Search for resonant production of tt[over ¯] decaying to jets in pp[over ¯] collisions at \( \sqrt{s} \)=1.96 TeV

T. Aaltonen et al. (CDF Collaboration)

Phys. Rev. D 84, 072003 — Published 11 October 2011
DOI: 10.1103/PhysRevD.84.072003
Search for resonant production of $t\bar{t}$ decaying to jets in $pp$ collisions at $\sqrt{s} = 1.96$ TeV
This Letter reports a search for non standard model topquark resonances, \( Z' \), decaying to \( t \bar{t} \rightarrow W^+ b W^- \bar{b} \), where both \( W \) decay to quarks. We examine the top-antitop quark invariant mass spectrum for the presence of narrow resonant states. The search uses a data sample of \( pp \) collisions at a center of mass energy of 1.96 TeV collected by the CDF II detector at the Fermilab Tevatron, with an integrated luminosity of 2.8 fb\(^{-1}\). No evidence for top-antitop quark resonant production is found. We place upper limits on the production cross section times branching ratio for a specific topcolor assisted technicolor model in which the top quark plays an important role in theoretical extensions of the standard model (SM). Its large mass gives the top quark a special position within the standard model and may shed light on the dynamics of

 PACS numbers: 13.85.Rm, 14.65.Ha, 14.70.Hp, 14.70.Pw, 14.80.Rt, 14.80.Tt

The discovery of the top quark in 1995 [1] completed the third generation of quarks. After years of its discovery, the top quark plays an important role in theoretical extensions of the standard model (SM). Its large mass gives the top quark a special position within the standard model and may shed light on the dynamics of

...
electroweak symmetry breaking. One SM extension, top-color assisted technicolor [2], predicts new strong dynamics which accounts for the spontaneous breaking of electroweak symmetry and explains the large mass of the top quark. This model predicts a vector particle (Z'), which couples primarily to the third generation of quarks and has no significant couplings to leptons. The existence of a narrow (Γ' = 0.012 M') width Z' resonance decaying to \( t\bar{t} \) pairs, using a lepto.phobic topcolor model [3] as the reference, has been probed since the beginning of Tevatron operations, both at CDF [4] and D0 [5]. Other theories of physics beyond the standard model predict new heavy particles that would be evident as a resonance in the \( t\bar{t} \) production invariant mass distribution [6].

This Letter presents a search for narrow resonant states decaying to top-antitop pairs. In the lepto.phobic topcolor model, top quarks decay as in the SM via the weak interaction, nearly always to a \( W \) boson and a \( b \) quark, \( W \) bosons decay into light fermion-antifermion pairs: a leptonic decay (32.4%) into a charged lepton and a neutrino; or hadronic decay (67.6%) into an up-type quark and a down-type quark. All previously reported searches have been analyses of top-antitop decays in the lepton plus jets channel, where one of the \( W \) bosons decays leptonically (to an electron or a muon) and the other \( W \) decays hadronically. This channel features a clean signature due to the presence of a lepton in the final state, and has a branching ratio of 29%. The result presented in this Letter is an analysis of \( t\bar{t} \) decays in the all-hadronic channel, where both \( W \)’s decay hadronically. Because this topology features only multiple hadronic jets in the final state, it is subject to a considerable multijet QCD background. We demonstrate that this background can be controlled and significantly suppressed with a careful event selection. Analysis of \( t\bar{t} \) decays in the all-hadronic channel is advantageous for several reasons: the channel offers the largest branching ratio (46%) of any of the \( t\bar{t} \) final states; there is no unobservable neutrino in the final state, which permits improved resolution of the \( t\bar{t} \) invariant mass; finally, this sample is orthogonal to that of previous analyses – the result presented in this Letter is complimentary to the previous results in the lepton plus jets channel.

\( p\bar{p} \) collision events analyzed in this paper were produced at the Tevatron collider at a center of mass energy of 1.96 TeV and were recorded by the CDF II detector [7]. The data sample corresponds to a total integrated luminosity of 2.8 \( fb^{-1} \). CDF II is a general purpose particle detector. It consists of high precision tracking systems for vertex and charged particle track reconstruction, surrounded by electromagnetic and hadronic calorimeters for energy measurement, and muon subsystems outside the calorimeter for muon detection. CDF II uses a cylindrical coordinate system with azimuthal angle \( \phi \), polar angle \( \theta \) measured with respect to the positive \( z \) direction along the proton beam, and the distance \( r \) from the beamline. The pseudorapidity, transverse energy, and momentum are defined as \( \eta = -\ln [\tan(\theta/2)] \), \( E_t = E \sin \theta \), and \( P_t = P \sin \theta \), where \( E \) and \( P \) are the energy and momentum of an incident particle.

The data were collected using a multijet online event selection (trigger), which is implemented in three stages. For triggering purposes, the calorimeter is subdivided into a \( 24 \times 24 \) grid of towers in \( \eta-\phi \) space. At level 1, we require at least one trigger tower with transverse energy \( E_{\text{trans}}^{\text{thr}} \geq 10 \geV \). At level 2, we require the sum of the transverse energies of all the trigger towers, \( \sum E_{\text{trans}}^{\text{thr}} \), to be \( \geq 175 \geV \) and the presence of at least four clusters of trigger towers with \( E_{\text{trans}}^{\text{thr}} \geq 15 \geV \). Finally, at level 3 we require four or more reconstructed jets with raw(uncorrected) \( E_t \geq 10 \geV \), where jets are identified as clusters of energy depositions in the calorimeter using a fixed cone ( \( \Delta R = \sqrt{\Delta \phi^2 + (\Delta \eta)^2} = 0.4 \) ) algorithm [8]. The efficiency of this trigger selection on all-hadronic \( t\bar{t} \) events is about 80%. The main background present in this data sample is QCD multijet production.

The jet energies are corrected for calorimeter response, multiple interactions, and energy radiated outside the jet cone [9]. Jets originating from a \( b \) quark are identified by the secvtx [10] algorithm, which searches for tracks with nonzero impact parameter that result from the displaced decay of \( b \) hadrons inside the jet, and fits the tracks to a common vertex. If this vertex is significantly displaced from the primary interaction point, the jet is tagged as a \( b \) jet.

Events compatible with the signal are selected by requiring six or seven jets with \( |\eta| < 2 \) and corrected \( E_t > 15 \geV \). To remove leptonic \( t\bar{t} \) decays, we veto events with well identified leptons [11] or with significant imbalance in transverse momentum [12]. After all the preselections defined above, the SM \( t\bar{t} \) contribution to the data sample is expected to be very small (0.3%). To enrich the signal presence in the data sample we have to apply additional cuts, which we describe later in the paper.

The distinctive feature of this analysis is the use of likelihoods calculated by integrating signal matrix elements both to perform \( t\bar{t} \) invariant mass reconstruction and to suppress the overwhelming background. The full expression of the likelihood for a given event with jet momenta configuration \( \vec{p}_j \) to be the result of SM \( t\bar{t} \) production and decay is given by:

\[
P(\vec{p}_j|m) = \sum_{\text{comb}_1} \int \frac{dz_a dz_b f(z_a) f(z_b)}{4E_a E_b |v_a - v_b|} \frac{d^3 \vec{p}_i}{(2\pi)^3 2E_i} \frac{|\mathcal{M}(m,p)|^2 (2\pi)^4 \delta^4(E_F - E_I) TF(\vec{p}_j | \vec{p}_i) P_i(\vec{p})}{\sigma_{\text{tot}}(m) e(m) N_{\text{comb}_1}}
\]

where \( z_a, b \), \( v_a, b \) and \( E_a, b \) are the fractional momenta, velocities and energies carried by partons \( a, b \) and \( f(z_a, b) \) are the parton distribution functions of colliding proton and antiproton, \( \vec{p}_j \) are jets (partons) four-momenta, \( m \) is the top quark mass, \( \mathcal{M}(m,p) \) is the SM \( t\bar{t} \) leading order matrix element, \( \sigma_{\text{tot}}(m) e(m) \) is the SM \( t\bar{t} \) production cross section times the selection efficiency both as a function of \( m \), and \( E_F(E_I) \) is a generic notation for the
The central four-vector of the final (initial) state, and $P_t(p)$ for the transverse momentum of the $t\bar{t}$ system. The sum is performed over all jet to parton assignments $N_{comb}$. The probability that a parton with energy $E_p$ is observed as a jet with energy $E_j$ is given by the transfer function, $TF(j|p)$, which is parameterized as a function of parton energy and pseudorapidity. Transfer functions are defined individually for the jets associated with $b$ and light jets, as they have different response in the calorimeter. We construct the transfer functions from the events that have all the jets uniquely matched to each individual parton within a cone of $\Delta R = 0.4$. For each energy and $\eta$ region we use smoothed histogrammed distributions of $1 - E_{jet}/E_{parton}$ as $TF(j|p)$ parameterization. A sample of fully simulated $t\bar{t}$ Monte Carlo (MC) events, generated using PYTHIA v6.2 [13] with parton showering followed by the full simulation of the CDF II detector, and assuming $m = 175$ GeV/$c^2$ is used to obtain the $P_t(p)$ and $TF(j|p)$ parametrizations.

The probability density, $P(j|m)$, can be expressed with respect to any variable that is a function of parton four-momenta, in this case the invariant mass of $t\bar{t}$ pairs, $M_{t\bar{t}}$. Integrating Eq. (1) over the parton variables times a delta function $\delta(x - M_{t\bar{t}}(p))$ we obtain the probability density function for each event once the jets are measured. We use the mean value of this distribution as an estimator for $M_{t\bar{t}}$.

To discriminate between SM background and $Z'$ signal events, we calculate event quantities which are sensitive to the presence of a signal and use them as inputs for a neural network which is trained to separate the signal and the background. Keeping in mind that SM $t\bar{t}$ is one of the background samples for $Z'$ events, here we will refer to SM $t\bar{t}$ as the signal sample for the event selection purpose only. We train the neural network to select events with the presence of $t\bar{t}$ pairs and to veto dominant QCD multijet production. Using SM $t\bar{t}$ events as signal events to optimize the event selection and enrich the $t\bar{t}$ content of the sample accomplishes reasonable results for $Z'$ events as shown later in the paper. In addition, this choice makes the search unbiased to a specific mass and the model of $Z'$ hypothesis used.

A first set of 10 kinematic variables, summarized in Table I, has already been shown to be effective [14] in reducing the QCD background. Significant distinguishing features of $t\bar{t}$ production in comparison to the QCD background are high $E_T$ jets, dijet resonances from $W$ decay and trijet resonances from $t$ decay. The centrality is $C = \frac{\sum E_T}{\sqrt{s}}$, where $\sqrt{s}$ is the invariant mass of the multijet system. The aplanarity is defined as $A = \frac{1}{3} Q_1$, where $Q_1$ is the smallest of the three normalized eigenvalues of the sphericity tensor, $M_{ab} = \sum_j P^a_j P^b_j$ calculated in the center of mass frame of all jets, where $a$ and $b$ refer to the spatial components of the jet four-momentum $P_j$. In Table I, $\theta^*$ is a jet emission direction, represented by the angle between the jet direction, measured in the center of mass frame of all jets, and the proton beam axis. For the last variable, MED (matrix element discriminant), we exploit the broad set of information from the event about its production and decay through the SM $t\bar{t}$ matrix element. For each event we calculate 'the minus log probability' of Eq. (1) at 9 different top mass points, $m_t = 155, 160...195$ GeV/$c^2$, and use their sum as the final discriminator.

### Table I. Neural network input variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum E_T$</td>
<td>Scalar sum of all jet $E_T$</td>
</tr>
<tr>
<td>$\sum_j E_T$</td>
<td>As above, excluding two highest $E_T$ jets</td>
</tr>
<tr>
<td>$C$</td>
<td>Centrality, defined in text</td>
</tr>
<tr>
<td>$A$</td>
<td>Aplanarity, defined in text</td>
</tr>
<tr>
<td>$M_{m}^{\min}$</td>
<td>Minimum dijet invariant mass</td>
</tr>
<tr>
<td>$M_{m}^{\max}$</td>
<td>Maximum dijet invariant mass</td>
</tr>
<tr>
<td>$M_{3j}^{\min}$</td>
<td>Minimum trijet invariant mass</td>
</tr>
<tr>
<td>$M_{3j}^{\max}$</td>
<td>Maximum trijet invariant mass</td>
</tr>
<tr>
<td>$E_T^{1/2}$</td>
<td>$E_T \sin^2 \theta^*$ for the highest $E_T$ jet</td>
</tr>
<tr>
<td>$E_T^2$</td>
<td>Geometric mean of $E_T$ of remaining $N - 2$ jets</td>
</tr>
<tr>
<td>MED</td>
<td>Constructed from matrix element</td>
</tr>
</tbody>
</table>

Having defined the variables, to separate $t\bar{t}$ from background events, we use them as inputs to a neural network [15] with two hidden layers and one output node. The neural network is trained on samples of signal and background events with $6 \leq N_{jets} \leq 7$. To model the signal events we use the PYTHIA MC generator at leading order (LO) to produce SM $t\bar{t}$ events assuming a top quark mass of $m_{top} = 175$ GeV/$c^2$ and the theoretical cross section of 6.7 pb [16]. We use the multijet data events as the background sample for training the neural net since the $t\bar{t}$ contribution is expected to be negligible. After the training, the value of the output node, $NN_{out}$, is used as a discriminator between signal and background. Its distribution is shown in Fig. 1. In addition, we show the comparison of the QCD dominated data, MC generated SM $t\bar{t}$ and MC generated $Z'$ events for one of the input variables $\sum E_T$.

In the final event selection we require a cut on the neural net output, $NN_{out} > 0.93$, and at least one jet tagged as having originating from a $b$ quark. The neural net requirement was optimized to suppress the QCD background while enhancing the content of $t\bar{t}$ events by maximizing the SM $t\bar{t}$ significance. Table II shows the selection efficiencies for SM $t\bar{t}$ and $Z'$ events after final event selection cut. There are 2086 events surviving these final selection criteria including 680 SM $t\bar{t}$ events as estimated from the simulated event sample and assuming the NLO theoretical cross section [16]. The remaining events are from QCD multijet processes plus a potential signal contribution from $Z'$ events.
have one or two rate matrix to estimate the probability for an event to associated to the jet, and the number of reconstructed on the transverse energy of the jet, the number of tracks we consider several control regions, defined by the neural requirement to predict the QCD background for events in the final selected sample. To test our background model, we consider several control regions, defined by the neural. 

![Image 1](https://example.com/image1.png)

**FIG. 1.** \(\sum E_T\) and neural net output value for QCD dominated data (black points), SM \(t\bar{t}\) (dark grey histogram), QCD prediction (light grey histogram) and 700 GeV/c\(^2\) \(t\bar{t}\) resonance normalized to the expectation (dashed histogram).

<table>
<thead>
<tr>
<th>(M_2) (GeV/c(^2))</th>
<th>(\sigma) (pb)</th>
<th>(\epsilon \pm \delta \epsilon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM (t\bar{t})</td>
<td>6.7</td>
<td>3.8 ± 0.5</td>
</tr>
<tr>
<td>450</td>
<td>8.96</td>
<td>4.2 ± 0.5</td>
</tr>
<tr>
<td>500</td>
<td>5.66</td>
<td>4.7 ± 0.5</td>
</tr>
<tr>
<td>550</td>
<td>3.40</td>
<td>5.3 ± 0.5</td>
</tr>
<tr>
<td>600</td>
<td>2.09</td>
<td>5.7 ± 0.5</td>
</tr>
<tr>
<td>650</td>
<td>1.31</td>
<td>5.8 ± 0.4</td>
</tr>
<tr>
<td>700</td>
<td>0.78</td>
<td>5.6 ± 0.4</td>
</tr>
<tr>
<td>750</td>
<td>0.47</td>
<td>5.2 ± 0.3</td>
</tr>
<tr>
<td>800</td>
<td>0.28</td>
<td>4.6 ± 0.3</td>
</tr>
<tr>
<td>850</td>
<td>0.16</td>
<td>4.0 ± 0.2</td>
</tr>
<tr>
<td>900</td>
<td>0.10</td>
<td>3.6 ± 0.2</td>
</tr>
</tbody>
</table>

**TABLE II.** Table of cross sections, \(\sigma\), and acceptances, \(\epsilon \pm \delta \epsilon\), for \(Z'\) and SM \(t\bar{t}\) events.

![Image 2](https://example.com/image2.png)

**FIG. 2.** Distributions of \(\sum E_T\) and \(M_{t\bar{t}}\) in the control region \(0.75 < N_{N_{\text{out}}} \leq 0.93\). To measure and set the confidence level intervals on resonant \(t\bar{t}\) production given the observed \(M_{t\bar{t}}\) spectrum, we start with the following likelihood,

\[
L(\vec{n}|\sigma, \bar{\nu}) = \prod_i \frac{\epsilon_i e^{-\mu_i}}{n_i!} \tag{2}
\]

which is the prior probability of observing \(\vec{n}\), where \(n_i\) is the observed number of events in the \(i\)th bin of the \(M_{t\bar{t}}\) distribution, and \(\mu_i\) is the expected number of events in the same bin and is given by \(\mu_i = \sigma_s (A_s - A_{s_{\text{cont}}}) T_{s_i} + \sigma_{t\bar{t}} (A_{t\bar{t}} - A_{t\bar{t}_{\text{cont}}}) T_{t\bar{t}_i} + N_{\text{QCD}} T_{QCD}\), which depends on the assumed signal cross section, \(\sigma_s\), the SM \(t\bar{t}\) cross section, \(\sigma_{t\bar{t}}\), the signal and SM \(t\bar{t}\) effective acceptances, \(A_s, A_{t\bar{t}}\), the number of expected QCD events, \(N_{\text{QCD}}\), and the fraction of events in \(i\)th \(M_{t\bar{t}}\) bin, \(T_{s_i}, T_{t\bar{t}_i}, T_{QCD}\) for the signal, SM \(t\bar{t}\) and QCD distributions, respectively. As the QCD background prediction is performed using the data sample itself, the presence of SM \(t\bar{t}\) and assumed signal events must be subtracted from QCD background estimation. We calculate the number of residual contamination events in QCD background prediction, using assumed and theoretical cross sections for signal and SM \(t\bar{t}\) events, and their effective acceptances for the residual contamination terms \(A_{s_{\text{cont}}}\) and \(A_{t\bar{t}_{\text{cont}}}\) by applying the tag rate matrix to the signal and SM \(t\bar{t}\) samples before the tagging requirement, and then.

We use Bayes’ theorem to connect the likelihood of the measurement to the posterior probability density, which is used to set the upper limits.

\[
p(\sigma|\vec{n}) = \int d\bar{\nu} p(\sigma, \bar{\nu}|\vec{n}) = \int d\bar{\nu} L(\vec{n}|\sigma, \bar{\nu}) \pi(\sigma, \bar{\nu})/p(\vec{n}) \tag{3}
\]

where \(\pi(\sigma, \bar{\nu})\) is the prior probability density, and \(p(\vec{n}) = \int d\bar{\nu} \int d\sigma L(\sigma, \bar{\nu}|\vec{n}) \pi(\sigma, \bar{\nu})\). There are two types of uncertainties we have to consider. The first type does not change the shape of the
$M_{t\bar{t}}$ distribution but only the acceptances. The second type affects both the shape and normalization; we refer as shape uncertainties. Uncertainties that do not change the shape of the templates (distributions) are incorporated as nuisance parameters and integrated over in Eq. (3). In this respect, Eq. (3) includes not only the statistical uncertainty of the data, but also the source of systematic uncertainties on: signal and SM $t\bar{t}$ acceptances, SM $t\bar{t}$ cross section, QCD normalization and integrated luminosity.

One of the shape uncertainties we consider is the jet energy scale corrections. After the jet energy corrections we are left with an uncertainty on the jet energy scale. A change in the jet energy scale modifies both the acceptances and the template shapes. To account for shape uncertainties and jet energy scale in particular, we generate a set of pseudoexperiments using the shifted templates and acceptances. This results in a shifted reconstructed cross section with respect to the nominal one. The mapping of this shift versus the input cross section provides an evaluation of the impact of the jet energy scale uncertainty at any given cross section. The complete list of the shape uncertainties sources we consider are those due to uncertainties in the jet energy scale, initial- and the final-state radiation, and proton and antiproton parton distribution functions.

Assuming that the nature of shape uncertainties follow a normal distribution, we convolute the posterior probabilities with a Gaussian whose shape uncertainties follow a normal distribution functions. Assuming that the nature of shape uncertainties follow a normal distribution, we convolute the posterior probabilities with a Gaussian whose width is equal to the quadrature sum of individual shape uncertainties.

After including the shape uncertainties, the posterior density function is used to define the upper and lower limits at any given confidence level (C.L.). If the lower limit is zero then the data is considered consistent with the SM at that level of confidence. We also extract the reconstructed cross section as the most probable value of the posterior density function. To obtain the sensitivity of the reconstruction algorithm we generate 1000 simulated experiments in the signal null hypothesis and extract the 95% C.L. expected upper limit, defined as the median of the upper limits distribution. This entire exercise is repeated for various resonance masses from 450 GeV/$c^2$ to 900 GeV/$c^2$. Together with the theoretical cross section versus mass curves these limits are used to exclude certain mass ranges.

In this analysis we consider the data gathered by CDF II between 2002-2008 at the Tevatron. The $M_{t\bar{t}}$ distribution for the 2086 events surviving the final event selection criteria is shown in Fig. 3 and is consistent with the SM expectations. The resulting $95\%$ C.L. upper limits on $\sigma(p\bar{p}\rightarrow Z') \cdot BR(Z'\rightarrow t\bar{t})$ as a function of $M_{t\bar{t}}$ are shown in Fig. 4 together with expected limits derived from pseudoexperiments that include the SM background hypothesis only. These limits can be used to exclude leptophobic topcolor resonance candidates with a mass less than 805 GeV/$c^2$ at $95\%$ C.L., assuming the width of the resonance is $\Gamma_{Z'} = 0.012 \, M_{Z'}$. The previous searches were performed in the lepton plus jets channel only, and the most recent results were conducted by CDF II [17] and D0 [18]. Using Tevatron data corresponding to 1 fb$^{-1}$ and 0.9 fb$^{-1}$ integrated luminosity respectively, they found no evidence for $t\bar{t}$ resonant production. For the same benchmark model of leptophobic topcolor $Z'$, the upper limits were set at 720 GeV/$c^2$ and 700 GeV/$c^2$ for CDF II and D0, respectively.

In conclusion, we have performed a search for a heavy resonance decaying into $t\bar{t}$ using data with 2.8 fb$^{-1}$ integrated luminosity in the all-jets channel. No evidence
for such a resonance is observed and we set upper limits on the production cross section times branching ratio at the 95% C.L. For one leptophobic topcolor production mechanism, we exclude masses up to 805 GeV/c^2.

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 World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; the Academy of Finland; and the Australian Research Council (ARC).