CDF note 8420

Measurement of the top mass in the all-hadronic channel using the Template Method with 1.02 fb^{-1}

The CDF Collaboration

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

Abstract

We present here the top mass measurement using Template Method in the all-hadronic channel, i.e. where both W's decay into a $q\bar{q}$ pair. The analysis is performed using 1.02 fb^{-1} of $p\bar{p}$ collisions collected with a multijet trigger at $\sqrt{s} = 1.96\ GeV/c^2$ with the Collider Detector at Fermilab. The method relies on a comparison of invariant mass distributions in data, taken after a neural network selection, to signal Monte Carlo and data-driven background events to extract the mass through an unbinned likelihood technique. The measurement gives as top mass $m_{top} = 174.0 \pm 2.2(stat) \pm 4.8(syst)\ GeV/c^2$.

1 Introduction

At Fermilab, top quarks are mainly pair produced in $p\bar{p}$ collisions via $q\bar{q}$ annihilation (85%) and gg fusion (15%). In the framework of Standard Model, top quarks decay into W bosons and b quarks with $BR \approx 1$.

In this analysis we search for events in which both W bosons coming from top quark decays give rise to quark-antiquark pairs, leading to an all-hadronic final state. This channel has the advantage of a large branching ratio, about 44%, and of fully reconstructed kinematics. The major disadvantage is the huge background from QCD multijet production which dominates the signal by three order of magnitude even after the application of a specific multijet trigger. A tight kinematical selection followed by the request of identified b-jets is thus needed in order to further improve the S/B,

The technique used here is based on a comparison trough likelihood maximization of invariant mass distribution for data with the expected signal and background probability density functions; the former is derived from Monte Carlo simulated samples, the latter from a data-driven background.

2 Event selection

An exhaustive explanation of the event selection is described elsewhere [1]. Here we summarize the essential features.

The multi-jet dataset we use, amounts to an integrated luminosity of about 1.02 fb^{-1}. All data/signal samples have to pass some prerequisites which require the event to have a well centered primary vertex, no tight leptons and no significant missing transverse energy. We require between 6 and 8 tight jets ($E_T \geq 15$ GeV, $|\eta| \leq 2.0$) and $\Delta R_{min} \geq 0.5$ for the minimum distance between jets.

The background after the b-tagging application is evaluated using a data-driven estimate based on a tag rate parametrization. A tag matrix derived in a control sample of events with 4 jets is then applied to events with a different jet multiplicity. This technique works because of the parametrization of the matrix in terms of the most important quantities like jet-E_T, jet track multiplicity, and number of primary vertices.

2.1 A neural net for kinematical selection

The neural net chosen here is the Multilayer perceptron (MLP[2]). This is a simple feed-forward network with 11 input layers, 2 hidden layers with 20 and 10 nodes respectively, and an output layer. The 11 inputs are

1. total transverse energy, $\sum E_T$
2. sub-leading transverse energy, $\sum_3 E_T = \sum E_T - E_T^1 - E_T^2$
3. centrality, $C = \sum E_T / \sqrt{\hat{s}}$
4. aplanarity, A
5. the minimum dijet mass, M_{2j}^{min}
6. the maximum dijet mass, M_{2j}^{max}
7. the minimum trijet mass, M_{3j}^{min}
8. the maximum trijet mass, M_{3j}^{max}
9. $E_T^{1\star} \equiv E_T^1 \cdot \sin^2 \theta_1^\star$
10. $E_T^{2\star} \equiv E_T^2 \cdot \sin^2 \theta_2^\star$
11. $\langle E_T^\star \rangle_{3N}$

where $\theta^\star$ is the cosine of the polar angle of the jet as measured in the all-jets rest frame. The last variable represent the geometric average of the $E_T^\star = E_T \cdot \sin^2 \theta^\star$ quantities over all jets but the first two (i.e. the 2 highest-E_T jets).

The performance of this neural net is described in Fig. 1 where we plot the signal-to-background ratio and the ratio between signal and expected candidates as a function of the cut on the neural net output.

The distributions of the neural network output, NN_{out} from now on, are shown in Fig. 2 for multi-jet events and the $t\bar{t}$ signal.

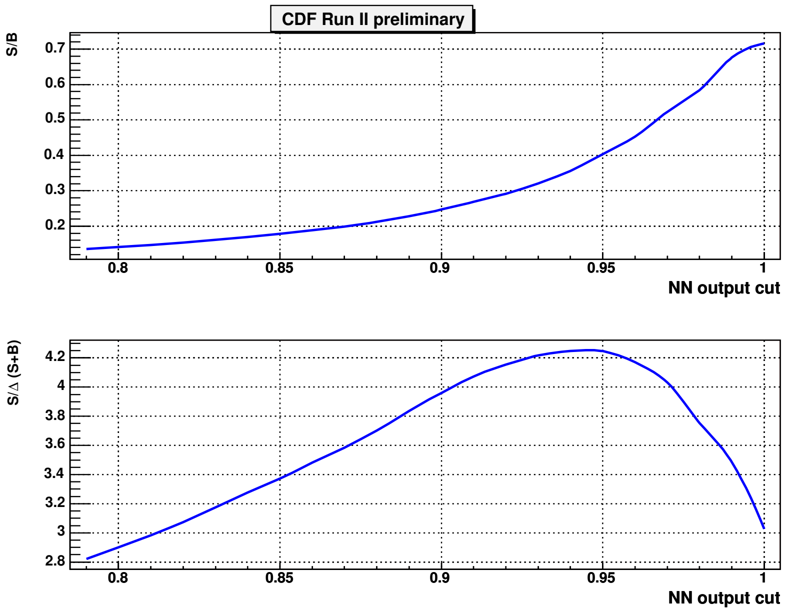

Figure 1: S/B and $\Delta(B+S)/S$ as a function of the cut on NN_{out}. Events with 6 to 8 jets and $\Delta R \geq 0.5$.

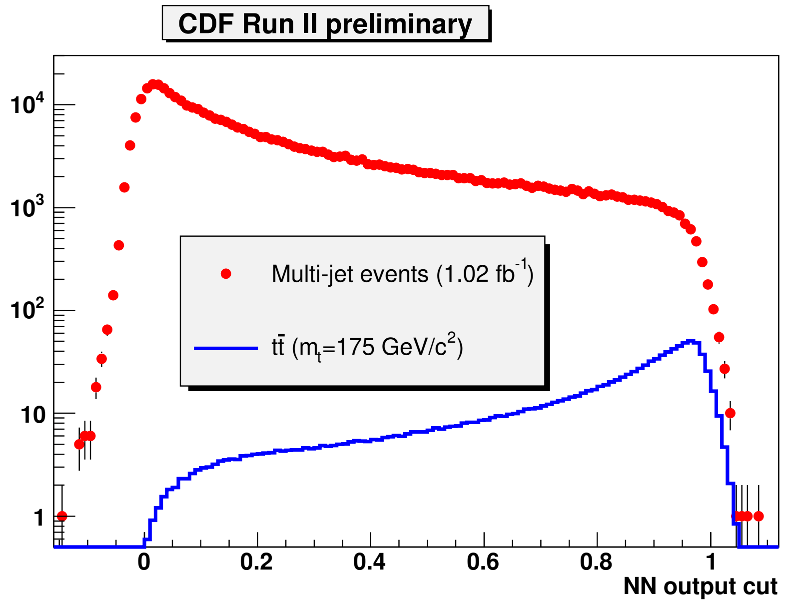

Figure 2: NN_{out} distribution for the whole samples ($\sigma_{t\bar{t}} = 6.7$ pb). Events with $6 \leq N_{jet} \leq 8$ and $\Delta R_{min} \geq 0.5$).

3 Background Modeling

The data-driven method used for the background estimate in the all-hadronic channel is based on the parametrization[1] of the jet tagging probability $\mathcal{P}$, evaluated on a signal-depleted sample of multi-jet events with exactly 4 jets and $\Delta R_{min} \geq 0.5$. The number of tags expected from background processes is then the sum over all events and all taggable jets of these probabilities. This background needs however to be corrected iteratively for the presence of $t\bar{t}$ events in the pre-tag sample.

We then produce a mass distribution for the background by running a kinematic fitter on each taggable jet in events passing the neural net cuts, but before the b-tagging request (pre-tag sample). Weighting each entry for the corresponding tag rate probability we normalize this distribution to the background expectation. The presence of signal in the pre-tag sample will lead not only to higher expectation, but will also modify the background mass distribution. Thus we apply the tag rate probability to taggable jets in signal events (with $m_t = 175\ GeV/c^2$) and subtract the resulting distribution, normalized to the tag rate expectation on signal events.

4 Kinematic Fitter

In this section we describe the procedure adopted to reconstruct a mass associated to the events surviving the cuts; we want to note here that this mass is not the event-by-event measurement, but serve as a top mass dependent quantity, which we will use to compare real data to simulated samples in order to extract the mass. The procedure is to append a mass to each tag that we find in the events surviving the neural net cut. We assume each event to be a $t\bar{t}$ event decaying into jets. We therefore combine the 6 jets to have the expected kinematics.

4.1 χ^2 fitter

After the neural net cut we proceed to reconstruct the event kinematics. Even if we have selected events with $6 \leq N_{jets} \leq 8$ and $\Delta R_{min} \geq 0.5$, in order to reduce combinatories, only the first 6 jets in order of decreasing E_T will be chosen to fully reconstruct the event. Now we define a χ^2 as a function of only one free parameter, the top mass itself, and ask MINUIT to minimize it. In order to reduce the number of possible permutations, we reject permutations which assign the tagged jet to the 4 jets reconstructing the W's.; thus we are left with 30 possible combinations between the 6 jets. We choose between them the combination with the "tag" mass being the one with the lowest χ^2. If we have more than one tag in the event, we start correcting the first tagged jet energy according to the b hypothesis, run the kinematic fitter, and obtain a mass. Then we iterate the procedure using as inputs the other tagged jets.

The χ^2 contains two terms which constrain the light quark jets to form the two W masses. Then

[1]The rate is parametrized as a function of the jet-E_T, the number of silicon tracks within the jet and the number of well reconstructed vertices.

a third jet four-momentum is added in order to form two objects closely spaced in the unknown mass. An additional term is added to account for the uncertainties on jet momentum measurements. The definition is thus as follows:

$$\chi^2 = \frac{(m_{jj_1} - m_W)^2}{\Gamma_W^2} + \frac{(m_{jj_2} - m_W)^2}{\Gamma_W^2} + \frac{(m_{jjj_1} - m_t)^2}{\Gamma_t^2} + \frac{(m_{jjj_2} - m_t)^2}{\Gamma_t^2} + \Sigma_{i=1}^{N} \frac{(p_i^{fit} - p_i^{data})^2}{\sigma_i^2}$$

where m_{jj} is the invariant mass of the di-jet, m_{jjj} is the invariant mass of the tri-jet, $m_W = 80.4\ GeV/c^2$ and $\Gamma_W = 2.1\ GeV/c^2$ [3] the W boson mass and width, $\Gamma_t = 1.5\ GeV/c^2$ [2] and m_t the free parameter, with the constraint that the two top masses have to be equal.

4.2 Background validation

We already mentioned that our background invariant mass distribution is derived through a tag rate parametrization. For each event passing the kinematical selection, but before the b-tag requirement (the pre-tag sample), we call our χ^2 fitter for each "taggable" jet in the event, that is to say, each jet with at least two silicon tracks associated to it. To be as more consistent as possible with the mass extraction in data, we follow the hypothesis that the taggable jet be a b-jet, and thus apply specific b-jet corrections to it. Thus we extract the mass value as the one with the lowest χ^2, and we weight the mass with the tag rate probability. In this way the integral of the distribution is normalized to the expected background amount.

We iterate over all the taggable jets in the event; thus we are left with as many masses as are the taggable jets in the event.

In order to check how properly these background describes the data, we choose controls region over the NN_{out} cut range where the signal presence is suppressed.

- $0.1 \leq NN_{out} < 0.2$;
- $0.2 \leq NN_{out} < 0.4$;
- $0.4 \leq NN_{out} < 0.6$;
- $0.6 \leq NN_{out} < 0.8$;

where NN_{out} is the output variable from the neural net. For each of these subsample, we plot the quantity which is essential to the measurement, i.e., the invariant mass distribution. We can see in fig. 3 that the agreement is good in the whole control region.

Additionally, we want to compare the χ^2 distribution associated with the mass, both as a double check of the method as well as to control that an additional cut on fit χ^2 will not introduce a bias. Comparison between pre-tag sample derived background and data is shown in fig. 4.

[2] The mass spectrum follows the Breit-Wigner distribution; we use here a Gaussian approximation for simplicity.

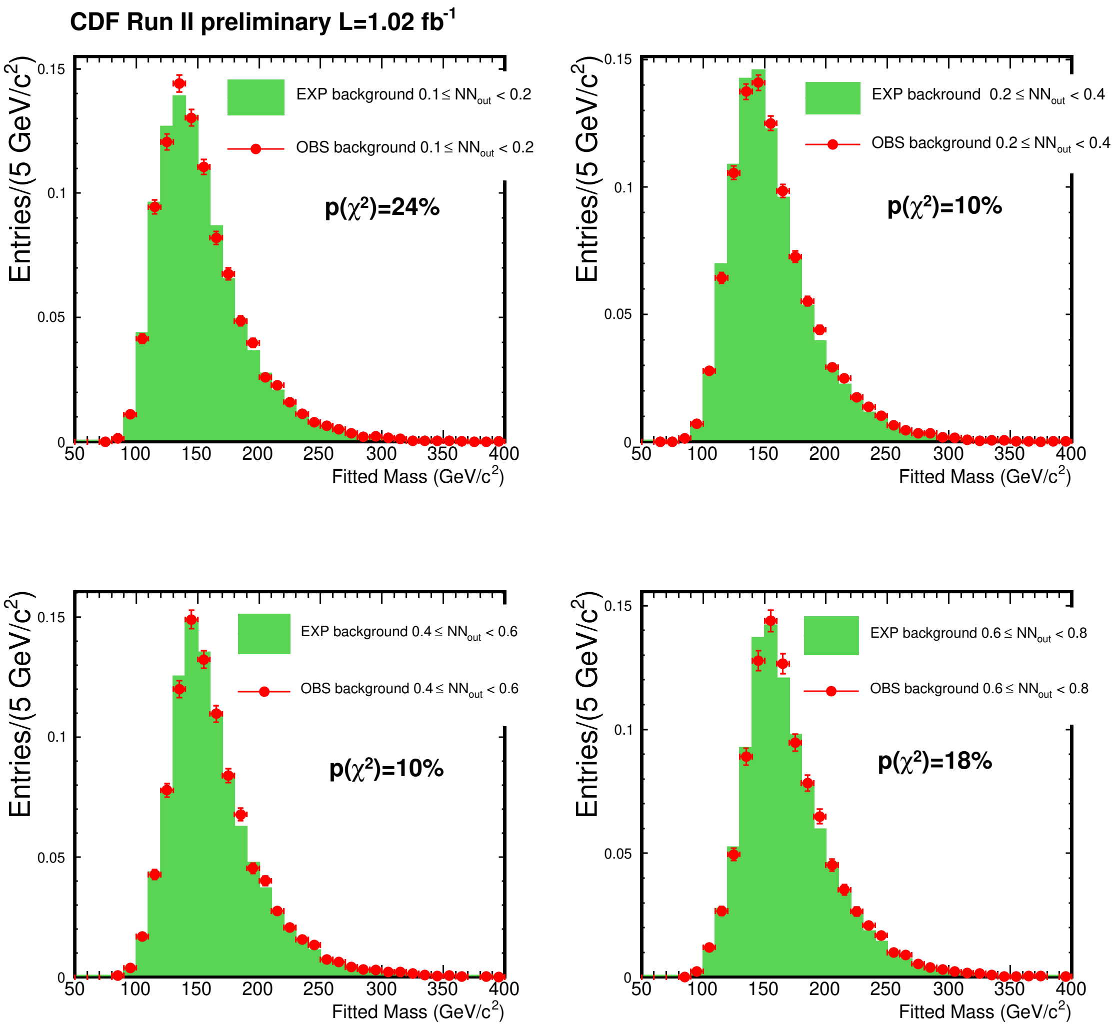

Figure 3: Comparison of the expected invariant mass with the one we get from the data in various control regions.

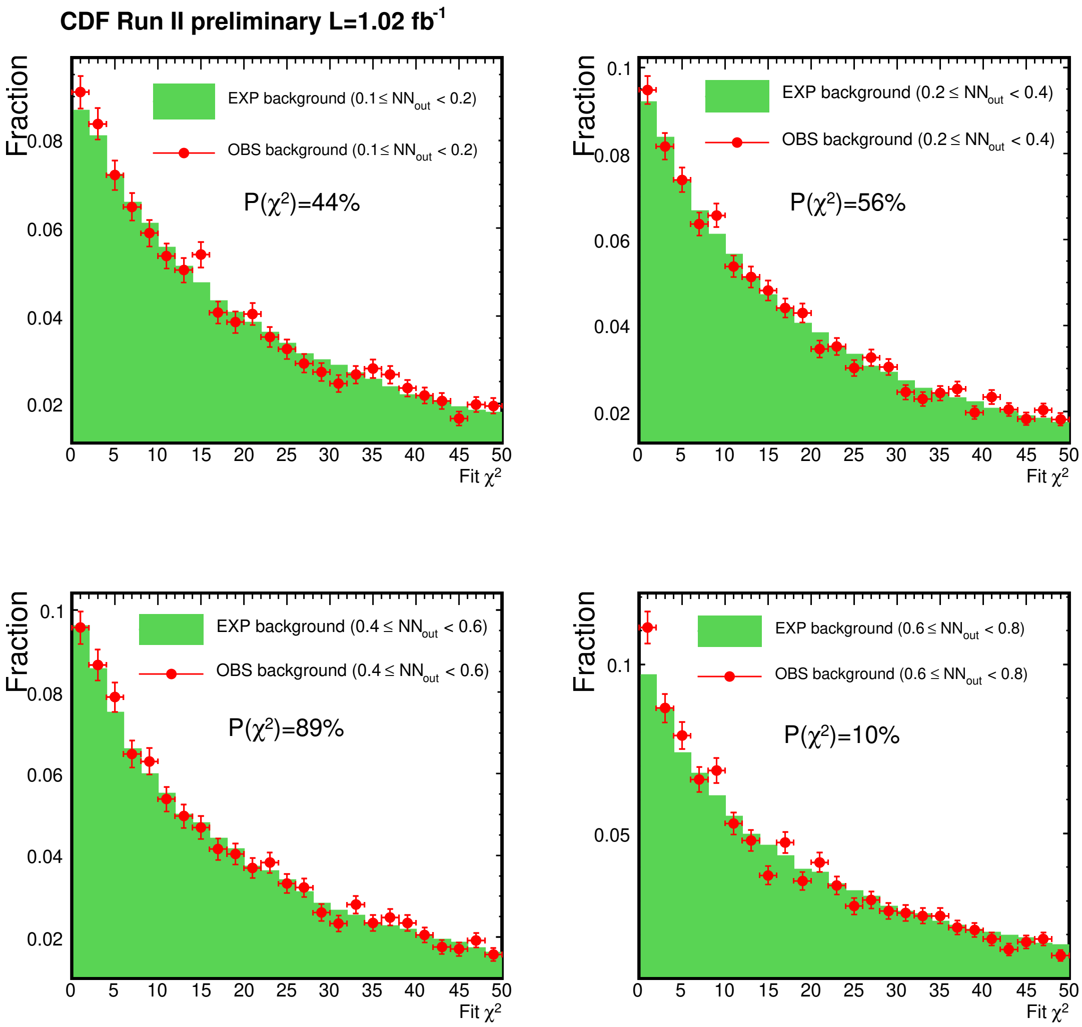

Figure 4: Comparison of expected and data χ^2 distribution

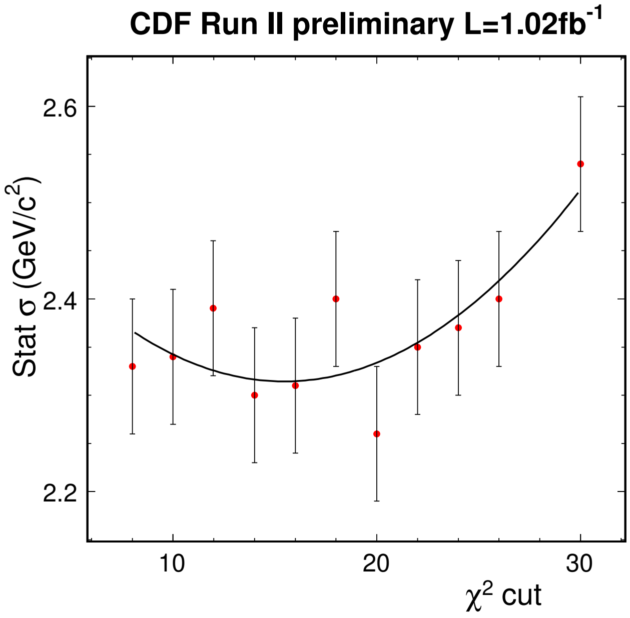

Figure 5: Expected statistical error vs χ^2 cut using $1.02 fb^{-1}$. Overlaid is a parabolic fit of the points.

In order to choose the cut on the neural network selection which provides the smallest statistical uncertainty expected on the signal we span the NN_{out} vs χ^2 region to look for statistical error reduction using pseudoexperiments. The smallest expected statistical uncertainty on the signal corresponds to the cuts $NN_{out} \geq 0.91$[3] and $\chi^2 \leq 16$, as shown in fig. 5, where on the y-axis we report the symmetrized statistical error, after the cut $NN_{out} \geq 0.91$ as a function of the χ^2 cut.

4.3 Template Parametrization

In this analysis 21 $t\bar{t}$ samples, generated herwig with with masses ranging from 150 GeV/c^2 to 200 GeV/c^2, in steps of 2.5 GeV/c^2 are used. With this method it is crucial to know with a certain precision which shape to expect from $t\bar{t}$ generated at different top quark masses. Since we have only a discrete knowledge of the signal templates, we try to achieve better signal - background separation power by interpolating the various samples with a continuous arbitrary curve which is function of only one parameter, the generated mass. The signal is described by a sum of two gaussian distribution, intended to account for the core of the distribution, where correct jet-parton combination are dominant, and a gamma distribution, which describes the combinatoric background. Each parameter in the function is given a linear dependence on generated top quark mass. The resulting probability density function is normalized to 1. Analytically,

[3] We notice that the best cut $NN_{out} \geq 0.91$ is different here with respect to the best cut $NN_{out} \geq 0.94$ used for the cross section measurement.

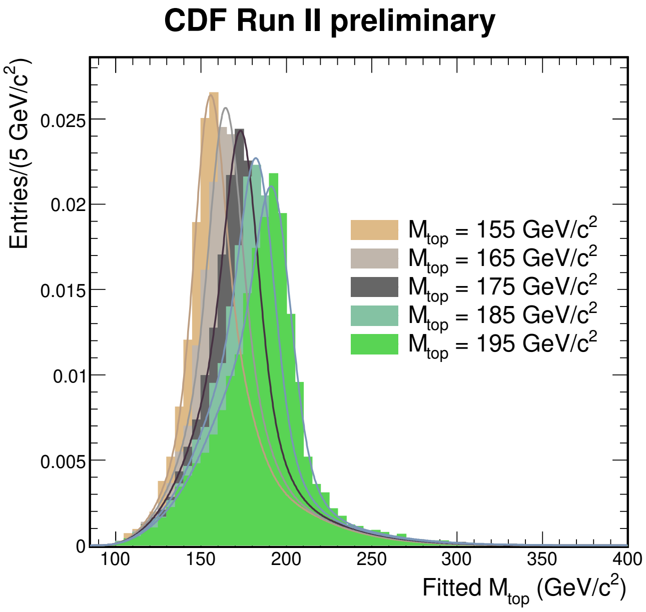

Figure 6: Signal templates with their respective fit overlaid. We see that the core of the distribution, that is where the kinematic fitter choose the correct jet-parton association, follows the signal input top mass.

$$P_{sig}(m|M_{top}) = \delta_7 \cdot \frac{\delta_2^{1+\delta_1}}{\Gamma(1+\delta_1)} \cdot (m-\delta_0)^{\delta_1} \cdot e^{-\delta_2(m-\delta_0)} + \delta_8 \cdot \frac{1}{\sqrt{2\pi}\delta_4} \cdot e^{\frac{-(m-\delta_3)^2}{2\delta_4^2}} + (1-\delta_7-\delta_8) \cdot \frac{1}{\sqrt{2\pi}\delta_6} \cdot e^{\frac{-(m-\delta_5)^2}{2\delta_6^2}}$$

where

$$\delta_i = \alpha_i + \beta_{i+9} \cdot (M_{top} - 175)$$

and M_{top} is the signal generated top mass. The dependence of the parameters on the top mass will allow us to interpolate between the discrete set of signal templates (see Fig. 6), giving a higher statistical power.

The background probability density function is parametrized using two gamma distribution and one gaussian distribution. Obviously, we do not expect any top mass dependence here.

$$P_{bkd}(m) = \delta_8 \cdot \frac{\delta_2^{1+\delta_1}}{\Gamma(1+\delta_1)} \cdot (m-\delta_0)^{\delta_1} \cdot e^{-\delta_2(m-\delta_0)} + \delta_9 \cdot \frac{\delta_5^{1+\delta_4}}{\Gamma(1+\delta_4)} \cdot (m-\delta_3)^{\delta_4} \cdot e^{-\delta_5(m-\delta_3)} +$$

$$+(1-\delta_8-\delta_9) \cdot \frac{1}{\sqrt{2\pi}\delta_7} e^{\frac{-(m-\delta_6)^2}{2\delta_7^2}}$$

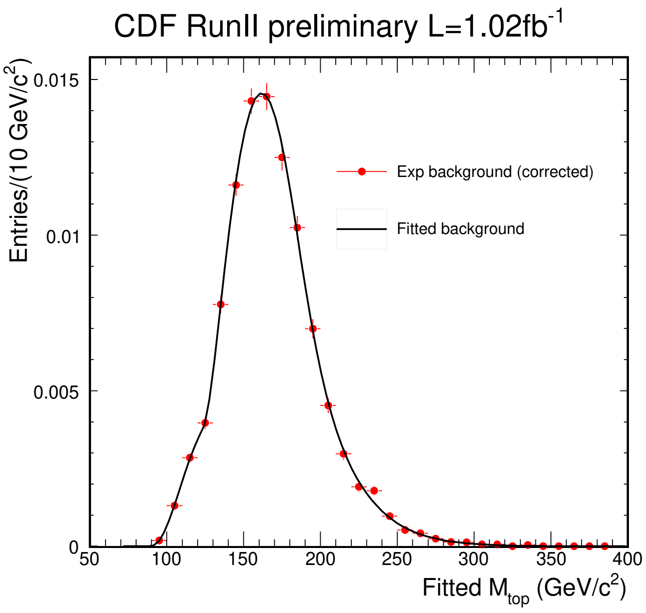

Figure 7: Expected background with fit superimposed.

5 Likelihood

Now that we have a model for signal and background mass templates, we use a likelihood function to compute the probability of the masses survivings the kinematic cuts to come from the the expected p.d.f.'s. The function is divided in two parts: the first accounts for the signal-background discrimination, and so include as parameters to be minimized the number of signal tags, n_s, the number of background tags n_b, and the "true" top quark pole mass; the second term is a constrain on the amount of background tags, which we obtain from the tag rate parametrization once applied to the pre-tag sample.

$$\mathcal{L} = e^{-\frac{(n_s+n_b-N)^2}{2\sigma_N^2}} \cdot \prod_{i=1}^{N} \frac{n_s \cdot P_{sig}(m_i|M_{top}) + n_b \cdot P_{bkd}(m_i)}{n_s + n_b} \cdot e^{-\frac{(n_b-n_{exp})^2}{2\sigma_{n_{exp}}^2}}$$

where $\sigma_{n_b} = 5.0\%$ is the uncertainty on the expected background and $\sigma_{n_b} = \sqrt{N}$ is the error on the total number of observed tags, N.

In order to facilitate the computation, we minimize the negative logarithm of the likelihood using MINUIT, instead of maximizing the likelihood itself. Minimization is performed with respect to the three free parameters present in the likelihood function: n_b, n_s and m_{top}. Statistical error on the parameters are given by MINOS taking positive and negative statistical error as the difference between the input top mass and the values M^+ and M^- for which stands the relation $-lnL(M^{\pm})+$

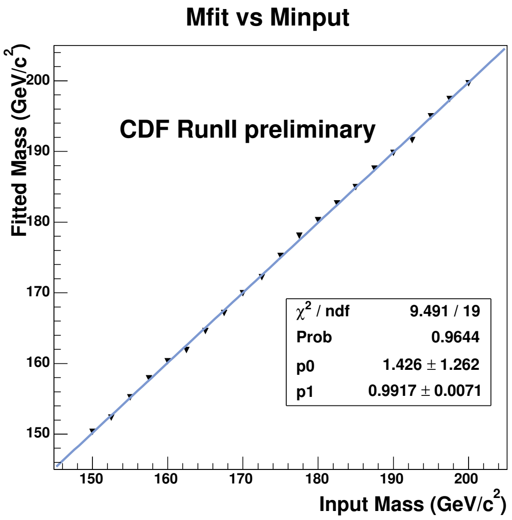

Figure 8: Measured mass as a function of the input MonteCarlo mass using pseudoexperiments. Fit slope is consistent with 1.

$lnL(M) = -1/2$; correlation between the parameters is properly taken into account by MINOS itself.

5.1 Sanity checks and expected performance

We want to check for possible biases in the top mass measurement introduced by our method, as well as to be as accurate as possible on our estimate of the statistical power of our method before performing the measurement on our data sample. To do so, we run pseudo-experiments, extracting pseudo-data with the predicted amount from our signal and background templates.

For each of the 21 input top masses, ranging from 150 GeV/c^2 to 200 GeV/c^2, we fix the total amount to be the one observed in the data, N_{tot}. Then we take our background expectation of tags passing the cuts, fluctuate it according to Poisson statistics, and get the number of signal tags as the difference between N_{tot} and the fluctuated background. We perform the measurement on 20 different sets of pseudodata, and plot output mass and pull means with respect to the input mass (see fig. 8 and fig. 9); since the number of tries is so limited, we plot the mean of the distribution, and use the $RMS/\sqrt{N_{PE's}}$ as the error estimator. We use a larger number of pseudo-experiments, $20^2 = 400$ PE's to evaluate the precision on uncertainties estimates plotting the pull widths with respect to the input mass, as in fig. 10. Here the width is extracted from a gaussian fit of the pull distribution, and the corresponding error is the one given by ROOT.

We proceed now to apply the method on the selected data, corresponding to an integrated luminosity of 1.02 fb^{-1}. We measure $n_s = 334 \pm 33$ signal tags and $n_b = 573 \pm 26$ background

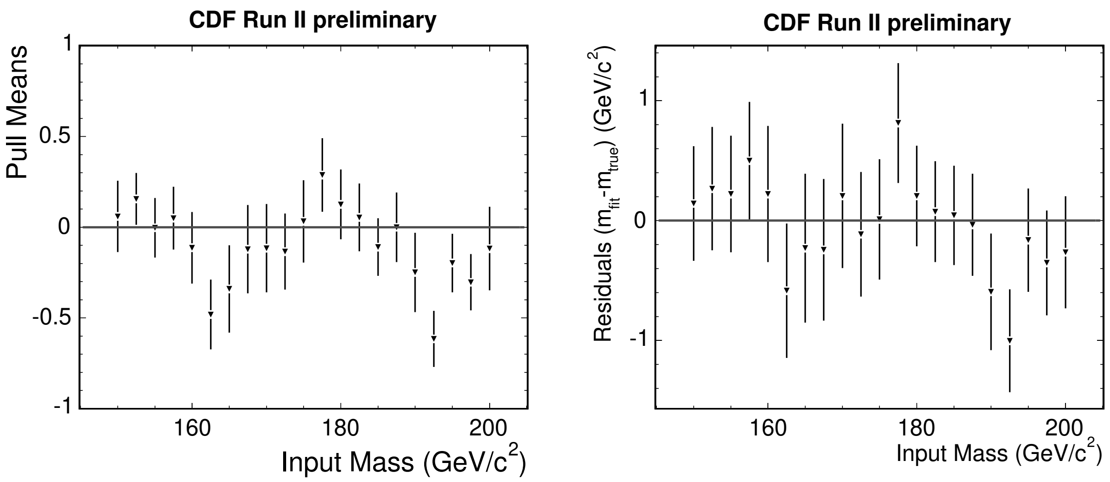

Figure 9: Mean of pull distributions, and respective residuals, with respect to different top quark generated masses.

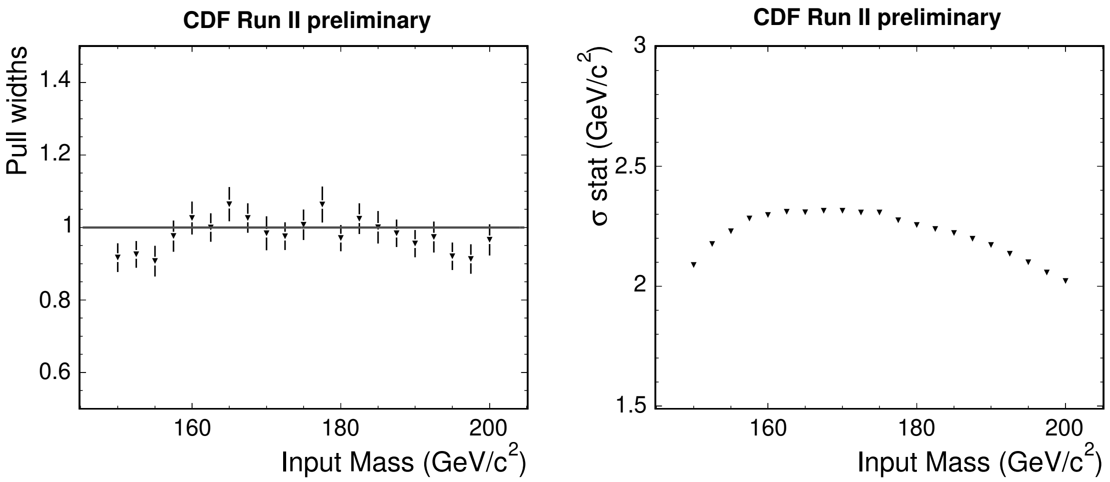

Figure 10: Pull widths and the corresponding errors (left), extracted from a gaussian fit of the pull distributions. Expected statistical error with respect to input masses (right).

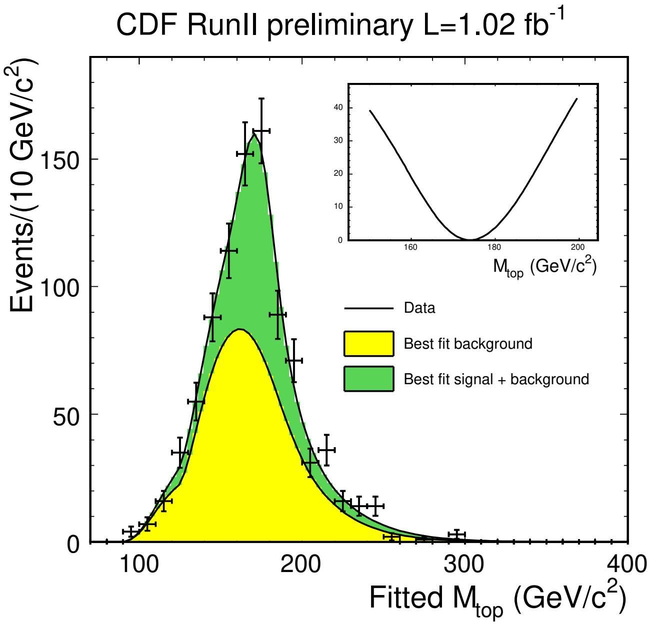

Figure 11: Invariant mass distribution after $NN_{out} \geq 0.91$, $\chi^2 < 16$ and the requirement of the presence of at least 1 tag. Superimposed are the background template (in yellow), and the combined, signal ($m_{top} = 174.0\, GeV/c^2$) plus background templates (in green). On the top-right corner is shown the negative log-likelihood, where the scale is changed to have the minimum equal to zero.

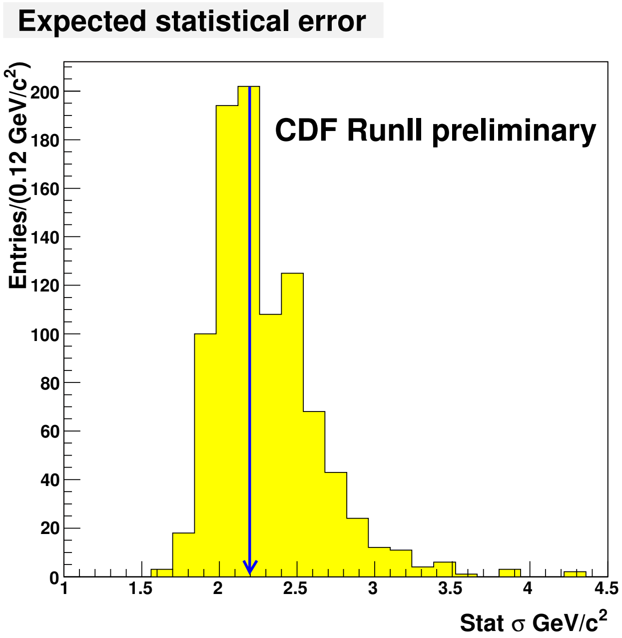

Figure 12: Expected statistical error using pseudoexperiments with input mass equal to 174.0 GeV/c^2. Above the distribution we draw the measured statistical error. About 40% of PE's have lower statistical error.

tags. We obtain the following value for the top quark mass and its statistical error : $M_t = 174.0 \pm 2.2\ GeV/c^2$. In fig. 11 we plot the observed data invariant mass distribution, where we superimpose the background and signal plus background template shapes. In fig. 12 we plot the observed statistical error on the distribution we get from pseudo-experiments using as input mass $M_t = 174.0\ GeV/c^2$.

6 Systematic uncertainties

We report here the estimate of the various sources of systematic error. Since we are dealing with events with high jet multiplicity, and the measurement technique relies on calorimetry-dependent observables, we expect the jet energy scale uncertainty to be the dominant one. The evaluation consists in performing pseudo-experiments where we vary each possible source of systematic uncertainties of $\pm1\sigma$, extract from the varied sample the expected amount of signal pseudodata, compare it with signal and background unchanged pdf's as described above. We take as estimation of the uncertainty the resulting top mass measurement shift. In table 1 is shown a summary of all the systematic uncertainty source and the total amount.

In order to evaluate the jet energy scale uncertainty we use a $t\bar{t}$ sample (herwig) with $m_t = 175\ GeV/c^2$ in which we artificially scale all jet energies by $\pm1\sigma$, run the kinematic selection on

Source	Systematics (GeV/c^2)
Jet energy scale	4.5
Generator	1.0
b-jet energy scale	0.5
Parton Distribution Function	0.5
Background shape	0.5
Background fraction	0.5
ISR	0.5
FSR	0.5
b-tag	0.5
MC statistics	0.1
Template parametrization	0.1
Total	4.8

Table 1: Breakdown of systematic uncertainty sources and respective amounts. We see that jet energy scale is by far the dominant contribution.

modified variables and obtain $M_{+1\sigma}$ and $M_{-1\sigma}$ invariant mass distributions. We use them as inputs to pseudoexperiments, and take half-difference of the two values, 4.5 GeV/c^2 as our estimate of jet energy scale uncertainty. We account for the different generation and fragmentation models implemented in Herwig and Pythia events, comparing Pythia $m_t = 175\,GeV/c^2$ extracted invariant masses with our Herwig-based likelihood technique. Consequently we get our generator systematic to be equal to 1.0 GeV/c^2. For the pdf's related uncertainty we consider the difference in measured top mass using dedicated Monte Carlo samples based on the MRST72 and MRST75 PDF parametrizations. This difference amounts to 0.1 ± 0.5 GeV/c^2. To be conservative we take the statistical uncertainty on the systematic as the systematic itself. Background normalization is known within a 5% error. We vary the background contribution using pseudoexperiments, increasing and decreasing its amount by 1σ, then take half difference as systematics. This amounts to 0.25 ± 0.50 GeV/c^2. We want to considere moreover the uncertainty associated with the presence of signal tags in the data-driven background. To do so we build two extra background samples, where we subtract to it the expected amount of signal, where the signal normalization comes from our iterative technique (see section 3), and the signal shape from the weighting of 172.5 GeV/c^2 and 177.5 GeV/c^2 signal pre-tag masses through tag rate parametrization. We perform pseudo-experiments and tak half-difference as background shape systematics; this amounts to 0.5 ± 0.5 GeV/c^2. Studies from the dilepton mass in Drell-Yann events showed that the measurement uncertainty can be reproduced by varying two showering parameters which are sensitive to the ISR and FSR spectrum: Λ_{QCD} and the K factor of the ISR/FSR transverse momentum scale. We fix the respective uncertainties as the semidifference of measured top masses from Pythia generated samples with more or less initial and final state radiation. A systematic uncertainty due to Monte Carlo statistics is determined by Poisson fluctuating the templates 100 times and using the fluctuated, reparametrized templates to run pseudoexperiments. The width of the resulting mass distribution, 0.1 GeV/c^2, is used as the systematic uncertainty. We attach a systematic uncertainty to the signal and background fit, by computing the mean of the difference between measured top mass and the mass we would obtain extracting pseudodata from the fit itself. This amounts to 0.1 GeV/c^2. B-tagging scale factor accounts for the difference in efficiencies between data and Monte Carlo. We evaluate here a systematic due to E_T dependence of b-tagging scale factor. Since background is data-driven, the analysis is sensitive to an overall uncertainty in the b-tagging scale factor only through signal shapes, where 1 tag and 2 tag contribution to the template can vary slightly. To estimate a systematic uncertainty, we allow a variation of $\pm5\%$ over an E_T range of $15 - 100$ GeV/c^2. The mass shift we observe is consistent with zero.

7 Summary

We presented the template method top quark mass measurement in the all-hadronic channel for the integrated luminosity of 1.02 fb^{-1}. We described the technique used to extract the mass, studied its stability and predicted reliably its statistical power; finally we studied the many different sources of systematic uncertainties. Applying the method to the available data, we obtain $m_{top} = 174.0 \pm 2.2(stat) \pm 4.8(syst)$ GeV/c^2.

References

[1] The CDF Collaboration, *Measurement of the $t\bar{t}$ production cross section in the all-hadronic channel (1.02 fb^{-}1)*, see http://www-cdf.fnal.gov/physics/new/top/confNotes/cdf8402_allhad_1invfb_xsec.ps

[2] See http://schwind.home.cern.ch/schwind/talks/ITseminar/index.htm and http://www.fynu.ucl.ac.be/users/c.delaere/level2/MLP/ for a description of MLP.

[3] S. Eidelman et al. (Particle Data Group), Phys. Lett. B592, 1 (2005).