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

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1103/PhysRevLett.102.152001">http://dx.doi.org/10.1103/PhysRevLett.102.152001</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Physical Society (APS)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Feb 15 04:58:29 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/114680">http://hdl.handle.net/1721.1/114680</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>

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


(CDF Collaboration)

1 Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China
2 Argonne National Laboratory, Argonne, Illinois 60439, USA
3 University of Athens, 157 71 Athens, Greece
4 Institut de Fisica d’Altes Energies, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain
5 Baylor University, Waco, Texas 76798, USA
6a Istituto Nazionale di Fisica Nucleare Bologna, I-40127 Bologna, Italy
6b University of Bologna, I-40127 Bologna, Italy
7 Brandeis University, Waltham, Massachusetts 02254, USA
8 University of California, Davis, Davis, California 95616, USA
9 University of California, Los Angeles, Los Angeles, California 90024, USA
10 University of California, San Diego, La Jolla, California 92093, USA
11 Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain
12 Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
13 Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA
14 Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia
15 Joint Institute for Nuclear Research, RU-141980 Dubna, Russia
16 Duke University, Durham, North Carolina 27708, USA
17 Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA
18 University of Florida, Gainesville, Florida 32611, USA
19 Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy
20 University of Geneva, CH-1211 Geneva 4, Switzerland
21 Glasgow University, Glasgow G12 8QQ, United Kingdom
We report a measurement of the top-quark mass $M_t$ in the dilepton decay channel $t\bar{t} \rightarrow b\ell^+\nu_b\bar{b}\ell^-\bar{\nu}_b$. 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 $\sqrt{s} = 1.96$ of $p\bar{p}$ collisions collected with the CDF II detector, yielding a measurement of $M_t = 171.2^{+2.7}_{-2.6}(\text{stat}) \pm 2.9(\text{syst})$ GeV/$c^2$. 

152001-3
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 = 100% to a \( W \) boson and a \( b \) quark, \( t\bar{t} \rightarrow W^+ b W^- \bar{b} \), 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 [4–6]. 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 parameterization 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 [9–11] 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 \( p_T > 18 \text{ GeV}/c \). We define a preselection which satisfies the basic signature of top dilepton decay: two oppositely charged leptons with \( p_T > 20 \text{ GeV}/c \), two or more jets with \( E_T > 15 \text{ GeV} \) [12] within the region \( |\eta| < 2.5 \), \( E_T > 20 \text{ GeV} \) [13], and dilepton invariant mass \( M_{ll} > 10 \text{ GeV}/c^2 \). 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 \textit{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 [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 the \( b \)-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 \textit{a priori} statistical uncertainty on \( M_t \) over the selection used in previous analyses [6] by 20%. This neural-network selection yields 344 candidate events (Fig. 2). Strikingly, the sample selected by the neural network is expected to be dominated by background events; the resulting measurement is expected to be more precise than previous measurements 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, \( \mathbf{x}_i \), as a function of the top-quark mass \( M_t \), as \( P_i(\mathbf{x}_i|M_t) \). We calculate \( P_i(\mathbf{x}_i|M_t) \) using the theoretical description of the \( \bar{t}t \) production and decay process with respect to \( \mathbf{x}_i \), \( P_i(\mathbf{x}_i|M_t) = \frac{1}{\sigma(M_t)} \times \frac{d\sigma(M_t)}{d\mathbf{x}_i} \), where \( \frac{d\sigma}{d\mathbf{x}_i} \) is the differential cross section \([20–22]\) and \( \sigma \) is the total cross section. The term \( \frac{1}{\sigma(M_t)} \) ensures that the probability density satisfies the normalization condition, \( \int d\mathbf{x}_i P_i(\mathbf{x}_i|M_t) = 1 \).

We evaluate \( P_i(\mathbf{x}_i|M_t) \) \([6]\) 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 parametrized detector transfer function \([5]\) \( W(p, j) \), defined as the probability of measuring jet energy \( j \) given quark energy \( p \).

We account for backgrounds using their probability densities \( P_{bg, i}(\mathbf{x}_i) \) and form the full per-event probability

\[
P_i^n(\mathbf{x}_i|M_t) = P_i(\mathbf{x}_i|M_t)p_i^n + \sum_k P_{bg, i}(\mathbf{x}_i)p_{bg, i}^n.
\]

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 \to ll )</td>
<td>116.5 ± 18.6</td>
<td>4.1 ± 1.8</td>
</tr>
<tr>
<td>( Z \to 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>
<tr>
<td>Observed (2.0 fb(^{-1}))</td>
<td>246</td>
<td>98</td>
</tr>
</tbody>
</table>

The functions \( P_{bg, i}(\mathbf{x}_i) \) are calculated using the differential cross section for each background. The proportions \( p_i^n \) and \( p_{bg, i}^n \) depend on whether the event has \( n \) secondary vertex tags, and are obtained from Table I. We evaluate background probability densities for \( Z/\gamma^*(\rightarrow ee, \mu\mu) + \text{jets} \), \( W+\rightarrow 3 \text{jets} \) where a jet is misidentified as a lepton, and \( WW + \text{jets} \). Probability densities for smaller backgrounds

FIG. 1. (a) 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. (b) Expected statistical uncertainty on \( M_t \) versus signal fraction after neural-network selection, for all evaluated networks. The selection [7] used in previous measurements is shown (×) for comparison. The arrows show the expected statistical uncertainty and signal fraction corresponding to the network used in the analysis.

FIG. 2. The output of the final network evaluated on the collected data (black triangles), with expected signal and background contributions (stacked solid histograms). The data show events passing the preselection. The evolution of the optimum selection network is performed with an \textit{a priori} threshold set at 0.5 for candidate selection. Of the 642 preselected events shown, 344 events pass this threshold and constitute the final candidate sample for mass fitting.
We find that 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. 3(a)] 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, \(M_t \to 175.0\,\text{GeV}/c^2 + (M_t - 171.0\,\text{GeV}/c^2)/0.86\) and \(\Delta M_t \to \Delta M_t/0.86\), from this response and apply them to the measured quantities in data.

From the pull distribution of our simulated experiments, we find that \(\Delta M_t\) is underestimated [Fig. 3(b)]. 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 \(M_t = 171.2 \pm 2.7\) (stat) 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 GeV/c^2 by varying the scale within its uncertainty [23]. An uncertainty specific to jets resulting from \(b\) partons contributes 0.4 GeV/c^2, while in-time pileup contributes 0.2 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 GeV/c^2. Uncertainties due to parton distribution functions are estimated using different parton distribution function sets (CTEQ6L [25] vs MRST72 [26]), using different values of \(\Lambda_{\text{QCD}}\), and varying the eigenvectors of the CTEQ6M [25] set; the quadrature sum of the last two (dominant) uncertainties is 0.6 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 GeV/c^2. Uncertainty due to imperfect modeling of initial (ISR) and final (FSR) QCD radiation is estimated by varying the amounts of ISR and FSR in simulated events [27] and is estimated to be 0.5 GeV/c^2. The uncertainty in the mass due to uncertainties in the

\begin{table}[h]
\centering
\caption{Summary of systematic uncertainties on the measured top-quark mass.}
\begin{tabular}{|l|c|}
\hline
Source & Size (GeV/c^2) \\
\hline
Generic jet energy scale & 2.5 \\
b-jet energy scale & 0.4 \\
In-time pileup & 0.2 \\
Generator & 0.9 \\
Parton distribution functions & 0.6 \\
Background statistics & 0.5 \\
Radiation & 0.5 \\
Response correction & 0.4 \\
Sample composition uncertainty & 0.3 \\
Background modeling & 0.2 \\
Lepton energy scale & 0.1 \\
Total & 2.9 \\
\hline
\end{tabular}
\end{table}
response correction is evaluated by varying the response within the uncertainties shown in Fig. 3(a) and is 0.4 GeV/c². 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 GeV/c². We estimate the uncertainty coming from modeling of the missing transverse energy in Z/γ events and the uncertainty in the data-derived model of misidentified leptons to be 0.2 GeV/c². The uncertainty in the lepton energy scale contributes an uncertainty of 0.1 GeV/c² to our measurement. Adding in quadrature yields a total systematic uncertainty of 2.9 GeV/c².

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 $M_t = 171.2 \pm 2.7 \text{(stat)} \pm 2.9 \text{(syst)}$ GeV/c². 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 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 the Royal Society, UK; the Institut National de Physique Nucléaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Educación 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 defined as $\eta = -\ln \tan \theta/2$, where $\theta$ (φ) is the polar (azimuthal) angle relative to the $z$ axis, while $p_T = |p| \sin \theta$, $E_T = E \sin \theta$.
[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 beam line to tower $i$. 

---

aDeceased.
bWith visitors from Universiteit Antwerpen, B-2610 Antwerp, Belgium.
cWith visitors from Chinese Academy of Sciences, Beijing 100864, China.
dWith visitors from University of Bristol, Bristol BS8 1TL, United Kingdom.
eWith visitors from University of California Irvine, Irvine, California 92697, USA.
fWith visitors from University of California Santa Cruz, Santa Cruz, California 95064, USA.
gWith visitors from Cornell University, Ithaca, New York 14853, USA.
hWith visitors from University of Cyprus, Nicosia CY-1678, Cyprus.
iWith visitors from University College Dublin, Dublin 4, Ireland.
jWith visitors from University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom.
kWith visitors from Universidad Iberoamericana, Mexico Distrito Federal, Mexico.
lWith visitors from University of Manchester, Manchester M13 9PL, England.
mWith visitors from Nagasaki Institute of Applied Science, Nagasaki, Japan.
nWith visitors from University de Oviedo, E-33007 Oviedo, Spain.
oWith visitors from Queen Mary, University of London, London, E1 4NS, England.
pWith visitors from Texas Tech University, Lubbock, Texas 79409, USA.
qWith visitors from IFIC (CSIC-Universitat de Valencia), 46071 Valencia, Spain.
rWith visitors from Royal Society of Edinburgh/Scottish Executive Support Research Fellow.
sWith visitors from Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09042 Monserrato (Cagliari), Italy.
tWith visitors from Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands.

[22] While \( \approx 15\% \) of \( t\bar{t} \) are produced by gluon-gluon fusion \( (gg \rightarrow t\bar{t}) \), it has negligible effect on our measurement.