Measurement of the $t\bar{t}$ cross section using high-multiplicity jet events

(D0 Collaboration)

1 Universidad de Buenos Aires, Buenos Aires, Argentina
2 LAFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil
3 Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
4 Universidade Federal do ABC, Santo André, Brazil
5 Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil
6 University of Alberta, Edmonton, Alberta, Canada;
   Simon Fraser University, Burnaby, British Columbia, Canada;
   York University, Toronto, Ontario, Canada
   and McGill University, Montreal, Quebec, Canada
7 University of Science and Technology of China, Hefei, People's Republic of China
8 Universidad de los Andes, Bogotá, Colombia
9 Center for Particle Physics, Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
10 Czech Technical University in Prague, Prague, Czech Republic
11 Center for Particle Physics, Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
12 Universidad San Francisco de Quito, Quito, Ecuador
13 LPC, Université Blaise Pascal, CNRS/IN2P3, Clermont, France
14 LPSC, Université Joseph Fourier Grenoble 1, CNRS/IN2P3, Institut National Polytechnique de Grenoble, Grenoble, France
15 CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France
16 LAL, Université Paris-Sud, IN2P3/CNRS, Orsay, France
17 LPNHE, IN2P3/CNRS, Universités Paris VI and VII, Paris, France
18 CEA, Ifri, SPP, Saclay, France
19 IPHC, Université de Strasbourg, CNRS/IN2P3, Strasbourg, France
20 IPNL, Université Lyon 1, CNRS/IN2P3, Villeurbanne, France and Université de Lyon, Lyon, France
21 III. Physikalisches Institut A, RWTH Aachen University, Aachen, Germany
22 Physikalisches Institut, Universität Bonn, Bonn, Germany
23 Physikalisches Institut, Universität Freiburg, Freiburg, Germany
24 II. Physikalisches Institut, Georg-August-Universität Göttingen, Göttingen, Germany
25 Institut für Physik, Universität Mainz, Mainz, Germany
26 Ludwig-Maximilians-Universität München, München, Germany
27 Fachbereich Physik, University of Wuppertal, Wuppertal, Germany
28 Punjab University, Chandigarh, India
29 Delhi University, Delhi, India
30 Tata Institute of Fundamental Research, Mumbai, India
31 University College Dublin, Dublin, Ireland
32 Korea Detector Laboratory, Korea University, Seoul, Korea
33 SungKyunKwan University, Suwon, Korea
34 CINVESTAV, Mexico City, Mexico
35 FOM-Institute NIKHEF and University of Amsterdam/NIKHEF, Amsterdam, The Netherlands
36 Radboud University Nijmegen/NIKHEF, Nijmegen, The Netherlands
37 Joint Institute for Nuclear Research, Dubna, Russia
38 Institute for Theoretical and Experimental Physics, Moscow, Russia
39 Moscow State University, Moscow, Russia
40 Institute for High Energy Physics, Protvino, Russia
41 Petersburg Nuclear Physics Institute, St. Petersburg, Russia
We present a measurement of the $t\bar{t}$ cross section using high-multiplicity jet events produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. These data were recorded at the Fermilab Tevatron Collider with the D0 detector. Events with at least six jets, two of them identified as $b$ jets, were selected from a 1 fb$^{-1}$ data set. The measured cross section, assuming a top quark mass of 175 GeV/c$^2$, is 6.9 ± 2.0 pb, in agreement with theoretical expectations.

DOI: 10.1103/PhysRevD.82.032002

PACS numbers: 14.65.Ha

I. INTRODUCTION

The top quark is the most massive fundamental particle ever observed. Its mass, $m_t = 173.1 ± 1.3$ GeV/c$^2$ [1], is approximately twice that of the next heaviest elementary particle, the $Z$ boson, and is approximately 35 times that of its weak-isospin partner, the bottom quark. Top quarks are primarily produced in pairs at the Fermilab Tevatron $p\bar{p}$ Collider via the $q\bar{q} \rightarrow t\bar{t}$ ($\approx 85\%$) and $gg \rightarrow t\bar{t}$ ($\approx 15\%$) quantum chromodynamic (QCD) processes. They decay to...

*Visitor from Augustana College, Sioux Falls, SD, USA.
†Visitor from Rutgers University, Piscataway, NJ, USA.
‡Visitor from The University of Liverpool, Liverpool, UK.
§Visitor from SLAC, Menlo Park, CA, USA.
¶Visitor from Centro de Investigacion en Computacion-IPN, Mexico City, Mexico.
Visitor from ECFM, Universidad Autonoma de Sinaloa, Culiacán, Mexico.
**Visitor from Universität Bern, Bern, Switzerland.
††Visitor from Universität Zürich, Zürich, Switzerland.
a W boson and a b quark with a branching fraction near one according to the standard model (SM). The W boson subsequently decays into a lepton and a neutrino or into a quark-antiquark pair. The decay products of the W bosons are used to classify the top quark decay channel. The all-hadronic decay channel, with a branching fraction of 46% [2], has a final state containing two b quarks and four lighter quarks and is shown schematically in Fig. 1. The top quark might also decay into non-SM particles (e.g., a charged Higgs boson) and the decay products of these new particles can change the branching fractions of the leptonic and all-hadronic t\(\bar{t}\) decay channels [3]. Comparing the t\(\bar{t}\) production cross section between different decay channels directly constrains the existence of beyond the standard model particles lighter than the top quark.

In this paper, we present a new measurement of the protons. Kinematic selection criteria were applied to further improve the signal-to-background ratio to approximately 1:7 (Sec. III D). The t\(\bar{t}\) production cross section was extracted using signal and background templates for a likelihood discriminant constructed from topological and kinematic observables. (Sec. IV).

II. DETECTOR AND RECONSTRUCTION

A. Detector

The D0 detector [12] has a central-tracking system consisting of a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located within a 2 T superconducting solenoidal magnet, with designs optimized for tracking and vertexing at pseudorapidities \(|\eta| < 3\) and \(|\eta| < 2.5\), respectively [13]. Central and forward pre-shower detectors are positioned just outside of the superconducting coil. The liquid-argon and uranium calorimeter has a central section (CC) covering pseudorapidities \(|\eta| \lesssim 1.1\) and two end calorimeters (EC) that extend coverage to \(|\eta| = 4.2\), with all three housed in separate cryostats [14]. Each calorimeter contains a four-layer electromagnetic (EM) section closest to the interaction region, followed by finely- and coarsely-segmented hadronic sections. Scintillators between the CC and EC cryostats provide sampling of developing showers at 1.1 \(\leq |\eta| \leq 1.4\). The luminosity is measured using scintillators placed in front of the EC cryostats [15]. An outer muon system, covering \(|\eta| < 2\), consists of a layer of tracking detectors and scintillation trigger counters in front of 1.8 T iron toroids, followed by two similar layers beyond the toroids. The trigger and data acquisition systems were designed to accommodate the high luminosities of Tevatron Run II.

B. Trigger

The events used in this analysis were collected using a multijet trigger. The first level of the trigger used dedicated hardware and preliminary information from the calorimeter to identify multijet events. This selection was refined in a second level with more complex algorithms. The third trigger level employed a fast reconstruction of the event with a simple cone jet algorithm [16]. This selection was further refined using the final reconstruction algorithms which included the midpoint cone jet algorithm [16]. Kinematic and jet-multiplicity requirements were applied at each stage to reduce the overall data rate.

The trigger required at least four reconstructed jets, independent of whether the four jets were associated to a single vertex. This specific requirement is only applied offline, in the event selection, as discussed below. The specific trigger requirements on the jets, particularly the energy thresholds, were changed several times during data collection to cope with the increasing instantaneous lumini-
The simulated all-hadronic $t\bar{t}$ tracks are associated, on average, with primary vertices in events. A distribution of the primary vertex $z$ position with respect to the center of the detector in the triggered data. The solid line is a fit to the region with $|z_{PV}| < 35$ cm, while the dotted line is an extrapolation of the fit outside that region. Displayed error bars represent statistical uncertainties only. The distribution is normalized to a unit area.

**D. Jets**

Jets were reconstructed from energy deposits in calorimeter cells using the Run II midpoint cone algorithm [16] with a cone radius $R = \sqrt{\Delta \phi^2 + (\Delta \eta)^2} = 0.5$ [19]. Only calorimeter cells with energies $\geq 2.5$ standard deviations (s.d.) above the average electronic noise are included in the calculation of jet energies. Cells with energies between 2.5 and 4 s.d. of the electronic noise are included in the sum only if there is a neighboring cell with $\geq 4$ s.d. Jets were required to have $<40\%$ of their energy in the coarse hadronic calorimeter, have at least half the remaining transverse energy matched to energy depositions identified by the hardware trigger, and have between $5\%$ and $95\%$ of their energy in the EM calorimeter. These requirements were for jets reconstructed in the CC; they were looser for jets in the rapidity ranges covered by the EC calorimeters.

Jet energies were corrected for the energy response of the calorimeter, for the effect of particles showering outside the jet cone, for overlaps due to multiple interactions and event pileup, and for calorimeter noise [20]. The calorimeter response was measured using the $p_T$ imbalance in $\gamma +$ jet and dijet events; the response of the calorimeter to electromagnetic showers was calibrated using the $Z \rightarrow e^+e^-$ mass peak and a detailed accounting of the material between the calorimeter and the interaction point. The jet energy calibration also used $Z +$ jet events and events acquired using low bias triggers. Jets that contained

The fit extrapolation outside this range is also shown. The total primary vertex acceptance was $79.5 \pm 2.0\%$.
Jets that contain a $b$ hadron are called “$b$ jets” as they typically originate from a $b$ quark. $b$ hadrons have relatively long lifetimes and so usually travel several millimeters before they decay. Secondary vertices, displaced from the primary vertex, are usually formed by the tracks associated with the decay products of the $b$ hadron.

An artificial neural network (NN) was used to identify $b$ jets [21]. Selected characteristics of secondary vertices and tracks associated with $b$ hadron decays were used as inputs to the NN. These included aspects of the secondary vertex such as its decay length significance, goodness of fit, number of tracks, mass of the system of particles associated with the vertex, and the number of secondary vertices found in the jet. Additionally, the weighted combination of track impact parameter significances and the probability that the jet originated from the primary vertex were also input into the NN.

The probability to identify a $b$ jet, the tag rate function, was measured in data and parametrized as a function of the jet $p_T$ and $\eta$. Similar functions were determined for charm jets. The fake rate, the probability to assign a $b$ tag to a non-$b$ jet, was dominated by light jets and long-lived particles (e.g., $K^0_s$, $\Lambda^0$). The $b$-tagging efficiency is $(57 \pm 2)\%$, the tagging efficiency for charm is $(15 \pm 1)\%$, and the fake rate is $(0.57 \pm 0.07)\%$ for the NN output threshold used in this analysis at $p_T = 40$ GeV/$c$ [21].

III. ANALYSIS TECHNIQUES

A. Data sample

The data used for this analysis were collected between August 2002 and February 2006 with the four-jet trigger described in Sec. II B. Quality requirements were imposed on the selected data; runs or parts of runs in which detector systems essential to this analysis had problems or significant noise were discarded. The integrated luminosity of the data sample, including these trigger and quality requirements, is $0.97 \pm 0.06$ fb$^{-1}$.

B. Background model

QCD multijet events that have at least two heavy-flavor jets are the dominant source of background to $t\bar{t}$ production in the all-hadronic decay channel. This large background is distinguished from the $t\bar{t}$ signal by exploiting differences between the kinematic and topological distributions of jets in $t\bar{t}$ and multijet events. Correlations between jets, particularly for $b$ jets, must be reproduced for the observables used in this analysis.

The background sample was created using triggered data events. Signal contamination in the background sample was minimized by selecting events with two $b$-tagged jets and low jet multiplicities. Samples of events with at least four taggable jets having $p_T > 15$ GeV/$c$ were selected from the triggered data. The $b$-jet identification

muons, assumed to originate from $c$- or $b$-hadron decays, were corrected to account for the energy of the muon and the accompanying neutrino. Muons with reconstructed $p_T > 60$ GeV/c were treated as having $p_T = 60$ GeV/c to avoid the impact from poorly reconstructed muon momenta. Jet energies were calibrated independently in the data and in the simulation using the same methodology. Jets in the simulation required additional corrections to reproduce the reconstruction efficiency and energy resolution in the data. The uncertainty on the jet energy calibration is $\approx 1.5\%$.

Jets were further required to be matched with at least two good quality tracks having $p_T > 1$ GeV/$c$ and $p_T > 0.5$ GeV/$c$, respectively, that included SMT hits and pointed to the primary vertex. These requirements are termed “taggability” and are important for identifying heavy-flavor jets (Sec. II E) and to reject jets produced by overlapping $p\bar{p}$ collisions. The taggability fraction depends nominally on the jet $p_T$, jet rapidity, $z_{PV}$, $\text{sign}(z_{PV} \times \eta_{jet}) \times |z_{PV}|$, and the flavor of the jet [18]. The fraction of jets that were taggable was measured using the selected sample of multijet events (Sec. III D) and is shown in Fig. 4 binned in jet $p_T$. Differences between the taggability determined with multijet data and with the $t\bar{t}$ signal simulation could bias the cross section measurement. The $t\bar{t}$ simulation yielded the same taggability fraction as a function of jet $p_T$ and $\eta$. Similar functions were determined for charm jets. The fake rate, the probability to assign a $b$ tag to a non-$b$ jet, was dominated by light jets and long-lived particles (e.g., $K^0_s$, $\Lambda^0$). The $b$-tagging efficiency is $(57 \pm 2)\%$, the tagging efficiency for charm is $(15 \pm 1)\%$, and the fake rate is $(0.57 \pm 0.07)\%$ for the NN output threshold used in this analysis at $p_T = 40$ GeV/$c$ [21].

FIG. 4 (color online). Comparison of the taggability fraction in selected multijet data after the selection with that in the $t\bar{t}$ simulation as a function of jet $p_T$. Displayed error bars represent statistical uncertainties only.

032002-6
criteria described in Sec. II E were applied to these samples; events were kept if there were at least two tagged jets. Our procedure for building the multijet background estimation uses a relatively pure multijet sample, with little contamination from $t\bar{t}$ events. The background sample was created by attaching low-$p_T$ jets selected from events with six or more jets to events with four or five jets. A reasonable distribution of the jets in the available phase space was ensured using a set of matching criteria. We have first validated this model building a five-jets “background” sample that was then compared to five-jets “signal events.” This procedure relies on the fact that for background events the lowest $p_T$ jets are most likely coming from gluon radiation off one of the other jets, and in an high-multiplicity environment they are essentially oriented randomly relative to the rest of the event.

One concern with basing the background distributions on a lower jet-multiplicity sample was that the relative contributions of different production diagrams might depend strongly on jet multiplicity. This was tested by examining distributions of the $\Delta R$ between the $b$ jets. We expect a peak near $\pi$ for $b\bar{b}$ produced in $2\rightarrow 2$ hard scatterings, whereas we expect a peak near one (twice the jet radius) for $b\bar{b}$ produced via gluon splitting, $g \rightarrow b\bar{b}$. This is illustrated for four and five jet events in Fig. 5. Figure 5(a) shows $\Delta R_{bb}$ for $b$ jets with $p_T > 15 \text{ GeV}/c$ while Fig. 5(b) is the $\Delta R_{bb}$ for $b$ jets with $p_T > 40 \text{ GeV}/c$. The relative height of the two peaks depends strongly on the $p_T$ requirement, but there is little difference between four- and five-jet events. The gluon-splitting contribution is significantly suppressed by increasing the $b$-jet $p_T$ requirement from 15 to 40 GeV/c.

To validate the model a background sample was constructed by adding the lowest $p_T$ jet from five-jet events to four-jet events. The two sources of jets were matched together to ensure compatible phase-space configurations. The leading jets in each sample were required to have a difference in $p_T$ ($\Delta p_T$) within 1 GeV/c. Matches resulting in unphysical configurations (e.g., spatially overlapping jets) were rejected. The background event statistics were enhanced by running 20 times over the four- and five-jet samples. In each step the $\Delta p_T$ requirement was relaxed by 1 GeV/c.

One issue with this matching scheme is that an initial four-jet event might not have sufficient phase space for an additional jet. Since QCD multijet events are not expected to contain significant missing transverse energy ($\not{E}_T$), the presence of $\not{E}_T$ implies the presence of unreconstructed or mismeasured jets which makes these events more suitable for use in the background sample. However, badly misreconstructed events or events containing hard neutrinos can skew the phase space. Requiring the ratio of $\not{E}_T$ to $H_T \equiv \sum_{i=1}^{4} p_T i$ to be small reduced these contributions. Agreement between the “signal” and background five-jet samples was best with $\not{E}_T > 5 \text{ GeV}/c$ and $\not{E}_T/H_T < 0.1$. Variations in this additional phase-space selection were included in the systematic uncertainty evaluation [22].

The resulting events were compared with the five-jet sample as illustrated in Fig. 6. Reasonable agreement was achieved with the individual jet $p_T$ distributions and with their sum. These manufactured background events are also compared against the five-jet events for several topological variables (defined in Sec. III E) in Fig. 7.

Both the original four-jet sample used to create these five-jet background events and the signal five-jet sample to which it was compared had little contamination from $t\bar{t}$ (0.2% and 0.7%, respectively), so this tests our ability to use one multijet sample to create a representation of a higher-multiplicity sample. This scheme was extended to produce the background sample for events with six or more jets. In this case, the lowest $p_T$ jets were added to either four-jet (fifth and lower $p_T$ jets) or five-jet (sixth and lower $p_T$ jets) samples. There was no reason to prefer the four-jet-initiated background over the one built from a five-jet.

FIG. 5. $\Delta R$ between the two leading $b$-tagged jets in four-jet and five-jet events with (a) $p_T > 15 \text{ GeV}/c$; (b) $p_T > 40 \text{ GeV}/c$. The peak near $\Delta R \approx \pi$ is dominated by direct $b\bar{b}$ production while the peak near $\Delta R \approx 1$ (twice the jet radius) is mainly $g \rightarrow b\bar{b}$. Displayed error bars represent statistical uncertainties only. Distributions are normalized to a unit area.
sample. Instead, an equal mix of the two was used for the final background sample and the difference between the two separate background samples and the mixed sample was used when evaluating systematic uncertainties. Variations between the two samples as a function of $H_T$ are shown in Fig. 8. Also shown is the change in the background due to systematic variations in the phase-space matching criteria described above.

**C. Signal model**

The $t\bar{t}$ signal was simulated with the ALPGEN event generator. Two inclusive $t\bar{t}$ samples were used in this analysis: one with $m_t = 170 \text{ GeV}/c^2$ and one with $m_t = 175 \text{ GeV}/c^2$ [23]. PYTHIA, with the tune A parameter settings, was used for the parton shower, hadronization, and underlying event aspects. The resulting events were processed through a GEANT [24] simulation of the DØ detector and underwent the full reconstruction and analysis procedure. Information from data events selected by a random beam crossing trigger were overlayed on the simulated events to reproduce experimental conditions including detector noise and overlapping $p\bar{p}$ interactions. The instantaneous luminosity distribution of the simulated events was weighted to match that of the triggered data. Several additional corrections were applied to the simulated events. First, the event generator used the leading order parton distribution functions (PDF) from CTEQ6L1 [25,26]. Events were reweighted to correspond to the CTEQ6.5M [27] PDF. Second, the default heavy-flavor fragmentation function in PYTHIA was reweighted to one that described the LEP $e^+e^-$ data [28]. In addition, the resolutions for jet energies in the simulation were better than in data. Smaller differences were observed for electrons and muons. The energies of all reconstructed objects in simulations were smeared to reproduce the resolutions observed in data [29]. The jet identification efficiency was $\approx 0.5\%$ higher in the simulation than in data. Therefore, jets in the simulation were randomly removed to make the efficiencies agree.

**D. Event selection**

Selection criteria were applied to triggered events to minimize background while retaining a relatively high signal efficiency. The selection criteria, together with the
number of events after each cut, the cut efficiency $\epsilon$, and the cumulative selection efficiency $\epsilon_{\text{cum}}$, are presented in Table I. Values are given for the all-hadronic $t\bar{t}$ signal, for the signal in all other $t\bar{t}$ decay channels, and for the data-based background. The signal fraction in the final selected sample corresponded to a purity of 12.5% (as found in Sec. IVA). As the background was derived from triggered data, the minimum set of requirements on that sample, which also included a reconstructed primary vertex with $|z_{\text{PV}}| < 35$ cm and $\geq 4$ jets having $p_T > 15$ GeV/$c$, are listed as the second line in Table I. This corresponded to a starting signal-to-background ratio of approximately 1:7700.

Events with isolated high-$p_T$ electrons and muons were removed to avoid overlap with other D0 $t\bar{t}$ cross section measurements [30,31]. This requirement had little effect on the all-hadronic $t\bar{t}$ signal, but did remove a considerable number of events from the background.

Events considered in this analysis were required to have at least six jets. Each jet was required to be taggable, have $p_T > 15$ GeV/$c$, and $|\eta| < 2.5$. Furthermore, at least four of the jets were required to have $p_T > 40$ GeV/$c$. At least
two of these high-\(p_T\) jets were required to be \(b\) tagged. These additional jet requirements improve the signal-to-background ratio by a factor of 100.

In total, 1051 data events satisfy the selection criteria. The efficiency for all-hadronic \(t\bar{t}\) events with \(m_t = 175\) GeV/\(c^2\) is \((4.04 \pm 0.02)\%\) while the overall efficiency for inclusive \(t\bar{t}\) events is \((1.94 \pm 0.01)\%\) (statistical uncertainties only). The equivalent efficiencies with \(m_t = 170\) GeV/\(c^2\) are \((3.65 \pm 0.04)\%\) and \((1.76 \pm 0.02)\%\), respectively. Given these efficiencies and the standard model branching fractions, \(\approx 93\%\) of the selected \(t\bar{t}\) events are from the all-hadronic decay channel. The surviving leptonic \(t\bar{t}\) events were primarily from the \(\ell + \)jets (\(= 60\%)\) and \(\tau + \)jets (\(= 40\%)\) decay channels. Few dileptonic events survived the full selection criteria (\(\approx 0.05\%\) of \(t\bar{t}\)).

The expected signal-to-background ratio, given the 12.5\% signal purity extracted during the cross section measurement, is 1:7.

**E. Maximum likelihood**

A likelihood discriminant based on topological observables was constructed to separate the all-hadronic \(t\bar{t}\) signal from the multijet background. The likelihood ratio, \(L\), for an event \(i\) is defined as

\[
L = \frac{L_S(i)}{L_S(i) + L_B(i)},
\]

where

![Graphs showing systematic variations in the background sample with six or more jets as a function of \(H_T\).](image)

**FIG. 8 (color online).** Systematic variations in the background sample with six or more jets as a function of \(H_T\). (a) Comparisons with the background samples created using only four-jet (4 + 2) or five-jet (5 + 1) events. (b) Comparisons including 1 standard deviation systematic variations in the phase-space matching criteria. The leading four jets were required to have \(p_T > 40\) GeV/\(c\). Distributions are normalized to unit area.

### Table I. The number of events after each selection requirement. Each selection is inclusive of the ones above it. Shown are the criteria, the number of events that pass the selection, the efficiency of the selection (\(\epsilon\)), and the cumulative selection efficiency (\(\epsilon_{\text{cum}}\)) for all-hadronic \(t\bar{t}\), all other \(t\bar{t}\) decay channels, and the data-based background. The \(m_t = 175\) GeV/\(c^2\) sample was used for the signal expectation. Signal and background numbers have been adjusted, using the 12.5\% signal fraction measured in this analysis, to sum to the number of candidate events selected in the data. The last entry in this Table shows the efficiency for \(t\bar{t}\) events assuming the SM branching fractions and considering also the contributions from leptonic decays of the \(t\) quarks. Statistical uncertainties are included for the overall signal efficiency.

<table>
<thead>
<tr>
<th>Selection</th>
<th>All-Hadronic (t\bar{t})</th>
<th>Other (t\bar{t})</th>
<th>Background</th>
<th>Approx. S:B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Num. (\epsilon(%)) (\epsilon_{\text{cum}}(%))</td>
<td>Num. (\epsilon(%)) (\epsilon_{\text{cum}}(%))</td>
<td>Num. (\epsilon(%)) (\epsilon_{\text{cum}}(%))</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3024 100.0 100.0</td>
<td>3712 100.0 100.0</td>
<td>18856 263 100.0 100.0</td>
<td>1:7000</td>
</tr>
<tr>
<td>Trigger, vertex, (\geq 4) jets with (p_T &gt; 15) GeV/(c)</td>
<td>1663 55.0 55.0</td>
<td>773 20.8 20.8</td>
<td>12679 185 67.2 67.3</td>
<td>1:5700</td>
</tr>
<tr>
<td>Lepton veto</td>
<td>1662 100.0 100.0</td>
<td>558 72.2 15.0</td>
<td>12679 185 67.2 67.3</td>
<td>1:5700</td>
</tr>
<tr>
<td>(\geq 6) jets with (p_T &gt; 15) GeV/(c)</td>
<td>913 55.0 55.0</td>
<td>165 29.6 4.5</td>
<td>1734 595 13.7 9.2</td>
<td>1:1600</td>
</tr>
<tr>
<td>(\geq 6) taggable jets with (p_T &gt; 15) GeV/(c)</td>
<td>628 68.8 20.8</td>
<td>60 36.3 1.6</td>
<td>5062 777 29.2 2.7</td>
<td>1:740</td>
</tr>
<tr>
<td>(\geq 2) (b)-tagged jets with (p_T &gt; 40) GeV/(c)</td>
<td>150 23.8 4.9</td>
<td>13 21.8 0.4</td>
<td>2562 0.5 0.014</td>
<td>1:16</td>
</tr>
<tr>
<td>(\geq 3) jets with (p_T &gt; 40) GeV/(c)</td>
<td>147 98.1 4.9</td>
<td>12 95.2 0.3</td>
<td>2059 80.4 0.011</td>
<td>1:13</td>
</tr>
<tr>
<td>(\geq 4) jets with (p_T &gt; 40) GeV/(c)</td>
<td>122 83.2 4.0</td>
<td>9 70.3 0.2</td>
<td>920 44.7 0.0049</td>
<td>1:7</td>
</tr>
<tr>
<td>Efficiency</td>
<td>((4.04 \pm 0.02)%)</td>
<td>((0.24 \pm 0.01)%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusive (t\bar{t}) Efficiency</td>
<td>((1.94 \pm 0.01)%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
for signal and similarly for background. Here, \( P_{S,k} \) is the signal probability density function, normalized to unit area, for the \( k \)th input variable \( x_k \), and \( n_{\text{var}} \) is the number of variables. The TMVA [32] package was used to build the probability distributions and the resulting likelihood ratio.

The criteria for the selection of observables to be input into the likelihood were: separation between signal and background, reasonable agreement in the five-jet background validation, little correlation with other chosen variables, and little dependence on jet energies (to minimize systematic uncertainty due to jet energy calibration). The following nine variables were used in the likelihood determination and are shown for simulated signal and data-based background events in Fig. 9:

- \( C \) is the centrality defined as the scalar sum of jet \( p_T \) divided by the sum of jet energies;
- \( H'_{T} \) is the scalar sum of jet \( p_T \) excluding the two highest \( p_T \) jets;
- \( B \) is the ratio of the dijet mass of the two leading \( b \)-tagged jets to the total mass of all the jets;

![Probability Distributions](image)

**FIG. 9** (color online). Probability distributions for the variables input into the likelihood ratio. The signal distributions were extracted from the sample with \( m_t = 175 \text{ GeV}/c^2 \). Displayed error bars represent statistical uncertainties only.
\[ \lambda_2, \lambda_3 \] are the smallest two eigenvalues of the momentum tensor \( M^{\alpha\beta} = \sum_i p_i^\alpha p_i^\beta / \sum_i |\hat{p}_i| \) where \( i \) runs over the number of jets and \( \alpha, \beta = 1, 2, 3 \) denote the three spatial components of the jet momenta [33];

\( y_{34} \) is the rapidity difference between the third and fourth leading jets;

\( A_{234} \) is the \( p_T \) asymmetry between the second and third jet and the fourth jet defined as \( (p_{T2} + p_{T3} - p_{T4})/(p_{T2} + p_{T3} + p_{T4}) \);

\( \langle y_b \rangle \) is the \( p_T \)-weighted average of the rapidities of the leading two \( b \)-tagged jets;

\( \langle y_l \rangle \) is the \( p_T \)-weighted average of the rapidities of the leading two light (not \( b \)-tagged) jets.

The \( y_{34}, \langle y_b \rangle, \) and \( \langle y_l \rangle \) variables exploit the difference in the correlations between jets in \( t\bar{t} \) events and the multijet background. The third and fourth jets in \( t\bar{t} \) events tend to come from the decay of the same \( W \) boson and are, therefore, close in rapidity. The two leading \( b \)-tagged jets are mostly central and more back-to-back in azimuthal angle than for QCD production of \( b\bar{b} \) events, and the same happens also for the two leading not \( b \)-tagged jets. The \( A_{234} \) variable exploits the fact that the asymmetry between the two jets from the \( W \) decay, which are assumed to be the third and fourth jet in the event, is small. To allow for cases in which the second jet is not \( b \)-tagged the second and third jet are treated equally in the calculation of this asymmetry.

Comparisons are shown in Fig. 7 for these variables in the five-jet background validation sample. The combined probability distributions for signal and background are shown in Fig. 10. The probability distributions and likelihoods were extracted independently for the \( m_t = 170 \text{ GeV}/c^2 \) and \( 175 \text{ GeV}/c^2 \) samples.

IV. RESULTS

A. Signal fraction

The signal and background likelihood templates were fit to the likelihood output, shown in Fig. 10, for the selected
data events using TMINUIT [34] from ROOT [35]. Results from the fit are shown in Fig. 11 and are in agreement with the data. The measured signal fractions are $(12.9 \pm 2.4)\%$ for $m_t = 170 \text{ GeV}/c^2$ and $(12.5 \pm 2.3)\%$ for $m_t = 175 \text{ GeV}/c^2$. Given 1051 data candidate events, this results in 136 and 131 $t\bar{t}$ events, respectively. Distributions for the observables included in the likelihood, using the signal and background fractions from the fit, are shown in Fig. 12 for $m_t = 175 \text{ GeV}/c^2$. There is reasonable agreement between the data candidates and the sum of signal and background normalized to the fit results.

Jets in an event can be associated with the decays of individual top quarks. A $\chi^2$ was constructed comparing the dijet masses with the $W$ boson mass and the two $b\bar{b}$ masses with each other. The combination with the lowest $\chi^2$ value was chosen. The results for the dijet mass and the $b\bar{b}$ mass are shown in Figs. 13(c) and 13(d). There is good agreement between data and the sum of signal and background. The comparison is also made in a region of phase space dominated by background ($L < 0.2$) and one which has a significantly larger signal fraction ($L > 0.8$), also shown in Fig. 13. The distributions were not renormalized.
Both the background-dominated and signal-enhanced distributions show reasonable agreement between data and the sum of signal and background.

**B. Systematic uncertainties**

The effects of systematic uncertainties and variations in input variables were studied using ensemble tests. Ten thousand pseudoexperiments were run for each source of uncertainty. Each pseudoexperiment drew events from the systematically shifted signal and background distributions and was fit using the standard signal and background likelihood templates. With the exception of the two background-related systematics, all of the systematic uncertainties are associated with the signal simulation only. All systematic uncertainties on the $t\bar{t}$ production cross section measured with $m_t = 175 \text{ GeV}/c^2$ are summarized in Table II. Many of these are described in more detail in earlier sections of this paper.

This analysis relies on ALPGEN + PYTHIA for the $t\bar{t}$ signal model used to determine the selection efficiency (Table I) and the kinematic shapes included in the likelihood determination (Fig. 9). It is possible that the $t\bar{t}$ simulation does not properly reproduce the properties of the $t\bar{t}$ system. Other analyses in the lepton + jets and dilepton decay channels published by the D0 Collaboration have found good agreement between the simulation and the reconstructed data [29–31,36]. Nevertheless, the simulation might misestimate the jet multiplicity.

<table>
<thead>
<tr>
<th>Source</th>
<th>Uncertainty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate statistics</td>
<td>±18.5</td>
</tr>
<tr>
<td>Background model</td>
<td>±10.7</td>
</tr>
<tr>
<td>Background model statistics</td>
<td>±3.8</td>
</tr>
<tr>
<td>Signal model</td>
<td>±3.2</td>
</tr>
<tr>
<td>Signal model statistics</td>
<td>±0.5</td>
</tr>
<tr>
<td>Trigger</td>
<td>$-2.0 + 3.9$</td>
</tr>
<tr>
<td>Jet identification efficiency</td>
<td>$-2.5 + 3.0$</td>
</tr>
<tr>
<td>Jet taggability</td>
<td>±8.8</td>
</tr>
<tr>
<td>Jet energy calibration</td>
<td>±10.8</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>$-3.1 + 2.2$</td>
</tr>
<tr>
<td>$b$ tagging</td>
<td>$-8.6 + 9.2$</td>
</tr>
<tr>
<td>Total statistical uncertainty</td>
<td>±18.9</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>±20.5</td>
</tr>
<tr>
<td>Luminosity uncertainty</td>
<td>±6.1</td>
</tr>
</tbody>
</table>

**TABLE II.** Uncertainties on the $t\bar{t}$ cross section categorized by source for the result corresponding to $m_t = 175 \text{ GeV}/c^2$. The uncertainties with $m_t = 170 \text{ GeV}/c^2$ are similar.
The total systematic uncertainty is 20.5%. The dominant sources of systematic uncertainty in the cross section measurement are the jet energy calibration (10.8%), construction of the data-based background (10.7%), $b$ tagging (9.2%), and jet taggability (8.8%). The total systematic uncertainty is 20.5%.

### C. Cross section measurement

The cross section is defined as

$$\sigma_{t\bar{t}} = \frac{fN}{L\varepsilon}$$

where $f$ is the measured fraction of $t\bar{t}$ signal, $N$ is the number of selected data events, $L$ is the integrated luminosity, and $\varepsilon$ is the inclusive $t\bar{t}$ efficiency given in Table I. This results in the following cross sections:

$$\sigma_{t\bar{t}}^{170 \text{ GeV/c}^2} = 7.9 \pm 1.5 \text{ (stat)} \pm 1.6 \text{ (sys)} \pm 0.5 \text{ (lum) pb},$$

$$\sigma_{t\bar{t}}^{175 \text{ GeV/c}^2} = 6.9 \pm 1.3 \text{ (stat)} \pm 1.4 \text{ (sys)} \pm 0.4 \text{ (lum) pb}.$$
V. M. ABAZOV et al.

[13] Pseudorapidity $\eta$ is defined as $\eta = -\ln(\tan(\theta/2))$, where $\theta$ is the polar angle with respect to the proton beam direction, with its origin at the center of the detector.
[19] Rapidity is defined as $y = -\ln[(E + p_z)/(E - p_z)]$, where $E$ is the energy and $p_z$ is the momentum along the proton beam direction. $\phi$ is defined as the azimuthal angle in the plane transverse to the proton beam direction.
[22] $E_T/H_T$ was varied between 0.12 and 0.08, covering 68% C.L., to allow more or less phase space for additional jets.
[23] Two top quark mass values were considered since the world-average mass varied between 170.9 ± 1.8 GeV/c² and 174.2 ± 3.4 GeV/c² during the time period covered by this analysis.
[33] Aplanarity is defined as $\langle \lambda \phi \Delta \phi \rangle$ and sphericity as $\langle \lambda^2 \phi \Delta \phi \rangle$.