CDF/ANAL/TOP/PUBLIC/7102
Version 1.3
July 12, 2004

Top Mass Measurement in the Lepton + Jets Channel using a Multivariate Template Method

The CDF Collaboration
URL http://www-cdf.fnal.gov

Abstract

We report a measurement of the top quark mass obtained from a sample of 162 pb^{-1} of data with CDFII. We use top candidate events in the lepton plus jets channel. For each event energy-momentum conservation is applied to reconstruct the top quark mass. The multivariate template method employs a likelihood procedure that evaluates the probability for a certain top mass value based on multidimensional templates obtained from Monte Carlo. We report an analysis that uses events with jets tagged by a displaced secondary vertex algorithm, SECVTX. We find a mass value of $M_{top} = 179.6^{+6.4}_{-6.3}(stat) \pm 6.8(syst)\ GeV/c^2$.

Preliminary results for Summer 2004 Conferences

1 Introduction

We report a measurement of the top quark mass obtained with the CDFII detector [1] using a multivariate template method on top candidate events in the lepton+jets channel with one or more jets tagged as b-jets by the displaced vertex algorithm (SECVTX).

The "template method" used in Run I [2] employs a likelihood procedure to extract a top mass value from the reconstructed top masses of the candidate events. Both signal (at a number of top masses) and background reconstructed mass distributions from Monte Carlo are used in the likelihood. These distributions are called "templates".

We measure the top mass using $t\bar{t}$ events in which one W from top decays into a charged lepton + neutrino and the other into 2-jets. A fit is made for each event by using energy-momentum conservation, then a likelihood procedure is used to determine the top mass value. The method discussed here differs from the traditional "template method" in several ways. We attempt to reduce jet systematics by calibrating the jet energy scale using the W mass constraint event-by-event. We reduce statistical uncertainty by estimating the probability to pick the correct permutation on an event-by-event basis and reweight the events according to this probability. We improve signal/background separation by utilizing other kinematic variables in addition to the reconstructed top mass. Kernel density estimation is employed to make multivariate templates and local polynomial regression is used to ensure likelihood continuity.

2 Description of the method

2.1 Event fitting

We reconstruct the process

$$p\bar{p} \to t\bar{t} + X \to (Wb)(W^-\bar{b}) + X \to l\nu bq\bar{q}\bar{b} + X$$

In our top quark mass reconstruction we consider only those parton assignments to the four leading jets which are consistent with the b tagging information. For each parton assignment we attempt to improve the resolution by integrating the top mass over the W mass Breit-Wigner distribution. For each of the two W bosons in the event we use a set of seven m_W^2 values equidistant in the cumulative Breit-Wigner probability.

A jet energy scale factor, JES, is included in the W mass kinematic fit with a Gaussian width constraint, σ_{JES}, which is a tunable parameter. The momenta of all jets in the event (including non-leading jets) for a particular permutation are multiplied by the jet energy scale obtained from the mass fit of the hadronically-decaying W. Figure 1 shows that for the correct permutation, scale shifts due to the W mass constraint compensate on average for the systematic jet energy shifts. Thus the systematic uncertainty is reduced, whereas the statistical error on the mass measurement is increased by this method because of fluctuations in jet fragmentation.

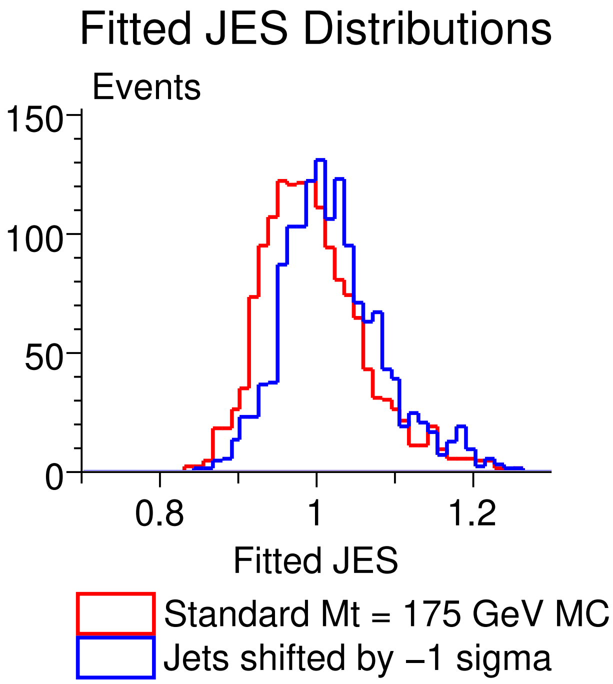


Figure 1: Distributions of the jet energy scale factor (JES) values for fits to Monte Carlo $t\bar{t}$ events. Red histogram: default energy scale. Blue histogram: energy scal shited by one $\sigma_{JES} = 0.035$.

The masses of the t and $\bar{t}$ quarks, m_t and $m_{\bar{t}}$, are then determined together with their errors, σ_{m_t} and $\sigma_{m_{\bar{t}}}$, and correlation coefficient, ρ. This procedure is carried out for each of the seven mass assignments for the leptonically-decaying W boson.

Jet, lepton, and unclustered energy measurement errors are propagated to t and $\bar{t}$ masses using the standard small error approximation formulae. Out of the two neutrino solutions we choose the one which results in a smaller value of $|m_t - m_{\bar{t}}|$. W mass assignments without neutrino solutions are discarded. Finally, we impose the $m_t = m_{\bar{t}}$ constraint on the measurement and calculate the top quark mass $M_{t,W}$, its error $\sigma_{M_{t,W}}$, and chi-square, $\chi^2_{t,W}$, for the top mass constraint as

$$M_{t,W} = \frac{\sigma^2_{m_{\bar{t}}} m_t + \sigma^2_{m_t} m_{\bar{t}} - \rho\, \sigma_{m_t} \sigma_{m_{\bar{t}}} (m_t + m_{\bar{t}})}{\sigma^2_{m_t} + \sigma^2_{m_{\bar{t}}} - 2\,\rho\, \sigma_{m_t} \sigma_{m_{\bar{t}}}} \tag{1}$$

$$\sigma^2_{M_{t,W}} = \frac{\sigma^2_{m_t} \sigma^2_{m_{\bar{t}}} \,(1-\rho)}{\sigma^2_{m_t} + \sigma^2_{m_{\bar{t}}} - 2\,\rho\, \sigma_{m_t} \sigma_{m_{\bar{t}}}} \tag{2}$$

$$\chi^2_{t,W} = \frac{(m_t - m_{\bar{t}})^2}{\sigma^2_{m_t} + \sigma^2_{m_{\bar{t}}} - 2\,\rho\, \sigma_{m_t} \sigma_{m_{\bar{t}}}} \tag{3}$$

The top mass for each jet to parton assignment is determined by summing over the points in the W mass integration grid weighted by the combined χ^2 of the mass constraints:

$$M_t = \frac{\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2} M_{t,W}}{\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2}} \tag{4}$$

We define the overall jet permutation chi-square as

$$\chi^2 = -2\log\left(\frac{1}{N}\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2}\right) \tag{5}$$

where N is the total number of points in the W mass integration grid (in the current analysis $N = 49$).

2.2 Handling of Combinatorics

The presence of four jets in the final state allows many permutations for the assignment of a jet to a parton from the $t\bar{t}$ process. Since we are using events with ≥ 1 tagged b jets in this analysis, at most six jet to parton assignments are possible.

We split our signal MC events into three subsamples:

1. Events in which the four leading jets correspond to the four daughter quarks in the $t\bar{t} \to l\nu bq\bar{q}\bar{b}$ process, and the jet permutation with the smallest χ^2 results in the correct jet to parton assignment. We call this GP (good permutation) subsample.

2. Events in which the four leading jets correspond to the four daughter quarks in $t\bar{t} \to l\nu bq\bar{q}\bar{b}$, but the smallest χ^2 permutation does not result in the correct jet to parton assignment. This is the BP (bad permutation) subsample. Later we will refer to the union of GP and BP subsamples as the CJ (correct jets) subsample.

3. Events in which one or more leading jets do not come from $t\bar{t} \to l\nu bq\bar{q}\bar{b}$. This subsample includes all $t\bar{t}$ MC events which do not belong to the first two subsamples. This is the IJ (incorrect jets) subsample.

The reconstructed mass distribution for these subsamples is illustrated in Figure 2 for 180 GeV/c^2 generated mass and jet energy scale constraint value 0.07. The dependence of the medians of these distributions for several values of the generated mass are also shown. The slope for the correct combination is one, as desirable.

These plots show that knowledge of the subsample to which each event belongs is crucial to the resolution of the reconstructed mass distribution. Perfect $t\bar{t}$ subsample separation can not be achieved in practice. Instead, to each data event we assign probabilities to belong to the different subsamples. We employ a model in which we consider differences between χ^2 values of the best permutation and all other permutations, as well as kinematic properties of $t\bar{t}$ events.

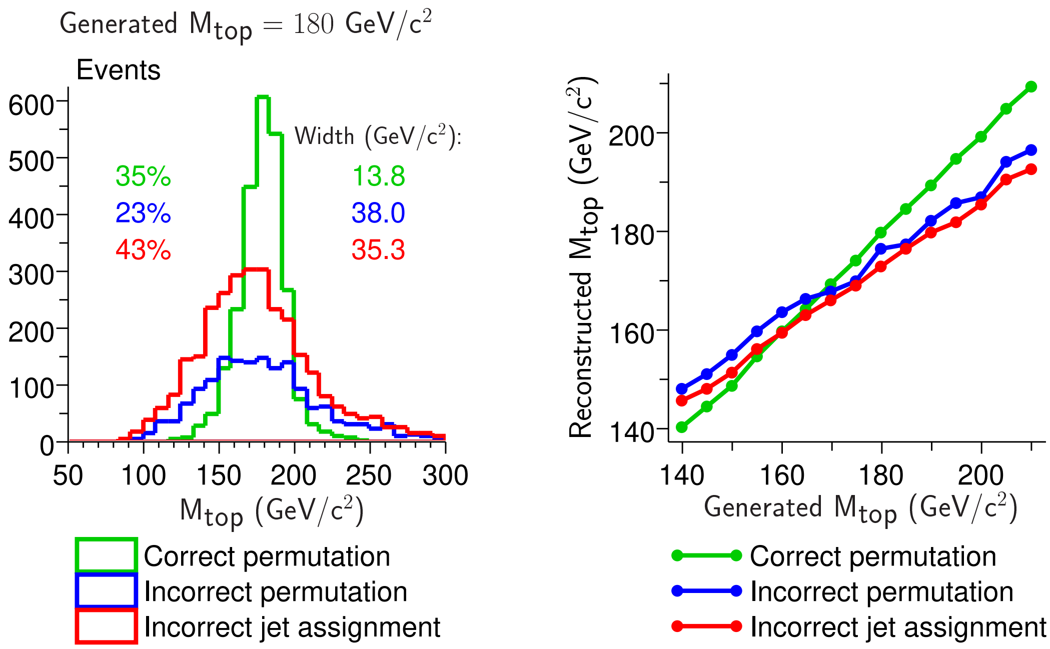


Figure 2: Left: reconstructed top mass distribution for the generated mass of 180 GeV/c^2 Right: median reconstructed mass vs generated mass. In both plots results for the three signal subsamples discussed in the text are shown. The slope for the GP subsample is 1.00, whereas the average slopes for the BP and the IJ subsamples are 0.67 and 0.68 respectively.

The fraction of incorrectly assigned permutations appears to decrease exponentially as a function of the difference between permutation χ^2 values. Therefore, we make the initial GP probability assignment assuming that the jets were correctly chosen as

$$P(GP|CJ) = \frac{a_b}{\sum_{i=1}^{12} a_i \exp\left(-\frac{\chi_i^2 - \chi_b^2}{w_e(\chi_i^2 + \chi_b^2)}\right)}, \tag{6}$$

$$\text{where} \tag{7}$$

$$w_e(y) = \exp(b_e + c_e y + d_e y^2) \tag{8}$$

Coefficient a_i models the event density in the correct region for jet permutation number i. Index b denotes the permutation with the smallest χ^2. Index e the function $w_e(y)$ refers to the error type assigned to the permutation with index b when the correct permutation is the one with index i. The constants in the above formulae (a_i, b_e, c_e, and d_e) are determined from the top signal Monte Carlo.

We also include additional kinematical information in order to improve the separation between the correct permutation sample and the other subsamples. This information is included by sequential application of the "approximate Bayesian update" formula. Suppose,

X is some kinematic variable whose value may be used to improve separation between the GP and BP subsamples. Then formula

$$P(GP|CJ, X) = \frac{\kappa P(GP|CJ)}{\kappa P(GP|CJ) + (1 - P(GP|CJ))} \tag{9}$$

follows from the Bayes theorem when $\kappa = \frac{P(X|GP)}{P(X|BP)}$ and $P(GP|CJ) + P(BP|CJ) = 1$. Rather than calculating κ from MC, we treat it as a decision surface to be optimized together with the constants in the initial $P(GP|CJ)$ assignment.

The first kinematic variable used to update $P(GP|CJ)$ is $\cos(\varphi)$ where φ is the angle between the lepton l and b_l (the b associated with the leptonic top) in the rest frame of the W which decays into $l\nu$. Distributions of this angle are shown in Figure 3 for both the correct and the incorrect permutations. Thus we try to improve the subsample separation near $\cos(\varphi)$ =-1 where the correct permutation is suppressed by the top decay matrix element.

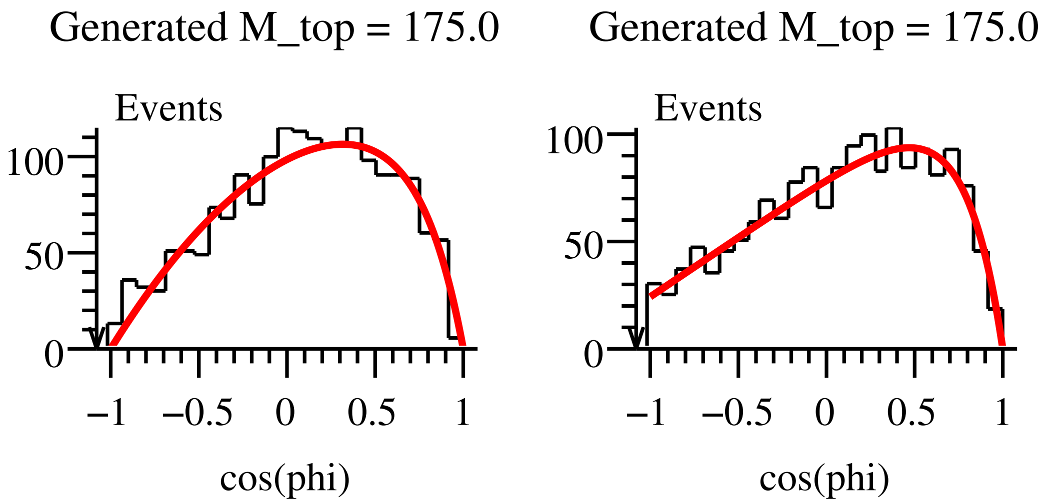


Figure 3: Distributions of the angle between l and b_l in the frame of the W which decays into $l\nu$ for the GP subsample one b tag. The function overlayed top of each plot has the same shape for each generated mass, only the height is adjusted. Left: correct permutations; right: incorrect permutation.

The second kinematic variable used to update $P(GP|CJ)$ is the product $\cos(\theta_1)\cos(\theta_2)$. θ_1 is the angle between the lepton momentum and the beam axis in the rest frame of the top quark which decays leptonically. θ_2 is the angle between the direction of the light quark and the beam axis in the rest frame of the top quark which decays hadronically. We use that light quark which, in the rest frame of the parent W, has smaller angle with the b quark originating from the hadronically decaying top. The $\cos(\theta_1)\cos(\theta_2)$ product is sensitive to the $t\bar{t}$ spin correlations [3]. The decision surface for this variable is modeled as

$$\kappa_{\rm spincorr} = D\,(1 + E\cos(\theta_1)\cos(\theta_2)) \tag{10}$$

where D and E are constants to be fitted.

The distributions of the correct permutation probabilities thus obtained from $t\bar{t}$ MC events is given by the top two histograms shown in Figure 4. The top left plot contains events with exactly one b tag, the top right plot contains events with two b tags. The latter plot shows that tagging information is obviously very useful to increase the correct permutation probability.

In order to check the reliability of the modeling of the correct permutation coefficients, we compare the predicted probabilities of the fit (which include samples at top masses M=140-210 GeV/c^2 in 5 GeV/c^2 bins), with the fractions of correct permutations found in $t\bar{t}$ MC events at M=177.5 GeV/c^2 (this sample was not included in the fit). The two bottom plots in Figure 4 show these comparisons for events with one and two b tags, respectively. The agreement between predicted and observed fractions is satisfactory.

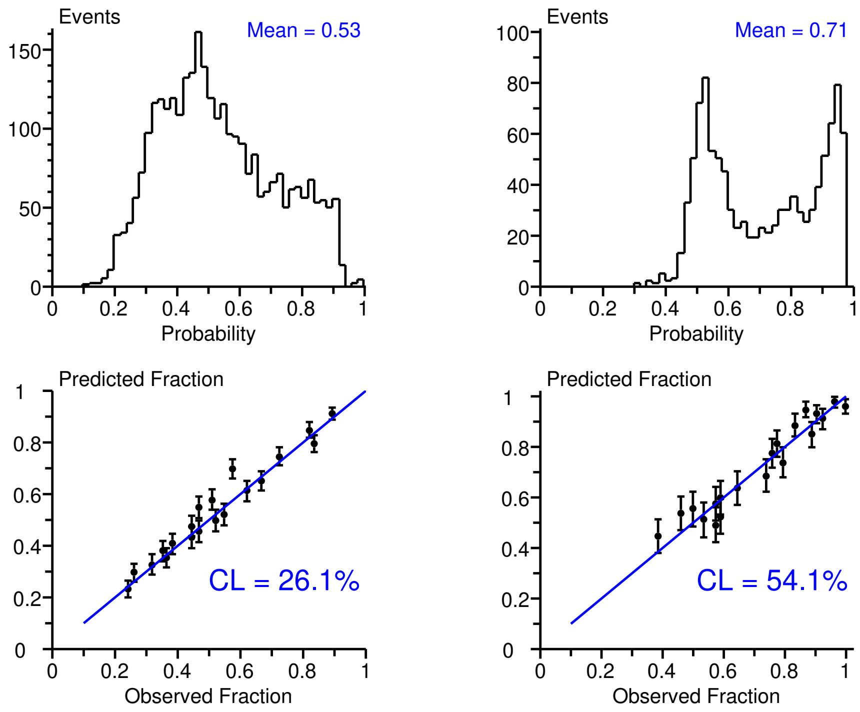


Figure 4: Distribution of correct permutation probabilities for one b-tag events (left) and two b-tag events (right). The bottom plots show the predicted probabilities vs the fraction of correct permutations expected from $t\bar{t}$ MC at a mass of 177.5 GeV/c^2. The lines shown are the 45^0 diagonals, not the result of a fit.

2.3 Signal-Background discrimination

For the multivariate template method we seek a set of "good" variables which both increase the sensitivity of the likelihood to the top quark mass and improve the discrimination between signal and background. The set of variables should also be optimal w.r.t. the systematic uncertainties of the measurement.

To make a reasonable pre-selection among all the possible kinematic and angular/shape variables (like acoplanarity) we evaluated various histogram distance measures which quantify the differences of probability density distributions.

Given two probability distributions $p_1(x)$, $p_2(x)$ of a variable x, if x is discrete in n bins, one can calculate the quantity

$$d(p_1, p_2) = \sum_{i=1}^{n} p_1(x_i) \log \frac{p_1(x_i)}{\alpha \cdot p_2(x_i) + (1-\alpha) \cdot p_1(x_i)} \tag{11}$$

where α is a free parameter. For a value of $\alpha = 0.5$, as proposed in [4], we evaluate a measure referred to as *K divergence*. To obtain a symmetric distance measure, we define

$$K = d(p_1, p_2) + d(p_2, p_1) \tag{12}$$

For comparison we also calculate the *Kolmogorov-Smirnov* distance KS, basically the maximum distance between the cumulative distributions of $p_1(x)$ and $p_2(x)$.

Figure 5 shows the numerical values of the distances between the signal, $t\bar{t}$ MC (HERWIG), and the W+4-jets background MC (ALPGEN+HERWIG) for the various event observables shown in Table 1. These, and also other distances measures, lead to similar conclusions: kinematic energy quantities discriminate signal from background better than angle or shape variables, whereas pure leptonic energy variables have almost no discriminative power, when the background is due to W events. We found that the sum of the transverse energies of the four leading jets is the top performer. The invariant top mass has only average discriminative power to distinguish signal from background.

2.4 Templates

For the construction of the templates we have chosen a nonparametric multivariate density reconstruction method called *Kernel Density Estimation* (KDE) [5]. The density at any given point in the multivariate space receives contributions from Monte Carlo events in such a way that events are weighted more heavily when they appear closer to the desired point. The distances between points are determined by the metric which is chosen as the inverse of the robust sample covariance matrix. In order to reduce the effect of outliers on the metric, we calculate the sample covariance matrix using the robust minimum covariance determinant (MCD) algorithm [6]. The MCD breakdown point was set at 25%.

We use a radially symmetric kernel function that behaves like a Gaussian probability

Table 1: Variables used for distance measures.

kinematic variables	
P_T^i $(i = 1 \dots 4)$	P_T of the four leading jets
$P_T^{i+j} \equiv P_T^i + P_T^{i+1}$ $(i = 1 \dots 3)$	Sums of P_T's
$P_T^{1-3} \equiv \sum_{i=1}^{3} P_T^i$	Sum of the three leading jets P_T's
$P_T^{1-4} \equiv \sum_{i=1}^{4} P_T^i$	Sum of the four leading jets P_T's
$E_T(\ell\pm)$	E_T of lepton
$\not{E}_{\mathrm{T}}$	Transverse Missing E_T
$H_T \equiv \sum_{jets} p_T(i) + E_T(\ell^\pm) + \not{E}_T$	Scalar sum of jets+lepton P_T+$\not{E}_{\mathrm{T}}$
P_T^{top}, $P_T^{t\bar{t}}$	Top and $t\bar{t}$ P_T calculated from the fit
M_{ttbar}	$t\bar{t}$ mass calculated from the fit
M_{ij}^{min}	Min di-jet mass among the 4 leading jets
Angular Variables	
$(\sum_{i=1}^{4} \lvert P_z(i)\rvert + P_z(l^\pm)/(\sum_{i=1}^{4} P_T(i) + P_T(\ell^\pm)$	Ratio of sum of long. and transverse P_T's
$R^{min} \equiv \Delta R_{ij}^{min} \times \min(P_z^{(i,j)})/P_T(\ell^\pm)$	Min di-jet separation $\times$ P_z(min jet) over $P_T(\ell)$
α^*	Leptonic top: ℓ and b angle, in $l^\pm\nu$ rest frame
Shape Variables	
$T = \max_{\vec{n}}(\sum_i \lvert\vec{p_i} \cdot \vec{n}\rvert / \sum_i \lvert\vec{p_i}\rvert)$. $\vec{n}$	Thrust
$S = \frac{3}{2}(\lambda_1 + \lambda_2)$	Sphericity
$Aco = 4\min(\sum_i \lvert\vec{p_i}\rvert^{out} / \sum_i \lvert\vec{p_i}\rvert)^2$. $\vec{p_i}^{out}$	Acoplanarity

density near the origin but decays slower at infinity. This function is defined by

$$K_H(r^2) = Norm(a, d) \times \begin{cases} e^{-r^2/2} & \text{if } r \leq a \\ e^{-ra+a^2/2} & \text{if } r > a \end{cases} \tag{13}$$

where $Norm(a, d)$ is the normalization term which depends on parameter a and dimensionality of the variable space. In this manner we could ensure that outlying points would contribute to the likelihood value but not have such low density that they would unduly affect the likelihood peak position. The value of parameter a is always chosen so that the Gaussian part of the distribution contains 90% of the density. We choose a global bandwidth which minimizes the asymptotic mean squared integrated error (AMISE) of the density estimator for a sample of points drawn from the multivariate Gaussian distribution.

Figure 6 is an illustration of 2-dimensional templates based on the invariant top mass and the scalar sum of the transverse momenta of the first four leading jets.

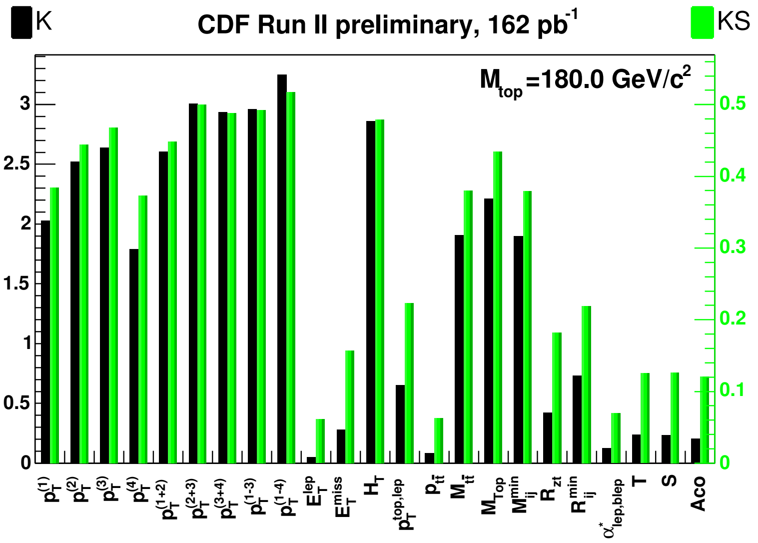


Figure 5: Distance between the signal MC (HERWIG) and the W+4-jets background for various event observables. Left scale (black bars): the symmetrized K divergence. Right scale (green bars): the KS distance.

2.5 Likelihood

We define the likelihood of the observed data sample as

$$\mathcal{L}(m_t) = \prod_{i=1}^{N} \left(f_b P_b(m_i, x_i) + (1 - f_b) P_s(m_i, x_i, m_t) \right) \tag{14}$$

where N is the number of observed events, m_i is the top mass in the ith event (determined for the jet permutation with lowest χ^2), x_i symbolizes all template variables besides measured mass, P_s and P_b are the signal and background densities in the (m, x) space, f_b is the background fraction, which is allowed to float freely. The signal density is composed as follows:

$$P_s(m, x, m_t) = \left(p_{\mathrm{GP}} S_{0,m_t}(m, x) + (1 - p_{\mathrm{GP}}) S_{1,m_t}(m, x) \right) p_{\mathrm{CJ}} + (1 - p_{\mathrm{CJ}}) S_{2,m_t}(m, x) \tag{15}$$

S_{0,m_t}, S_{1,m_t}, and S_{2,m_t} are the three signal templates for the given generated m_t. p_{GP} is just a short notation for $P(GP|CJ, X)$, which is an extension of $P(GP|CJ)$ defined in Eq. 6 to accomodate additional variables X. p_{CJ} is the probability to pick correct jets and is evaluated by Monte Carlo. We model the background density as

$$P_b(m, x) = \sum_{\text{bg types}} a_j B_j(m, x), \qquad \sum_{\text{bg types}} a_j = 1 \tag{16}$$

where B_j are the templates for different background types, and a_j are the background composition coefficients obtained from the $t\bar{t}$ cross section studies [7].

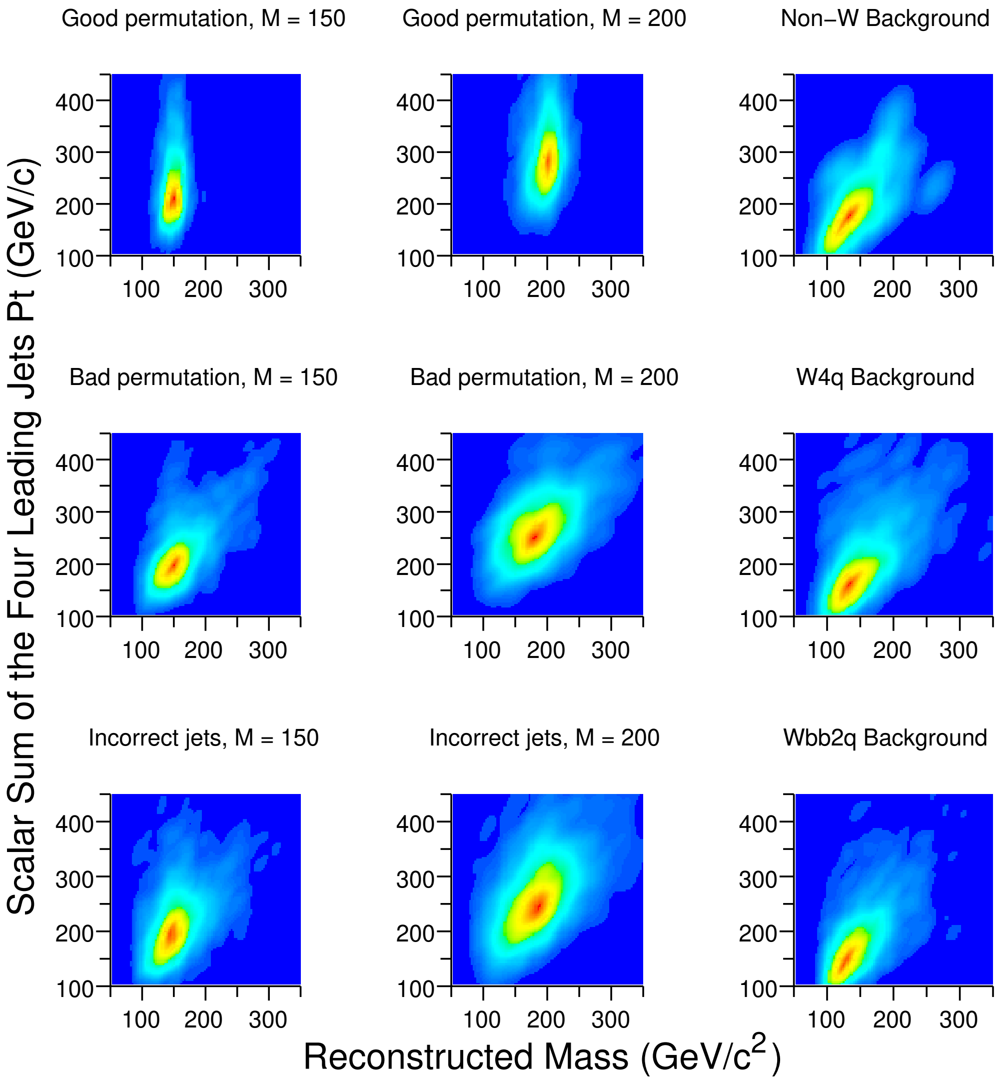


Figure 6: Two-dimensional templates obtained using the reconstructed mass and the scalar sum of the four leading jet p_T values. The first and the second column shows the templates for m_t=150 GeV/c^2 and 200 GeV/c^2, respectively. The top row is for events with the correct jet combination, the second row is for combinations that use the correct jets but not the correct assignment, the last row is for events for which one or more of the 4 jets are not decay products of the two top candidates. The templates for the three types of backgrounds are shown in the rightmost column. The color maps are not synchronized between the plots.

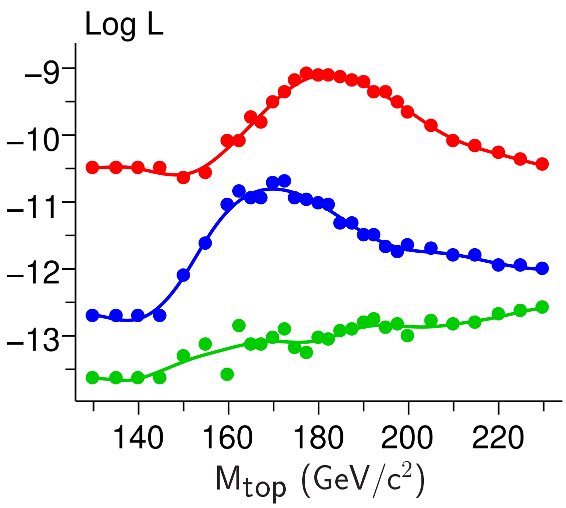


Figure 7: Smoothing and interpolation of the likelihood with local polynomial regression for three random Monte Carlo events. Raw likelihood values are shown with circles. Continuous lines display regression results.

The likelihood in Eq. 14 is not normalized. Instead of trying to impose an explicit normalization condition, we note that the normalization factor would be unnecessary if p_{CJ} and p_{GP} probabilities were completely independent from m and x. We maintain an approximate normalization by estimating p_{CJ} and p_{GP} in a manner which introduces very little dependence on m. However, we do expect that, in general, the width of the pull distribution obtained from pseudo-experiments will not be equal to 1, and the pull mean will not be exactly 0.

The likelihood in Eq. 14 is defined only for a discrete set of m_t values for which signal templates were generated. We expect that dependence of the likelihood on m_t should be continuous and smooth for each event. However, our signal templates do not guarantee likelihood continuity because each template is generated using an independent set of Monte Carlo events with finite statistics. To introduce the assumption of continuity and smoothness into the likelihood formula, we use local quadratic polynomial least squares regression with m_t as a predictor and $\log\mathcal{L}$ as a response. Interpolation is performed for each event separately, and resulting log likelihood curves are then added to obtain the likelihood of the whole sample. Polynomial regression calculations are performed using routines DPOLFT and DP1VLU from the SLATEC software package [8]. At each interpolation point m, template likelihoods are weighted in the polynomial fit according to the formula $w_i = (1 - ((m_{t,i} - m)/h)^2)^4$ if $|m_{t,i} - m| < h$, and $w_i = 0$ if $|m_{t,i} - m| \geq h$. The bandwidth h is set to 25 GeV/c^2. The effect of local regression is illustrated in Fig. 7.

3 Data Set and Background

Top event candidates are required to have a lepton, electron (muon) with $E_T(P_T) \geq 20$ GeV, $\not{E}_T \geq 20$ GeV, and four jets. If the event has more then four jets, the four jets with the highest transverse energy (E_T) are considered for event fitting. $E_T > 15$ GeV is required for the first three, $E_T > 8$ GeV for jet 4. Events with the fourth jet in the 8-15 GeV range are called 3.5-jet events. At least one jet tagged as a bottom jet using a displaced vertex algorithm is required. Backgrounds from non-top events come from several sources: from lepton misidentification (non-W, QCD events), from W $b\bar{b}$ events, from W+jets events for which a light quark jet has been tagged as a b-jet (mistag), from W or Z events which have the same topology as top events and from single top production events. The background rates are derived from those obtained for the $t\bar{t}$ cross section measurement done with the SECVTX ≥ 1 b-tagged jets sample [7]. Table 2 shows the results for 162 pb^{-1} of data: the number of events with a mass fit ($\chi^2 < 10$), and the expected background events for each category. The number of fitted events were obtained with a JES constraint of 0.07, which is the result of the optimization discussed in the next section.

Table 2: Summary of backgrounds for the 3.5 jets, the ≥ 4 jets and total sample (≥ 3.5 jets). The number of events with a mass fit were obtained for a value of the JES constraint of 0.07 and a $\chi^2 < 10$.

Background	W+3.5jets	W+$\geq$4jets	Total
mistags	0.90±0.15	0.85±0.13	1.75±0.28
W $b\bar{b}$, W $c\bar{c}$	0.83±0.29	0.60±0.22	1.43±0.51
Wc,WW/WZ etc.	0.41±0.13	0.17±0.07	0.58±0.20
Single top	0.21±0.04	0.09±0.02	0.30±0.06
Sum of above 3	1.45±0.34	0.86±0.24	2.31±0.58
non-W (QCD)	1.24±0.22	1.32±0.25	2.56±0.47
Total Backg	3.59±0.43	3.03±0.37	6.62±0.80
Have a mass fit	10	23	33

4 Optimization

We optimize the JES constraint parameter by generating many pseudo-experiments for the number of observed data events, N, having a mass fit. We draw N events from the $t\bar{t}$ Monte Carlo templates and from the background templates (three different ones: mistag, non-W and the W+ $b\bar{b}$+2jets Monte Carlo for the other sources of background). The total number of background events (34% of the 32 fitted events) is obtained from a likelihood fit of the data using a JES constraint, σ_{JES}, of 0.05.

The optimization is done using 2-dimensional templates, one variable being the reconstructed top mass, the other chosen among the variables discussed in Sec. 2.3. Four choices for the second variable are shown in the plots of Figure 8. The top plot shows the statistical error versus the JES constraint parameter. The middle plot show the expected systematic uncertainty due to systematic uncertainties on jets momenta. The lower plot shows the two uncertainties added in quadrature. The vertical bars show the expected uncertainties for one pseudo-experiment for the four cases, at each value of the JES constraint. At high values of JES constraint, top mass alone results in a smaller systematic error than other variable sets because it is less sensitive to jet transverse energy cuts applied in the event selection procedure. The four choices for the second variable have similar performance, therefore it is difficult to choose one with the present statistics. Also, from Figure 8 it is not clear that we need to use two dimensional templates, however, based on the studies discussed in Sec. 2.3, we choose the sum of the transverse momenta of the four leading jets. This variable provides the best discrimination between signal and background. A value of the JES constraint of 0.07 appears to be optimal.

5 Results

The likelihood procedure applied to the data (using a JES constraint of 0.07) provides a mass value and a background fraction as a function of the top mass. Figure 9 shows on the left the negative log-likelihood curve as obtained for our data sample, and on the right the background as a function of the mass as obtained from the likelihood procedure. Figure 10 shows the median and width of the pull distributions as a function of the generated mass as obtained from pseudo-experiments. The pull width on the average is 1.10 and the bias is small. We multiply the statistical error, as obtained from the likelihood fit, by a factor 1.10. The expected statistical error (multiplied by the factor 1.10) is shown on the right in Figure 10. The arrows show the error value obtained from the data.

The results of the fit for the top mass and for the statistical error of the top mass is:

$$M_{top} = 179.6\ ^{+6.4}_{-6.3}\ \ GeV/c^2$$

From the background fraction versus top mass plot shown in Figure 9 we extract the most probable background fraction by interpolating between the two templates closest to the obtained mass. The result for the most probable background fraction and for its statistical error is

$$f_b = 0.34 \pm 0.14\ \%$$

Figure 11 below shows the mass distribution for the 33 events reconstructed with a JES constraint of 0.07. The points are the data, the other distributions are for the different backgrounds and the $t\bar{t}$ Monte Carlo. It also shows the JES values obtained for each

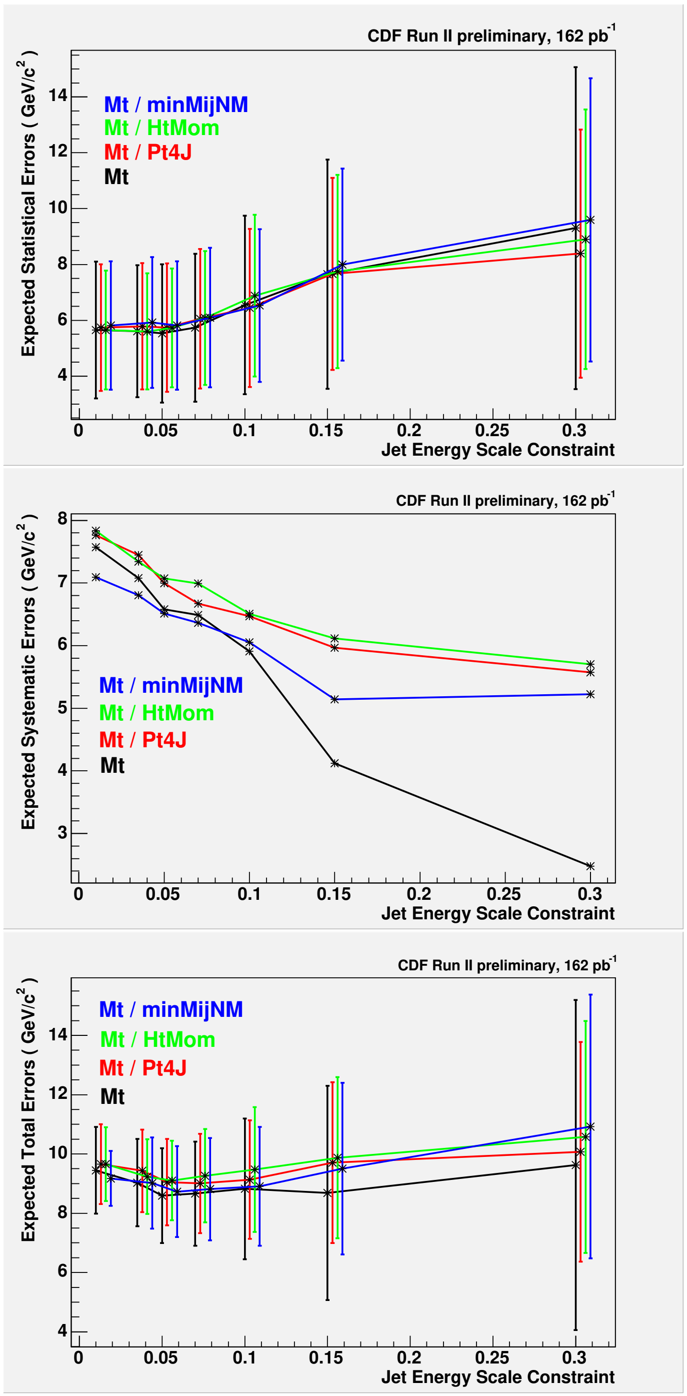
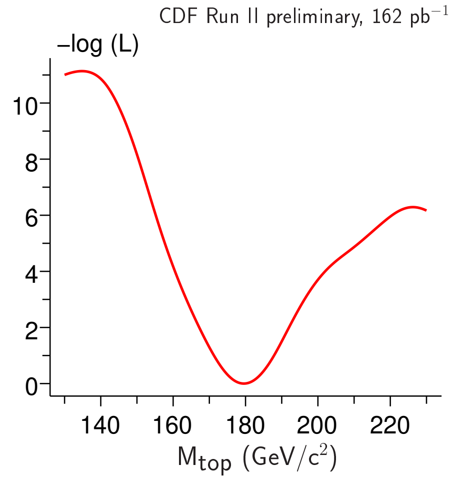


Figure 8: Expected top mass statistical (top), systematic from jet energy uncertainty (middle) and total (bottom) mass uncertainty as a function of the JES constraint in the fitting procedure. Four different sets of variables are shown. Vertical bars show the range of expected uncertainty from one pseudo-experiment.



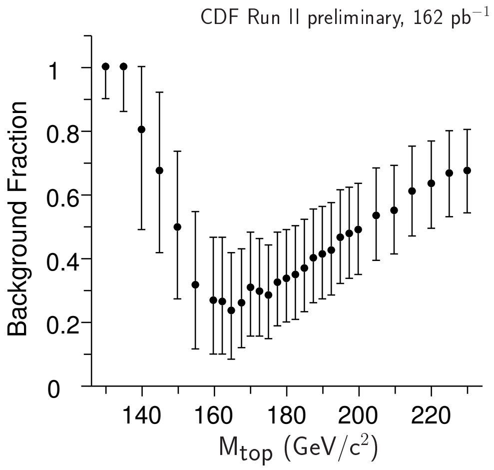
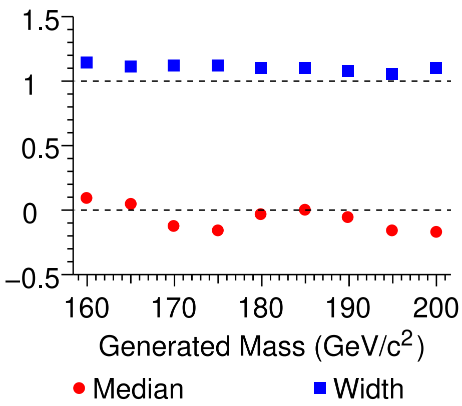


Figure 9: Top mass likelihood from data in the $\geq$ 3.5 jets sample. The background as a function of the reconstructed top mass is shown at right.



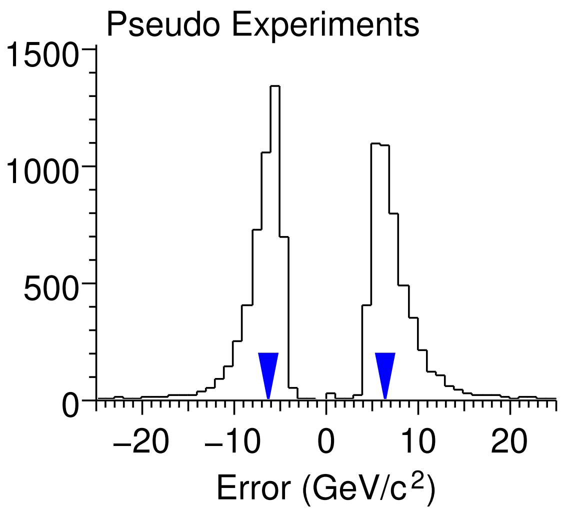


Figure 10: Left: parameters of the reconstructed pull distributions in the $\geq$ 3.5 jets sample. Width is defined as half of the distance between 16$^{\text{th}}$ and 84$^{\text{th}}$ percentiles. Right: distribution of the expected top mass statistical errors from pseudo-experiments in the $\geq$ 3.5 jets sample. The horizontal scale has been widened according to the pull width. Arrows show the observed error values.

event for the fit with a JES constraint of 0.07. The scatter plot on Figure 12 shows the sum of the transverse momenta of the four leading jets versus the reconstructed top mass, superimposed to the expectation from Monte Carlo. The squares represent the data.

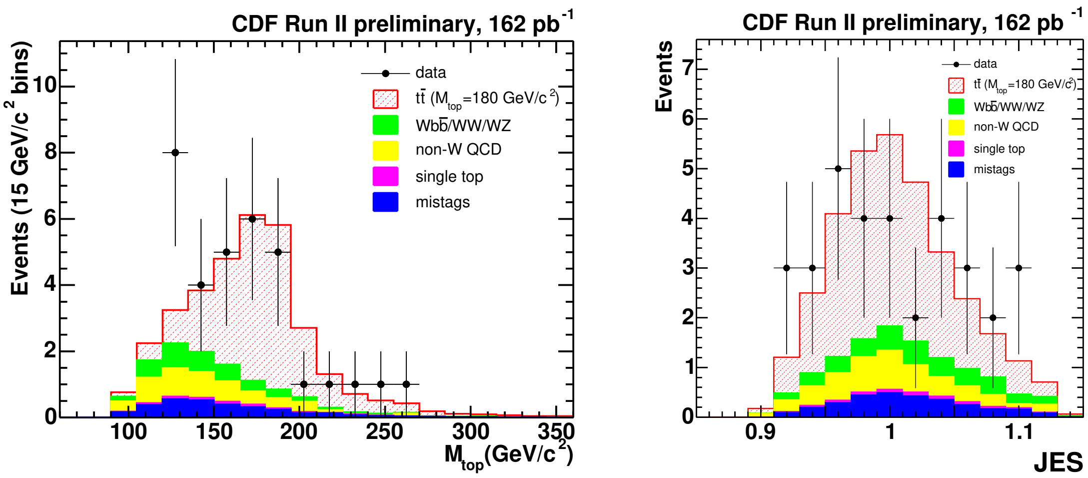


Figure 11: Left: Top mass distribution for the 33 events in the ≥ 3.5 jets data sample. The background fraction is set to 34%. Right: the JES values obtained for the data and Monte Carlo when the constraint was set to 0.07.

Systematic uncertainties shown in Table 3 have been determined from many pseudo-experiments of 33 events each, by modifying the default values for a given parameter by one standard deviation and evaluating a mass shift. The contribution to the systematic uncertainties on the jet energy are the largest ones. The other listed sources contribute little when they are added in quadrature.

6 Conclusions

The top quark mass measured by the multivariate templates method in the ≥ 3.5 jets sample is

$$M_{top} = 179.6^{+6.4}_{-6.3}(stat) \pm 6.8(syst)\ GeV/c^2.$$

This is the first CDF top mass measurement with freely floating background fraction. Our method allows for optimal trading of statistical and systematic errors, and will be especially powerful as the CDF accumulated luminosity increases.

7 Acknowledgments

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

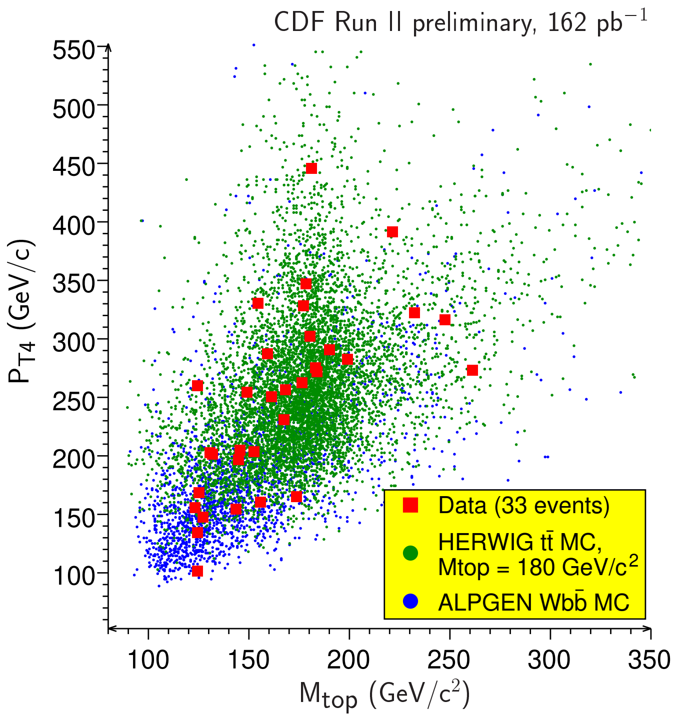


Figure 12: Scatter plot of the two variables used in the likelihood to measure the top mass: sum of the P_T of the 4 leading jets in the event and the reconstructed top mass. Dots are the background (blue) and the top signal (green) Monte Carlo, the squares are the data.

Table 3: Summary of top mass systematic errors.

Systematic	ΔM_{top} (GeV/c^2)
Jet Energy	6.7
Generators	0.2
ISR	0.2
FSR	0.6
PDF	0.6
Background Shape	0.4
b Tagging	0.3
Fitting Procedure	0.7
Total	6.8

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 fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

References

[1] F. Abe, et al., Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988); D. Amidei, et al., Nucl. Instum. Methods Phys. Res. A **350**, 73 (1994); F. Abe, et al., Phys. Rev. D **52**, 4784 (1995); P. Azzi, et al., Nucl. Instrum. Methods Phys. Res. A **360**, 137 (1995); The CDFII Detector Technical Design Report, Fermilab-Pub-96/390-E

[2] T. Affolder et al. *Measurement of the Top Quark Mass with the Collider Detector at Fermilab*, Phys. Rev. D, **63**, 032003 (2001).

[3] G. Mahlon and S. Parke, *Maximizing spin correlations in top quark pair production at the Tevatron*, Physics Letters B **411**, 173 (1997).

[4] J. Lin, *Divergence Measures Based on the Shannon Entropy*, IEEE Transactions on information theory, vol 37 no. 1 (1991) 145.

[5] D. Scott, *Multivariate Density Estimation : Theory, Practice, and Visualization* Wiley-Interscience, 1992.

[6] P.J. Rousseeuw, K. Van Driessen, *A Fast Algorithm for the Minimum Covariance Determinant Estimator*, Technometrics, **41**, 212.

[7] H. Bachacou, J. Nielsen and W-M Yao *Optimized Measurement of the $t\bar{t}$ Cross Section in the SECVTX Tagged Sample*, CDF-6902 (March 2004).

[8] The SLATEC Common Mathematical Library is a collection of Fortran subroutines for a wide variety of mathematical problems. It is available from http://www.netlib.org.