Measurement of the Top-Quark Mass with Dilepton Events Selected Using Neuroevolution at CDF

Aaltonen, T.

2009


http://hdl.handle.net/10138/24482
https://doi.org/10.1103/PhysRevLett.102.152001

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
Measurement of the top quark mass with dilepton events selected using neuroevolution at CDF

We report a measurement of the top quark mass $M_t$ in the dilepton decay channel $t\bar{t}\rightarrow b\ell^+\nu\ell^-$. Events are selected with a neural network which has been directly optimized for statistical precision in top quark mass using neuroevolution, a technique modeled on biological evolution. The top quark mass is extracted from per-event probability densities that are formed by the convolution of leading order matrix elements and detector resolution functions. The joint probability is the product of the probability densities from 344 candidate events in 2.0 $fb^{-1}$ of $pp$ collisions collected with the CDF II detector, yielding a measurement of $M_t = 171.2 \pm 2.7$ (stat.) $\pm 2.9$ (syst.) GeV/$c^2$.

PACS numbers: 14.65.Ha, 13.85.Ni, 13.85.Qk, 12.15.Ff
Over ten years after the discovery of the top quark, its mass, $M_t$, remains a quantity of great interest. $M_t$-dependent terms contribute to radiative corrections to precision electroweak observables, thus providing information on the unobserved Higgs boson [1] and other particles in possible extensions to the standard model [2] (SM). Top quarks are produced only at the Fermilab Tevatron, primarily in pairs and decay $\approx 100\%$ to a $W$ boson and a $b$ quark, $t\bar{t} \rightarrow W^+bW^-\bar{b}$, in the SM. The dilepton channel, where both $W$ bosons decay to charged leptons (electrons and muons, including leptonic decays of $\tau$ leptons) and neutrinos, has the smallest branching fraction, but also has the least number of hadronic jets in the final state and hence a smaller sensitivity to their energy calibration. Significant differences in the measurements of $M_t$ in different decay channels could indicate contributions from sources beyond the SM [3].

Reconstruction of $M_t$ in the dilepton channel presents unique challenges, as the two neutrinos in the final state result in a kinematically underconstrained system. We utilize a likelihood-based estimator that convolutes leading order SM matrix elements and detector resolution functions and integrates over unmeasured quantities. Prior applications of this method to dilepton events have yielded the most precise measurements of $M_t$ in this channel [4, 5, 6]. These prior measurements utilize event selection criteria that were designed to maximize signal purity for a measurement of the $t\bar{t}$ production cross section [7]. The selection optimization for precision in $M_t$ is hampered by the difficulty of searching the space of arbitrary multivariate selections. Well established multivariate algorithms such as neural networks are typically limited to minimization of a specific metric, such as misclassification error. They are not designed to optimize an event ensemble property, such as the uncertainty on the top quark mass. In contrast, the technique of neuroevolution [8] combines the parametrization of an arbitrary multivariate selection described by a neural network with an evolutionary minimization approach to search for the network weights and topology which optimizes an arbitrary metric. In this Letter, we present a measurement using an improved matrix element analysis technique and an event selection optimized with neuroevolution to minimize the expected statistical uncertainty in the top quark mass measurement. We utilize 2.0 fb$^{-1}$ of data collected between March 2002 and May 2007 with the CDF II detector at the Fermilab Tevatron.

CDF II [9, 10, 11] contains a charged particle tracking system consisting of a silicon microstrip tracker and a drift chamber immersed in a 1.4 T magnetic field. Surrounding electromagnetic and hadronic calorimeters measure particle energies. Outside the calorimeters, drift chambers and scintillators detect muons.

We use lepton triggers that require an electron or muon with $p_T > 18$ GeV/c. We define a preselection which satisfies the basic signature of top dilepton decay: two oppositely charged leptons with $p_T > 20$ GeV/c, two or more jets with $E_T > 15$ GeV [12] within the region $|\eta| < 2.5$, $E_T > 20$ GeV [13], and dilepton invariant mass $M_{ll} > 10$ GeV/c$^2$. Suppression of the $Z \rightarrow ll$ background is performed by the subsequent neural-network selection.

Neuroevolution, an approach modeled on biological evolution, is used to search directly for the optimal neural network. Beginning with a population of 150 networks with random weights, the statistical precision of $M_t$ is evaluated for each network by performing experiments using the simulated signal and background events which survive a threshold requirement on the network output. The events are simulated using the PYTHIA [14] and ALPGEN [15] generators and a full detector simulation [16]. Poorly performing networks are culled and the 30 strongest performers are bred together and mutated in successive generations until performance reaches a plateau in a statistically independent pool of events, which occurs after 15 generations. The statistical uncertainty obtained from the best performer in each generation is shown in Fig. 1(a). In the context of an arbitrary but a priori fixed choice of network threshold, the networks evolve to optimize the selection regardless of the threshold’s value. Because we have optimized directly on the final statistical precision rather than some intermediate or approximate figure of merit, the best-performing network is the one which gives the most precise measurement. This approach has been shown to significantly outperform traditional methods in event selection [17]. In particular, we use neuroevolution of augmenting topologies (NEAT) [18], a neuroevolutionary method capable of evolving a network’s topology and weights.

Some of the events passing this selection have secondary vertex tags [19], which enhance $b$-quark fraction and thus signal purity. We exploit this enhancement by separately fitting events with and without secondary vertex tags, and combining the fits. The predicted number of signal and background events is shown in Table I. Using the optimized selection improves the a priori statisti-
FIG. 1: Top, expected statistical uncertainty for the best network in each successive generation of network evaluation. The points show the average performance for each generation; the error bars show the variation due to the randomly generated networks in generation 0. Bottom, expected statistical uncertainty on $M_t$ versus signal fraction after neural network selection, for all evaluated networks. The selection used in previous measurements is shown (•) for comparison. The arrows show the expected statistical uncertainty and signal fraction corresponding to the selection network used in the analysis.

TABLE I: Expected sample composition after neural network selection for events with and without secondary vertex tags.

<table>
<thead>
<tr>
<th>Source</th>
<th>$N(0\text{-tag})$</th>
<th>$N(\geq 1\text{-tag})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z\rightarrow ll$</td>
<td>116.5 ± 18.6</td>
<td>4.1 ± 1.8</td>
</tr>
<tr>
<td>$Z\rightarrow ll + c\bar{c}/b\bar{b}$</td>
<td>9.3 ± 1.4</td>
<td>10.1 ± 4.0</td>
</tr>
<tr>
<td>WW, WZ, ZZ, Wγ</td>
<td>17.3 ± 5.9</td>
<td>0.7 ± 0.7</td>
</tr>
<tr>
<td>Misidentified leptons</td>
<td>29.0 ± 8.7</td>
<td>4.5 ± 1.1</td>
</tr>
<tr>
<td>$t\bar{t}$ ($\sigma = 6.7\text{ pb}$, $M_t = 175\text{ GeV}/c^2$)</td>
<td>43.8 ± 4.4</td>
<td>78.0 ± 6.2</td>
</tr>
<tr>
<td>Total</td>
<td>215.8 ± 21.9</td>
<td>97.5 ± 7.2</td>
</tr>
</tbody>
</table>

Observed (2.0 fb$^{-1}$) 246 98

due to the increase in $t\bar{t}$ acceptance and the suppression of background effects as described below. The distribution of expected statistical uncertainty versus signal purity for all evaluated networks can be seen in Fig. 1(b).

We express the probability density for the observed lepton and jet measurements, $x_i$, as a function of the top quark mass $M_t$ as $P_s(x_i|M_t)$. We calculate $P_s(x_i|M_t)$ using the theoretical description of the $t\bar{t}$ production and decay process with respect to $x_i$, $P_s(x_i|M_t) = [1/\sigma(M_t)][d\sigma(M_t)/dx_i]$, where $d\sigma/dx_i$ is the differential cross section and $\sigma$ is the total cross section. The term $1/\sigma(M_t)$ ensures that the probability density satisfies the normalization condition, $\int d x_i P_s(x_i|M_t) = 1$.

We evaluate $P_s(x_i|M_t)$ by integrating over quantities that are not directly measured, such as neutrino momenta and quark energies. The effect of simplifying assumptions is estimated using simulated experiments. We integrate over quark energies using a parameterized detector transfer function $W(p,j)$, defined as the probability of measuring jet energy $j$ given quark energy $p$.

We account for backgrounds using their probability densities $P_{bg}(x_i)$ and form the full per-event probability

$$P^a(x_i|M_t) = P_s(x_i|M_t)p_s^a + \sum_k P_{bg_k}(x_i)p_{bg_k}$$

The functions $P_{bg_k}(x_i)$ are calculated using the differential cross-section for each background. The proportions $p_s^a$ and $p_{bg_k}$ depend on whether the event has a secondary vertex tag, and are obtained from Table I. We evaluate background probability densities for: $Z/\gamma^*(\rightarrow ee, \mu\mu)+\text{jets}$, $W+\geq 3\text{ jets}$ where a jet is misidentified as a lepton, and $WW+\text{jets}$. Probability densities for smaller backgrounds ($WZ, ZZ, W\gamma$ and $Z\rightarrow \tau\tau$) provide negligible gain in sensitivity and are not modeled.

The posterior joint probability for the sample is the product of the per-event probability densities,

$$P(x|M_t) = \prod_{i=0}^{n} P^0(x_{i0}|M_t) \times \prod_{i=1}^{n} P^{+1}(x_{i1}|M_t)$$

over all untagged ($i_0$) and tagged ($i_1$) events. The measured mass $M_t$ is taken as the mean $\langle M_t \rangle$ computed using the posterior probability, and the measured statistical uncertainty $\Delta M_t$ is taken as the standard deviation.
The response of our method for simulated experiments (Fig. 3a) is consistent with a linear dependence on the true top mass. Its slope is less than unity due to the presence of unmodeled background. We derive corrections, $M_t \rightarrow 175.0 \text{ GeV}/c^2 + (M_t - 171.0 \text{ GeV}/c^2)/0.86$ and $\Delta M_t \rightarrow \Delta M_t/0.86$, from this response and apply them to the measured quantities in data.

![Graph](a) Mean measured $M_t$ in simulated experiments versus top quark masses. The solid line is a linear fit to the points. (b) Pull widths from simulated experiments versus top quark masses. The solid line is the average over all points.

From the pull distribution of our simulated experiments, we find that $\Delta M_t$ is underestimated (Fig. 3b). This is due to simplifying assumptions made in the probability calculations for computational tractability [5]. These assumptions are violated in small, well-understood ways in realistic events. We scale $\Delta M_t$ by an additional factor, $S = 1.16$, derived from our simulated experiments. Applying this method to the 344 candidate events, we measure $M_t = 171.2 \pm 2.7(\text{stat.}) \text{ GeV}/c^2$. The posterior probability is Gaussian within the statistical accuracy of the Monte Carlo integration.

There are several sources of systematic uncertainty in our measurement, which are summarized in Table II. The single largest source of systematic error comes from the uncertainty in the jet energy scale, which we estimate to be $2.5 \text{ GeV}/c^2$ by varying the scale within its uncertainty [23]. An uncertainty specific to jets resulting from $b$ partons contributes $0.4 \text{ GeV}/c^2$ while in-time pileup contributes $0.2 \text{ GeV}/c^2$. Uncertainty due to the Monte Carlo generator used for $t\bar{t}$ events is estimated as the difference in $M_t$ extracted from PYTHIA events and HERWIG [24] events and amounts to $0.9 \text{ GeV}/c^2$. Uncertainties due to PDFs are estimated using different PDF sets (cteq5l [25] vs. MRST72 [26]), different values of $\Lambda_{QCD}$, and varying the eigenvectors of the cteq6m [25] set; the quadrature sum of the latter two (dominant) uncertainties is $0.6 \text{ GeV}/c^2$. The limited number of background events available for simulated experiments results in an uncertainty on the shape of the background distributions, which yields an uncertainty on $M_t$ of $0.5 \text{ GeV}/c^2$. Uncertainty due to imperfect modeling of initial and final state QCD radiation (ISR and FSR, respectively) is estimated by varying the amounts of ISR and FSR in simulated events [27] and is estimated to be $0.5 \text{ GeV}/c^2$. The uncertainty in the mass due to uncertainties in the response correction is evaluated by varying the response within the uncertainties shown in Fig. 3a and is $0.4 \text{ GeV}/c^2$. The contribution from uncertainties in background composition is estimated by varying the background normalizations from Table I within their uncertainties and amounts to $0.3 \text{ GeV}/c^2$. We estimate the uncertainty coming from modeling of the missing transverse energy in $Z/\gamma^* \rightarrow t\bar{t}$ events and the uncertainty in the data-derived model of misidentified leptons to be $0.2 \text{ GeV}/c^2$. The uncertainty in the lepton energy scale contributes an uncertainty of $0.1 \text{ GeV}/c^2$ to our measurement. Adding in quadrature yields a total systematic uncertainty of $2.9 \text{ GeV}/c^2$.

In summary, we have presented a new measurement of the top quark mass in the dilepton channel. We have applied the technique of neuroevolution, for the first time in particle physics, to devise an event selection criterion which optimizes statistical precision. We measure $M_t = 171.2 \pm 2.7(\text{stat.}) \pm 2.9(\text{syst.}) \text{ GeV}/c^2$. This is the single most precise measurement of $M_t$ in this channel to date, is in good agreement with measurements in other channels [28, 29], and represents a $\sim 30\%$ improvement in statistical precision over the previously published measurements in this channel [6, 30, 31].

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 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 für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and

<table>
<thead>
<tr>
<th>Source</th>
<th>Size (GeV/$c^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic jet energy scale</td>
<td>2.5</td>
</tr>
<tr>
<td>$b$-Jet Energy Scale</td>
<td>0.4</td>
</tr>
<tr>
<td>In-time pileup</td>
<td>0.2</td>
</tr>
<tr>
<td>Generator</td>
<td>0.9</td>
</tr>
<tr>
<td>PDFs</td>
<td>0.6</td>
</tr>
<tr>
<td>Background statistics</td>
<td>0.5</td>
</tr>
<tr>
<td>Radiation</td>
<td>0.5</td>
</tr>
<tr>
<td>Response correction</td>
<td>0.4</td>
</tr>
<tr>
<td>Sample composition uncertainty</td>
<td>0.3</td>
</tr>
<tr>
<td>Background modeling</td>
<td>0.2</td>
</tr>
<tr>
<td>Lepton energy scale</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td>2.9</td>
</tr>
</tbody>
</table>
the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Educaci´on y Ciencia and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

[11] CDF uses a cylindrical coordinate system with the z axis along the proton beam axis. Pseudorapidity is η ≡ − ln(tan θ/2), θ (φ) is the polar (azimuthal) angle relative to the z axis, while p_T = |p| sin θ, E_T = E sinθ.
[13] Missing transverse energy E_T ≡ | − \sum_i E_T^i \hat{n}_i|, where E_T^i is the transverse energy in calorimeter tower i, and \hat{n}_i is the unit transverse vector from the beamline to tower i.
[22] While ≈15% of \hat{t} are produced by gluon-gluon fusion (gg → \hat{t}), it has negligible effect on our measurement.