First measurement of the fraction of top-quark pair production through gluon-gluon fusion
We present the first measurement of $\sigma(gg \rightarrow t\bar{t})/\sigma(p\bar{p} \rightarrow t\bar{t})$. We use 0.96 fb$^{-1}$ of $\sqrt{s} = 1.96$ TeV $p\bar{p}$ collision data recorded with the CDF II detector at Fermilab. Using charged particles with low transverse
Many studies have been dedicated to the understanding of the top quark, motivated in part by its large mass that may give it a unique role in the generation of mass for the quarks, leptons, and force carriers in the standard model (SM) of particle physics. In \( p\bar{p} \) collisions at a center-of-momentum energy of \( \sqrt{s} = 1.96 \) TeV, (15 ± 5)% of \( t\bar{t} \) pairs are expected to be produced through gluon-gluon fusion and the rest through quark-antiquark annihilation [1,2], based on next-to-leading-order (NLO) quantum chromodynamics (QCD) calculations. The inclusive \( t\bar{t} \) production cross section has been measured by both CDF [3,4] and D0 [5] collaborations using various methods and decay modes of the \( t\bar{t} \) pairs, and the results are in agreement with SM predictions. However, the details of the production process have never been investigated.

Here we report the first measurement of the fractional cross section of \( t\bar{t} \) production through gluon-gluon fusion, \( \sigma(gg \rightarrow t\bar{t})/\sigma(p\bar{p} \rightarrow t\bar{t}) \). For the first time in high-energy hadron-hadron collisions, two processes with identical final states are experimentally discriminated based on their initial state differences. To discriminate between the similar final state signatures of \( gg \rightarrow t\bar{t} \) and \( q\bar{q} \rightarrow t\bar{t} \), we take advantage of the higher probability for a gluon than for a quark to radiate a low-momentum gluon [6]. Therefore, on average we expect a larger low-momentum charged particle multiplicity in \( gg \rightarrow t\bar{t} \) compared to \( q\bar{q} \rightarrow t\bar{t} \). Given the large theoretical uncertainties associated with gluon radiation, we do not rely on theoretical calculations for the modeling of the charged particle multiplicity. Instead, we use two different processes, \( W + n\)-jet and two-jet (dijet) production, with well-understood production mechanisms, as calibration samples to relate the observed charged particle multiplicity to the fraction of processes involving more gluons [7].

We use a data sample of \( \sqrt{s} = 1.96 \) TeV \( p\bar{p} \) collisions with an integrated luminosity of \( 0.96 \pm 0.06 \) fb\(^{-1} \) recorded by the CDF II detector at Fermilab between March 2002 and February 2006. The CDF II detector is described in detail in [8]; here, we briefly discuss the components essential for this analysis. The detector consists of a tracking system immersed in a solenoidal magnetic field of 1.4 T and electromagnetic and hadronic calorimeters surrounding the solenoid, followed by the muon system. Electrons, photons, and hadronic jets are identified using calorimeters and the tracking information. Muons are identified by the muon system together with tracking and calorimeter information. The data are collected using a three-level trigger system.

According to the SM top quarks almost always decay to a \( W \) boson and a bottom quark, and so in \( t\bar{t} \) events we expect to have two \( W \) bosons and two \( b \) quarks. We select \( t\bar{t} \) candidate events where one of the \( W \) bosons decays to two jets and the other decays to a lepton (\( l \)) and the corresponding neutrino. In this analysis \( l \) is either an electron or a muon. Our first calibration data set is a set of \( W(\rightarrow l\nu) + n\)-jet \((n = 0, 1, 2, 3)\) candidate events, for which the number of gluons involved in the production process increases with the number of jets [9]. The second is a set of events with two back-to-back, high-energy jets. The average number of gluons involved in dijet production [10] falls with increasing transverse energy \( (E_T) \) [11] of the highest \( E_T \) jet (leading jet), as the relative rate of the \( qq \rightarrow qq, qg \rightarrow qg \), and \( gg \rightarrow gg \) subprocesses change. The number of gluons in each subprocess is 0, 2, and 4, respectively, as we count the gluons regardless of their being in the initial or in the final state. Similarly, in the case of the \( W + n\)-jet processes, the \( W + 0\)-jet process has no gluon, the \( W + 1\)-jet process has one gluon which can be either in the initial or in the final state, and the \( W + \geq 2\)-jet processes have larger number of gluons involved in the production process.

The \( W + n\)-jet data are collected with an inclusive lepton trigger that requires an electron with \( E_T > 18 \) GeV or a muon with \( p_T > 18 \) GeV/c. We select events with a reconstructed isolated electron (muon) candidate with \( E_T > 20 \) GeV \((p_T > 20 \) GeV/c\) and a missing \( E_T \) \((E_T) > 20 \) GeV. We categorize the \( W + n\)-jet samples by \( n \), the number of jet candidates with \( E_T > 15 \) GeV and pseudorapidity \( |\eta| < 2 \). Jets are defined using an iterative cone algorithm [12] with a cone of \( \Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2} = 0.4 \) and are corrected for absolute energy response, \( \eta \) dependence of calorimeter response, and multiple interactions. For the \( t\bar{t} \) data sample, in addition to the above, we require four or more jets where at least one is identified as originating from a \( b \) quark (\( b \) tag). To define a \( b \) tag, we identify within a jet a long-lived \( B \)-hadron candidate through the presence of a displaced secondary vertex [3]. In both \( t\bar{t} \) and \( W + n\)-jet samples, we remove any event with a second lepton candidate consistent with arising from a \( Z \) boson decay or a \( t\bar{t} \) event in which both \( W \) bosons
decay to leptons. We also veto the events in which the electron (muon) is consistent with coming from a conversion photon (cosmic ray) [3]. The dijet data are collected using two inclusive jet triggers that require a jet with $E_T$ of at least 50 GeV or at least 100 GeV (Jet50 and Jet100 data sets). We require a minimum leading jet $E_T$ of 75 and 130 GeV for Jet50 and Jet100 data sets, respectively, to avoid any trigger bias. We remove events containing an electron (muon) candidate with $E_T > 20$ GeV ($p_T > 20$ GeV/c). We also require exactly two jets with $|\eta| \leq 2$ and a minimum $E_T$ of 20 GeV and with the two jets back to back, having a $|\Delta \phi| \geq 2.53$ rad. 

The background processes in our $\bar{t}t$ sample consist of $W + j$ events, electroweak processes ($WW$, $WZ$, $ZZ$), single top quark, and multijet QCD processes (non-$W$). For non-$W$ and $W + j$ background, we can have a real $b$ tag (heavy flavor background, HF) or have a $b$ tag due to misidentification (light flavor background, LF). We estimate LF and HF in events with a real $W$ boson using various calibration data sets. For the small fraction of events from non-$W$ sources, we assume the non-$W$ background is equal parts HF and LF. The results of the analysis are insensitive to this assumption. Single top-quark processes are part of HF, while diboson backgrounds, ignoring the few $Z \rightarrow b\bar{b}$ events, are included in LF. We find 240 $\bar{t}t$ candidates with an estimated background contamination of $(13 \pm 2\%)$. The background estimates are found using the method explained in [3]. 

The number of low-$p_T$ charged particles $N_{\text{trk}}$ is affected by low-energy particles arising from jet fragmentation as well as multiple interactions within the same $p\bar{p}$ bunch crossing. The track is in our definition of $N_{\text{trk}}$, we require it to have a $p_T$ in the range $0.3–3$ GeV/c and $|\eta| \leq 1.1$, to have a reliable and efficient track reconstruction, and to originate from the vertex associated with the charged leptons and jets. We reject the track if it falls within $\Delta R = 0.6$ and $\Delta R = 0.4$ of jets with $E_T \geq 15$ GeV (high-$E_T$ jets) and $6 \leq E_T < 15$ GeV (low $E_T$ jets), respectively. Excluding these tracks results in a different available tracking area for each event. We therefore correct the observed multiplicity to the total tracking coverage in $\eta$ and $\phi$ event by event. The resulting track multiplicity still has a modest dependence on the number of high $E_T$ jets in the event. We therefore make a further correction to $N_{\text{trk}}$ by measuring this dependence in multijet QCD candidate events and using this as a per-jet correction ($\sim 1$ track per jet) to the multiplicity for all jets with $|\eta| \leq 1.1$.

We show that there is a correlation between the average number of low-$p_T$ charged particles $\langle N_{\text{trk}} \rangle$ and the average number of gluons involved in the production process $\langle N_g \rangle$ in a given sample. We count the number of high-energy gluons involved in the production process using Monte Carlo (MC) calculations for both dijet and the $W + n$-jet data samples, where we only consider the incoming and outgoing high-energy gluons participating in the production process. The $W + n$-jet MC sample is created using the ALPGEN [13] program followed by PYTHIA [14] to perform the jet fragmentation. The MC dijet events are created using the PYTHIA MC. We plot the observed $\langle N_{\text{trk}} \rangle$ in data against the expected $\langle N_g \rangle$ from MC calculations for the calibration samples in Fig. 1. This demonstrates an approximately linear dependence between $\langle N_{\text{trk}} \rangle$ and $\langle N_g \rangle$. We do not use this plot to obtain our result, but rather directly fit the observed $N_{\text{trk}}$ distributions as described below.

The $\langle N_{\text{trk}} \rangle$ and $\langle N_g \rangle$ correlation enables us to define $N_{\text{trk}}$ distributions each representing a specific average number of gluons involved in the production process. We use this correlation and the observed $N_{\text{trk}}$ distributions in the $W + 0$-jet sample and the dijet sample with leading jet $E_T$ of 80–100 GeV to define a no-gluon and a gluon-rich $N_{\text{trk}}$ distribution, respectively. The $W + 0$-jet sample is largely composed of the Drell-Yan $q\bar{q}$ process with a small QCD background of order 4% and contribution from $W$ production in association with other partons where none of the final state jets are detected. The fraction of $W + 0$-jet candidates with production processes involving gluons is estimated to be $(5 \pm 4\%)$. The no-gluon contribution of dijet candidates with leading jet $E_T$ of 80–100 GeV comes from $qq \rightarrow qq$ processes and is estimated to be $(27 \pm 3\%)$. An iterative procedure is adopted in order to remove the no-gluon (gluon-rich) contribution from the $N_{\text{trk}}$ distribution of the 80–100 GeV dijet ($W + 0$-jet) sample. We start with the normalized (to unity) dijet 80–100 GeV and $W + 0$-jet $N_{\text{trk}}$ distributions. We subtract the normalized $W + 0$-jet $N_{\text{trk}}$ distribution from the normalized dijet sample $N_{\text{trk}}$ distribution with a factor of 0.27. Afterward, we normalize

![FIG. 1. The correlation between the average number of low-$p_T$ charged particles (data) and the average number of gluons (MC). The dotted line is from a linear fit to the points.](image-url)
the subtracted dijet sample distribution to unity and subtract it from the originally normalized $W + 0$-jet sample with a factor of 0.05. We then iterate this procedure. There are no significant changes in the distributions after the first iteration. Figure 2 shows the comparison between the no-gluon and gluon-rich parametrization. The $\langle N_g \rangle$ of the gluon-rich $N_{\text{trk}}$ distribution, defined as described above, is comparable to the $\langle N_g \rangle$ of the $gg \rightarrow t\bar{t}$ process.

To verify that the no-gluon or gluon-rich distribution can model the $N_{\text{trk}}$ distribution of any process with comparable $\langle N_g \rangle$ regardless of the center-of-momentum energy, we check the $W + 1$-jet data sample, and we see no dependence on jet $E_T$ in $\langle N_{\text{trk}} \rangle$. We also compare the $N_{\text{trk}}$ distribution of dijet 80–100 GeV, with $\langle N_g \rangle$ of ~2.4, with the $N_{\text{trk}}$ distribution of dijet events with leading jet $E_T$ of at least 180 GeV, with $\langle N_g \rangle$ of ~2.1 and see negligible differences. This is similar to the case of $\langle N_g \rangle$ of dijet 80–100 GeV and the $\langle N_g \rangle$ of the $gg \rightarrow t\bar{t}$ process. Therefore, we can use the no-gluon and gluon-rich distributions to model the $q\bar{q} \rightarrow t\bar{t}$ and $gg \rightarrow t\bar{t}$, respectively.

The gluon-rich fraction associated with a given $N_{\text{trk}}$ distribution can be found using a binned likelihood fit of the $N_{\text{trk}}$ distribution of the form

$$N [f_g F_g (N_{\text{trk}}) + (1 - f_g) F_q (N_{\text{trk}})]$$

where $N$ is the normalization factor and one of the free parameters, $f_g$, is the fraction of gluon-rich components of the sample and the second free parameter, and $F_g (N_{\text{trk}})$ and $F_q (N_{\text{trk}})$ are the normalized gluon-rich and no-gluon parametrizations, respectively. We check this technique is free of bias using 1000 simulated experiments for each true gluon-rich fraction, every 10% between (5–95)%.

To produce these simulated experiments we randomly generate $N_{\text{trk}}$ distributions samples from the no-gluon and gluon-rich $N_{\text{trk}}$ distributions based on the given true gluon-rich fraction. We then measure $f_g$ for each simulated experiment and look at the measured $f_g$ distribution for each true gluon-rich fraction. We do not see any bias. To verify that the method works well, we measure the $f_g$ in dijet data samples. Table I shows the comparison between the measured $f_g$ in data calibration samples and the MC predictions. The good agreement between data and MC calculations confirms that the method works well.

The $N_{\text{trk}}$ distribution of the $t\bar{t}$ candidates, shown in Fig. 3, has a mean of 10.6 ± 0.5. The fit, shown in the figure, models the data distribution very well, based on a goodness of fit test with 92% probability. The measured gluon-rich fraction in $t\bar{t}$ candidates determined by fitting

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Leading jet $E_T$ & MC expectation & $f_g$ from fit to data \\
\hline
80–100 GeV & $0.73 \pm 0.03$ & $0.73 \pm 0.01$ \\
100–120 GeV & $0.69 \pm 0.03$ & $0.69 \pm 0.01$ \\
120–140 GeV & $0.63 \pm 0.04$ & $0.66 \pm 0.01$ \\
140–160 GeV & $0.57 \pm 0.04$ & $0.63 \pm 0.01$ \\
160–180 GeV & $0.52 \pm 0.04$ & $0.57 \pm 0.01$ \\
\geq 180 GeV & $0.42 \pm 0.05$ & $0.49 \pm 0.01$ \\
\hline
\end{tabular}
\caption{The fraction of gluon-rich events in each sample as predicted by MC calculations and the fraction of gluon-rich events as found using the likelihood fit to track multiplicity distributions in dijet calibration samples. Uncertainties for the MC fractions include both statistical and systematical contributions. The uncertainties on the fit results to the data are only statistical.}
\end{table}
the $N_{u_k}$ distribution consists of two components, the $t\bar{t}$ gluon-rich fraction and the background gluon-rich fraction. Therefore, knowing the background fraction in our sample $f_b$ and the measured $f_g$ from the fit, we can write

$$f_g = f_b f_g^{bkg} + (1 - f_b) f_g^{tt},$$

where $f_g^{bkg}$ and $f_g^{tt}$ are the gluon-rich fraction of the background and $t\bar{t}$ signal, respectively.

The $\langle N_g \rangle$ for each background process is unique to that process and is not necessarily the same as the $\langle N_g \rangle$ of the $gg \to t\bar{t}$ process. However, for this analysis, we do not need to know the details of each background process. We only need to know the total contribution of all the background processes to the measured $f_g$, the first term in the right-hand side of Eq. (2). We note that the $N_{u_k}$ distributions can be empirically characterized as the superposition of different $N_{u_k}$ distributions with different $\langle N_g \rangle$. For example, in the case of a sample of $\langle N_g \rangle$ of 1, one can have 50% $N_{u_k}$ distribution with $\langle N_g \rangle$ of 0 and 50% $N_{u_k}$ distribution with $\langle N_g \rangle$ of 2. To estimate $f_g^{bkg}$, we measure $f_g$ in the $W + 1$, $W + 2$, and $W + 3$-jet data samples with no $b$ tag and with at least one $b$ tag using the fit to the $N_{u_k}$ distribution for each of these six samples. We then extrapolate the $f_g$ values from the $W + 1$, $W + 2$, and $3$-jet samples to $W + 4$ or more jet bins for $b$-tag and no $b$-tag samples separately. We note that the extrapolation from $W + 1$, $W + 2$, and $W + 3$ jets to $W + 4$ jets is based on the assumption of a linear evolution of the gluon content of the $W + n$ jet and of the QCD background from $W + 1$ jet to $W + 4$ jets. We consider the $b$-tag sample as representative of HF and the no $b$-tag sample as representative of LF. Using these extrapolations and the LF and HF fractions, we find $f_g^{bkg} = 0.54 \pm 0.09$ and $f_g^{tt} = 0.09 \pm 0.16$.

Given $f_g^{tt}$, we measure $\sigma(gg \to t\bar{t})/\sigma(p\bar{p} \to t\bar{t})$ as

$$\left[1 - \frac{A_{gg}}{A_{gq}} + \left(\frac{A_{gg}}{A_{gq}}\left(\frac{1}{f_g^{tt}}\right)\right)^{-1}\right]^{-1} = 0.07 \pm 0.14\text{(stat)},$$

where $A_{gg}$ and $A_{gq}$ are the acceptance for $gg \to t\bar{t}$ and $q\bar{q} \to t\bar{t}$, respectively. Using PYTHIA MC calculations, we find $(14.1 \pm 0.5)\%$ and $(11.5 \pm 0.4)\%$ for $A_{gg}$ and $A_{gq}$, respectively. The acceptance uncertainties include the systematic uncertainties. The result is equivalent to a $\sigma(q\bar{q} \to t\bar{t})/\sigma(p\bar{p} \to t\bar{t})$ of 0.93 $\pm 0.14\text{(stat)}$.

The systematic uncertainties of this measurement, a total of 0.07, arise from uncertainties in the measurement of $N_{u_k}$ and the subsequent calculations. The uncertainties in $N_{u_k}$ are due to the per-jet correction (0.05), the estimated gluon content of the $W + 0$-jet sample (0.04), and the choice of the low $E_T$ jet cut (0.02). In addition to these sources, there are uncertainties associated with the estimated $qg \to qg$ fraction of the 80–100 GeV dijet sample, the background fraction, the modeling of the background gluon-rich fraction, the non-$W$ background fraction, and the acceptances; these are all negligible. To estimate the effects of all the above uncertainties, we changed the central values and measured the change in the relevant variables. Given the fact that data from the same data-taking period is used for both calibration and $t\bar{t}$ samples, no systematic uncertainty is associated to the effects of particles produced within the same $p\bar{p}$ collision, instantaneous luminosity, multiple interactions, or the track reconstruction.

The result corresponds to an upper limit of 0.33 at 95% confidence level. We use a classical statistical technique to set the limit by simulating the possible outcomes for a given true value taking into account the systematic effects.

In conclusion, we have presented the first measurement of $\sigma(gg \to t\bar{t})/\sigma(p\bar{p} \to t\bar{t})$ and found $0.07 \pm 0.14\text{(stat)} \pm 0.07\text{(syst)}$, corresponding to an upper limit of 0.33 at 95% confidence level, in 0.96 fb$^{-1}$ of data collected at CDF. This is in agreement with the SM prediction of 0.15 $\pm$ 0.05, and does not suggest that non-SM processes [15] contribute to top-quark pair production at the Tevatron.

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 the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; the European Community’s Human Potential Programme; the Slovak R&D Agency; and the Academy of Finland.

[1] M. Cacciari et al., J. High Energy Phys. 04 (2004) 068.
[2] N. Kidonakis and R. Vogt, Phys. Rev. D 68, 114014 (2003); E. L. Berger and H. Contopanagos, Phys. Lett. B 361, 115 (1995); W. Bernreuther et al., Nucl. Phys. B690, 81 (2004).
[3] D. Acosta et al. (CDF Collaboration), Phys. Rev. D 71,
052003 (2005).

[4] D. Acosta et al. (CDF Collaboration), Phys. Rev. D 71, 072005 (2005); 72, 032002 (2005); Phys. Rev. Lett. 93, 142001 (2004); A. Abulencia et al. (CDF Collaboration), Phys. Rev. D 74, 072006 (2006); 74, 072005 (2006); Phys. Rev. Lett. 96, 202002 (2006).

[5] V. M. Abazov et al. (D0 Collaboration), Phys. Rev. D 74, 112004 (2006); 76, 052006 (2007); 76, 092007 (2007); 76, 072007 (2007).

[6] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. 15, 438 (1972); G. Altarelli and G. Parisi, Nucl. Phys. B126, 298 (1977); P. Abreu et al. (DELPHI Collaboration), Z. Phys. C 70, 179 (1996); G. Abbiendi et al. (OPAL Collaboration), Eur. Phys. J. C 11, 217 (1999).

[7] For details see S. Pashapour, Ph.D. thesis, [Report No. FERMILAB-THESIS-2008-11], 2008.

[8] D. Acosta et al. (CDF Collaboration), Phys. Rev. D 71, 032001 (2005).

[9] F. A. Berends, H. Kuijf, B. Tausk, and W. T. Giele, Nucl. Phys. B357, 32 (1991).

[10] B. L. Combridge, J. Kripfganz, and J. Ranft, Phys. Lett. B 70. 234 (1977).

[11] We use a coordinate system with an origin at the center of the detector, $\phi$ and $\theta$ as the azimuthal and polar angles, respectively, and a $z$ axis along the proton beam direction. The transverse energy is defined as $E_T = E \sin \theta$ and particle momentum transverse to the beam is $p_T = p \sin \theta$. The missing $E_T$ ($\vec{E}_T$) is defined by $\vec{E}_T = -\sum_i E_i \hat{n}_i$, where $i$ is the calorimeter tower number with $|\eta| < 3.6$, and $\hat{n}_i$ is a unit vector perpendicular to the beam axis pointing at the $i$th calorimeter tower. We also define $E_T = |\vec{E}_T|$.

[12] F. Abe et al. (CDF Collaboration), Phys. Rev. D 45, 1448 (1992).

[13] M. L. Mangano et al., J. High Energy Phys. 07 (2003) 001.

[14] T. Sjostrand et al., High-Energy-Physics Event Generation with PYTHIA 6.1, Comput. Phys. Commun. 135, 238 (2001).

[15] L. Zhang et al., Phys. Rev. D 61, 115007 (2000).