatly reducing the backgrounds.

In the $b$-tagged cross section analysis the data is split into electrons and muons, three and four or more jets, and one or two or more $b$-tags. The number of expected signal and observed data events for these eight groups is shown in Table 1. If the experiment is performed many times the cross sections measured in each group will be distributed
as a Gaussian with an rms given by the cross section $\sigma$ times the relative error in that group $r_i = \sqrt{N_i}/S_i$, where $N_i$ ($S_i$) is the number of data (signal) events in group $i$. Since the measurements in each group are independent the cross section’s statistical error can be derived from the product of the Gaussian distributions in each group. This product has an rms $s$ given by $1/s^2 = \sum_i 1/(\sigma r_i)^2$, which leads to a total relative error of $(s/\sigma)^2 = (\sum_i 1/r_i^2)^{-1} = (\sum_i S_i^2/N_i)^{-1}$. An error $s/\sigma=6\%$ is obtained using the numbers in Table 1. This corresponds to an equivalent gain of a factor of four in data relative to the un-tagged measurement. This gain is almost exclusively due to the background rejection achieved by b-tagging. Figure 2 shows the predicted sample compositions for the $l+\text{jets}$ sample with exactly one and with two or more b-tagged jets.

The final cross section results are

$$l+\text{jets un-tagged: } \sigma_{\ell p p \to \ell+X} = 6.3^{+0.9}_{-0.8} \text{(stat)} \pm 0.7 \text{(syst)} \pm 0.4 \text{(lumi)} \text{ pb} \quad (1)$$

$$l+\text{jets b-tagged: } \sigma_{\ell p p \to \ell+X} = 8.3^{+0.9}_{-1.0} \text{(stat)}^{+0.6}_{-0.5} \text{(syst)} \pm 0.5 \text{(lumi)} \text{ pb} \quad (2)$$

It is clear from the b-tagged result that the cross section measurements are already not dominated by statistics. So a future increase in the amount of data will reduce the total error only if it can be used to reduce the systematic errors. The two largest systematic errors in the un-tagged analysis are the modeling of $W+\text{jets}$ and the discriminant template shapes. In the b-tagged analysis the two largest systematic errors are the large error in the $W+b\bar{b}+\text{jets}$ cross section and the uncertainty in the b-tagging probability.

### 2.0.2. Top quark mass

The most accurate top quark mass measurements are done in the lepton+jets channel. In this channel one top quark decays into three jets and the other top quark decays to a jet, a lepton and a neutrino. Because the neutrino can not be measured, the top quark mass measurement in the $l+\text{jets}$ channel is dominated by the invariant mass of the three jets coming from the decay of a top quark. This invariant mass has a typical width of about 20 GeV/c$^2$ and its mean, which is directly related to the top quark mass, has an error of $20/\sqrt{N}$, where $N$ is the number of signal events. For 1 fb$^{-1}$ of data $N \approx 150$, which means that the expected statistical error in the top quark mass is $\approx 1.6$ GeV/c$^2$. The reason it is possible to achieve a 1% measurement in the top quark mass but only a 10% measurement in the cross section is that the three jet invariant mass is very localized in comparison to its allowed range (say from 50 to 450 GeV/c$^2$).
Due to the presence of backgrounds and to the impossibility of knowing which jet comes from which parton (which gives rise to 24 possible combinations in the assignment between jets and partons), a fit to the three jet invariant mass is not the most effective way of extracting the top quark mass. In the analysis presented here a probability \( P(x|\alpha) \) is calculated for every event. The four momenta of the lepton and jets are labeled by \( x \) and the top quark mass \( M_{\text{top}} \), Jet Energy Scale \( JES \) and fraction of signal events \( f \) are labeled by \( \alpha \). Given that all events are independent, the probability (or likelihood) for a sample of \( N \) events is the product of the probabilities for each event \( L = \prod_{i=1}^{N} P(x_i|\alpha) \). If \( \alpha \) is known then \( L \) will be very close to maximal, otherwise a different set of events would have been observed. Or if \( \alpha \) is unknown it can be determined by maximizing \( L \). A given event in the final sample has a certain probability \( P_T(x|\alpha) \) of being signal and a probability \( P_B(x) \) of being background. Therefore the probability for an individual event is \( P(x|\alpha) = f P_T(x|\alpha) + (1 - f) P_B(x) \).

A large sample of \( \gamma\text{+jets} \) events was used to perform an independent study of the Jet Energy Scale. A Gaussian likelihood \( G(JES) \) with a mean \( JES = 1 \) and an rms of 0.037 is derived from this study. Since the \( t\bar{t} \) and \( \gamma\text{+jets} \) samples are independent, the likelihood for the combination of both samples is just the product of \( L \) and \( G \). The top quark mass is extracted by projecting this product onto the \( M_{\text{top}} \) axis:

\[
L(M_{\text{top}}) = \int df \, dJES \, L(M_{\text{top}}, JES, f) \, G(JES) \tag{3}
\]

The normalized likelihoods \( L(M_{\text{top}})/L_{\text{max}} \) are shown in Figure 3. The figure on the left shows the likelihood for the combined \( e\text{+jets} \) and \( \mu\text{+jets} \) samples for the un-tagged analysis. The center (right) figure shows the likelihoods for the \( e\text{+jets} \) (\( \mu\text{+jets} \)) sample in the b-tagged analysis. The top quark masses derived from these plots are:

\[
\begin{align*}
\text{1+jets un-tagged: } m_{\text{top}} & = 170.5 \pm 2.5(\text{stat+JES}) \pm 1.4(\text{syst}) \text{ GeV}/c^2 \tag{4} \\
\text{1+jets b-tagged: } m_{\text{top}} & = 170.5 \pm 2.4(\text{stat+JES}) \pm 1.2(\text{syst}) \text{ GeV}/c^2 \tag{5} \\
& = 170.5 \pm 1.8(\text{stat}) \pm 1.6(\text{JES}) \pm 1.2(\text{syst}) \text{ GeV}/c^2
\end{align*}
\]

With a weight of 39.7% measurement (5) has the largest single contribution to the world top quark mass combination [5].
2.0.3. W boson helicity

In the $W^+$ boson rest frame, the general angular distribution of the down type decay product ($e^+, \mu^+, d$ or $\bar{s}$) has the form

$$w(\theta^*) = |d^0_{11}(x)|^2 f_0^2 + |d^1_{11}(x)|^2 f_-^2 + |d^2_{11}(x)|^2 f_+^2,$$

where $x = \cos \theta^*$, $d^f_{m,m'}$ are the rotation matrices, $f^f_\lambda$ is the probability of finding the $W^+$ boson in a state of helicity $\lambda = (0, -1, +1)$ and $f_0 + f_- + f_+ = 1$. In the $W^+$ rest frame the z-axis is defined in the $W^+$ momentum direction before the boost that brings the $W$ boson to rest and $\theta^*$ is the angle of the down type decay product with the z-axis. In the case of the $W^-$ decay all helicities are reversed and the angular distribution remains the same. Using the rotation matrices the angular distribution is

$$w(\theta^*) \propto 2(1-x^2)f_0 + (1-x)^2 f_- + (1+x)^2 f_+.$$

In the SM $V-A$ current interaction the $W$ bosons originating from top quarks have well defined values of $f_0 = m_t^2/(m_t^2 + 2M_W^2 + m_b^2) \approx 0.70$ and $f_+ = m_b^2/(m_t^2 + 2M_W^2 + m_b^2) \approx 0$. For any linear combination of $V$ and $A$ currents the value of $f_0$ stays constant but $f_+$ changes. For example for a pure $V+A$ charge current $f_+ = 0.3$.

In this analysis [6] the $\cos \theta^*$ distribution is compared with templates for different values of $f_+$ obtained for constant $f_0 = 0.70$. The data selection involves both the use of a discriminant and b-tagging. The different assignments between partons and jets are resolved by performing a constrained fit to each event and selecting the combination with the smallest $\chi^2$. Figure 4 shows the $\cos \theta^*$ distribution for (a) $l+\text{jets}$ and (b) dilepton events. The SM prediction is shown as a solid line, the dashed line shows a pure $V+A$ interaction. Templates were calculated for seven values of $f_+$ and a binned Poisson likelihood was used to extract $f_+$ from the $\cos \theta^*$ distributions shown in Figure 4. The likelihood as a function of $f_+$ is shown in Figure 4 (right plot). The final result is [6]

$$f_+ = 0.056 \pm 0.080 \text{ (stat)} \pm 0.057 \text{ (syst)}$$

Since $f_0 = 0.70$ the value of $f_+$ is bounded between 0 and 0.3. In the absence of prior knowledge or a measurement, all values of $f_+$ between 0 and 0.3 are equally probable (flat likelihood). Then without performing the experiment it is possible to say that $f_+$ has a 68% probability of being inside the interval $0.68 \times 0.3 = 0.20$ or 34% probability (1 sigma) of being inside a 0.10 interval. The comparison of this number
with the statistical error in Eq. 6 clearly indicates that not much has been gained yet by performing the measurement. The reason for this is that contrary to, for example three jet invariant masses, angular distributions cover the entire allowed range and are usually fairly flat. This means that larger statistics are needed in order to achieve the same precision as that found in measurements like the top quark mass. Therefore analyses that involve measurements of angular distributions will be the main beneficiaries of the future increase in statistics.

2.0.4. Single Top Quark Production

Discovering single top quark production essentially involves a cross section measurement like the ones described in Subsection 2.0.1. The main difference is that the number of background events is much larger than the signal ones and therefore the use of both b-tagging and powerful discriminants is needed. D0 has recently announced for the first time evidence for single top quark production [3]. Three different algorithms were used in the analysis, Boosted Decision Trees (BDT), Matrix Element and Bayesian Neural Networks. Only the analysis with the largest significance (BDT) will be described here.

A BDT algorithm builds a sequence of “trees”. In each tree a group of events starting in the tree “trunk” will, after a series of decisions, temporarily occupy different “branches” until they all end up in different “leaves.” The decisions made to bifurcate a branch are optimized using signal and background MC events. Once optimized (or trained) the tree can be used to classify events. After a data or MC event that is run through the tree reaches a leaf a “leaf purity” \( T_n(i) \) is assigned to it. \( T_n(i) \), the output or leaf purity of tree \( n \) for event \( i \), is defined as \( s/(s+b) \) where \( s \) and \( b \) are the sum of signal and background weights for that leaf. An event is called signal if \( T_n(i) \) is greater than a specified number (usually 0.5) and background otherwise. During the training session an error \( \text{err}_n \) is assigned to each tree based on the number of misclassified events, e.g. a MC signal event called background. The \( n+1 \) tree in the sequence is built weighting (or boosting) the misclassified events with a function of \( \text{err}_n \). Until stability is reached, and as a consequence of weighting, each tree in the sequence will have a reduced number of misclassified events. The BDT output \( T(i) = \sum_n \alpha_n T_n(i) \) for event \( i \) is a linear combination of the outputs \( T_n(i) \) of all the trees in the sequence. The coefficients \( \alpha_n \) are functions.
Figure 5 shows the BDT output for a $t\bar{t}$ enriched sample (left plot) and for the final sample used in the analysis (center plot). The $tb$ and $tqb$ labels in the plots stand for s-channel and t-channel production of single top quarks which, as pointed out in Section 2, are produced together with a b-quark in the s-channel and with a b- and a light quark in the t-channel. A binned likelihood fit over the decision tree output or discriminant is used to extract the single top quark signal. The right plot on Figure 5 shows the posterior probability density obtained from the likelihood assuming a flat nonnegative prior probability. The cross section extracted from this plot is

$$\sigma(p\bar{p}\to tb+X, tqb+X) = 4.9 \pm 1.4 \text{ (stat+syst) pb}$$

(7)

Using ensemble test studies, the probability that the background would fluctuate up to produce a measured cross section of 4.9 pb or greater was calculated to be 0.035%, which corresponds to a 3.4 $\sigma$ significance of this result.

ACKNOWLEDGMENTS

I would like to thank the organizers for an exciting and stimulating conference. And I would also like to express my sincere thanks to Linda Stutte, Lisa Shabalina and Yann Coadou for their corrections and suggestions after carefully reading the manuscript.

REFERENCES

3. The D0 Collaboration. FERMILAB-PUB-06/475-E. To be published in PRL. More information can be found in http://www-d0.fnal.gov/Run2Physics/top/public/fall06/singletop/