Measurement of the fraction of $\bar{t}t$ production via gluon-gluon fusion in $pp$ collisions at $\sqrt{s} = 1.96$ TeV

MEASUREMENT OF THE FRACTION OF $i \, \ldots$

PHYSICAL REVIEW D 79, 031101(R) (2009)

$^{23}$Harvard University, Cambridge, Massachusetts 02138, USA

$^{24}$Division of High Energy Physics, Department of Physics, University of Helsinki
   and Helsinki Institute of Physics, FI-00014, Helsinki, Finland

$^{25}$University of Illinois, Urbana, Illinois 61801, USA

$^{26}$The Johns Hopkins University, Baltimore, Maryland 21218, USA

$^{27}$Institut für Experimentelle Kernphysik, Universität Karlsruhe, 76128 Karlsruhe, Germany

$^{28}$Center for High Energy Physics: Kyungpook National University, Daegu 702-701, Korea;
   Seoul National University, Seoul 151-742, Korea; Sungkyunkwan University, Suwon 440-746, Korea;
   Korea Institute of Science and Technology Information, Daejeon, 305-806, Korea;
   Chonnam National University, Gwangju, 500-757, Korea

$^{29}$Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

$^{30}$University of Liverpool, Liverpool L69 7ZE, United Kingdom

$^{31}$University College London, London WC1E 6BT, United Kingdom

$^{32}$Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, E-28040 Madrid, Spain

$^{33}$Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

$^{34}$Institute of Particle Physics, McGill University, Montréal, Canada H3A 2T8;
   and University of Toronto, Toronto, Canada M5S 1A7

$^{35}$University of Michigan, Ann Arbor, Michigan 48109, USA

$^{36}$Michigan State University, East Lansing, Michigan 48824, USA

$^{37}$Institut fur Theoretische und Experimental Physik, ITEP, Moscow 117259, Russia

$^{38}$University of New Mexico, Albuquerque, New Mexico 87131, USA

$^{39}$Northwestern University, Evanston, Illinois 60208, USA

$^{40}$The Ohio State University, Columbus, Ohio 43210, USA

$^{41}$Okayama University, Okayama 700-8530, Japan

$^{42}$Osaka City University, Osaka 588, Japan

$^{43}$University of Oxford, Oxford OX1 3RH, United Kingdom

$^{44a}$Istituto Nazionale di Fisica Nucleare, Sezione di Padova-Trento, I-35131 Padova, Italy;
   $^{44b}$University of Padova, I-35131 Padova, Italy

$^{45}$LPNHE, Université Pierre et Marie Curie/IN2P3-CNRS, UMR7585, Paris, F-75252 France

$^{46}$University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

$^{47a}$Istituto Nazionale di Fisica Nucleare Pisa, I-56127 Pisa, Italy;
   $^{47b}$University of Pisa, I-56127 Pisa, Italy;
   $^{47c}$University of Siena, I-56127 Pisa, Italy;
   $^{47d}$Scuola Normale Superiore, I-56127 Pisa, Italy

$^{48}$University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

$^{49}$Purdue University, West Lafayette, Indiana 47907, USA

$^{50}$University of Rochester, Rochester, New York 14627, USA

$^{51}$The Rockefeller University, New York, New York 10021, USA

$^{52a}$Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1, I-00185 Roma, Italy;
   $^{52b}$Sapienza Università di Roma, I-00185 Roma, Italy

$^{53}$Rutgers University, Piscataway, New Jersey 08855, USA

---

*a* Deceased.

bVisitor from Universiteit Antwerpen, B-2610 Antwerp, Belgium.

cVisitor from Chinese Academy of Sciences, Beijing 100864, China.

dVisitor from University of Bristol, Bristol BS8 1TL, United Kingdom.

eVisitor from University of California, Irvine, Irvine, CA 92697, USA.

fVisitor from University of California, Santa Cruz, Santa Cruz, CA 95064, USA.

gVisitor from Cornell University, Ithaca, NY 14853, USA.

hVisitor from University of Cyprus, Nicosia CY-1678, Cyprus.

iVisitor from University College Dublin, Dublin 4, Ireland.

jVisitor from University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom.

kVisitor from Universidad Iberoamericana, Mexico D.F., Mexico.

lVisitor from University of Manchester, Manchester M13 9PL, England.

mVisitor from Nagasaki Institute of Applied Science, Nagasaki, Japan.

nVisitor from Universidad de Oviedo, E-33007 Oviedo, Spain.

oVisitor from Queen Mary, University of London, London, E1 4NS, England.

pVisitor from Texas Tech University, Lubbock, TX 79409, USA.

qVisitor from IFIC (CSIC-Universitat de Valencia), 46071 Valencia, Spain.

rVisitor from Royal Society of Edinburgh.

sVisitor from Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09042 Monserrato (Cagliari), Italy.
We present a measurement of the ratio of the $t\bar{t}$ production cross section via gluon-gluon fusion to the total $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV at the Tevatron. Using a data sample with an integrated luminosity of 955 pb$^{-1}$ recorded by the CDF II detector at Fermilab, we select events based on the $t\bar{t}$ decay to lepton + jets. Using an artificial neural network technique we discriminate between $t\bar{t}$ events produced via $q\bar{q}$ annihilation and $gg$ fusion, and find $G_f = \sigma(gg \rightarrow t\bar{t})/\sigma(p\bar{p} \rightarrow t\bar{t}) < 0.33$ at the 68% confidence level. This result is combined with a previous measurement to obtain the most stringent measurement of this quantity by CDF to date, $G_f = 0.07^{+0.15}_{-0.07}$.


In hadron colliders, such as the Tevatron, the pair production of heavy quarks has contributions from the different partons present in the initial-state hadrons. While for a given quark flavor the total pair production cross section can be measured simply by counting events in specific final-state channels, the contribution from the different primary partons to this cross section is difficult to estimate. For beauty, charm, and light-flavor production, the hadronization process does not normally allow the spin and kinematic properties of a quark to be observed through the analysis of the final-state particles. The situation is different for $t\bar{t}$ production. The top quark, with a mass of about 175 GeV/c$^2$, has a lifetime that is an order of magnitude shorter than the typical hadronization time of $\approx 5 \times 10^{-24}$ s [1]. As a consequence, the spin and kinematic information of the top quark are preserved in its decay products, allowing the different $t\bar{t}$ production processes to be distinguished based on the kinematic characteristics of the final-state particles.

The standard model (SM) predicts the $t\bar{t}$ production processes to be $q\bar{q}$ annihilation ($q\bar{q} \rightarrow t\bar{t}$) and $gg$ fusion ($gg \rightarrow t\bar{t}$), occurring at the Tevatron with relative fractions of $\approx 85\%$ and $\approx 15\%$, respectively, and having significantly different kinematic properties [2]. Predictions for the relative fraction of $t\bar{t}$ production from $gg$ fusion range from $10\%$ to $20\%$ due to uncertainties in the parton density functions [3,4]. A measurement of this fraction tests the SM prediction and our understanding of gluon parton distribution functions (PDFs) in the proton. Disagreement with this prediction could reveal the possible existence of new mechanisms of top-quark production and decay. For instance, production of top pairs at the Tevatron could be affected by a new vector particle associated with top color [5,6]. Such a resonance would affect the angular correlations between the top and antitop, and the relative mixture of $q\bar{q}$ and $gg$ initiated $t\bar{t}$ production. Additionally, new physics in the decay of the top quark, such as a $t \rightarrow H^+ b$, would also affect these correlations [5].

This article describes the first measurement of the fraction $G_f = \sigma(gg \rightarrow t\bar{t})/\sigma(p\bar{p} \rightarrow t\bar{t})$ based on the kinematics of $t\bar{t}$ production, its decay products, and their correlations. We use the $t\bar{t}$ event kinematics in an artificial neural network (NN) to distinguish between the two modes of production. This analysis is described in detail in [7]. We use data in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV collected by the multipurpose Collider Detector at Fermilab (CDF II) [8] from February 2002 to March 2006, corresponding to an integrated luminosity of 955 pb$^{-1}$. The result of this analysis is then combined with a complementary measurement of this fraction [9], which takes advantage of the higher probability for a primary gluon, compared to a quark, to radiate a low energy gluon in the production process.

The production of $t\bar{t}$ is expected to be followed by the decay of each top quark to a $W$ boson and a $b$ quark with a branching ratio of approximately 100% [1]. We select events according to the topology of the $t\bar{t}$ decay to lepton + jets, in which one $W$ decays leptonically and the other one hadronically, $t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow l\nu b q\bar{q}$. We require events to have an electron or muon candidate with $p_T > 20$ GeV/c and $|\eta| < 1$, an imbalance of transverse energy of $E_T > 20$ GeV [10] as expected from the undetected neutrino, and four or more jets with $p_T > 20$ GeV/c and $|\eta| < 2$ [10]. A jet is defined as a cluster of energy in the calorimeter and is reconstructed using an algorithm with a fixed cone of radius 0.4 in $\eta$-$\phi$ space [11]. To account for nonlinearities in the detector response and multiple $p\bar{p}$ collisions in an event, we correct jet energies and $E_T$ [12]. Furthermore, to increase the purity of the sample, at least one jet in the event is required to have a displaced vertex ($b$ tag), which is indicative of the likely $b$-quark production.
origin of the jet [13]. We find 167 candidate events with one b tag, and 65 candidate events with two or more b tags that pass our event selection.

Background processes in the \( t \bar{t} \) candidate sample originate primarily from direct \( W + \) jets production, with minor contributions from diboson production (WW, WZ, ZZ), and multijet production (non-W). The expected background estimates in the one b-tag and two or more b-tags categories are shown in Table I and were determined in [14] with 318 pb\(^{-1}\) of data and scaled to 955 pb\(^{-1}\) of data. We also show the number of events observed in the data and the signal fraction \((\tilde{S}_f)\).

Using the HERWIG version 6.5 [15] leading-order (LO) Monte Carlo (MC) generator, we simulate \( t \bar{t} \) signal samples for the two production processes with a top-quark mass of 175 GeV/c\(^2\). We model the dominant background of \( W + \) light and \( W + \) heavy flavor jets events with the ALPGEN MC generator [16], using HERWIG to model parton showers. All generated events are passed through the CDF II detector simulation [8].

We consider only the four jets with highest transverse energy in each event and include all possible permutations associating jets with partons consistent with the b-tag information. The reconstruction is performed using a kinematic fitter [17] which compares the jet-to-parton association to the \( t \bar{t} \) hypothesis assuming the masses of the \( W \) bosons and the top quarks to be 81 and 175 GeV/c\(^2\), respectively. In the fitter the energy scale of the jets is varied according to its uncertainty. The agreement between each permutation and the \( t \bar{t} \) hypothesis is quantified by the \( \chi^2 \) value of the fit. For each event the permutation with the lowest \( \chi^2 \) is used to extract kinematic variables as described below.

We calculate eight variables that are sensitive to the production mechanism; two of these describe the production and the other six describe the decay of a given \( t \bar{t} \) event. At leading order the \( t \bar{t} \) production rate depends on two variables evaluated in the \( t \bar{t} \) rest frame; the cosine of the angle between the top-quark momentum and the direction of the incoming proton \( \cos \theta^* \) and the top-quark velocity relative to c, \( \beta \) [18]. Since the functional form of the \( t \bar{t} \) production rate to these variables is different for gg fusion than it is for \( q \bar{q} \) annihilation, these variables contain information that could allow us to distinguish the production mechanism. The remaining six variables describe the \( t \bar{t} \) decay and contain information about the correlations between the spins of the top quarks. These variables are the cosines of the angles with respect to the “off-diagonal” spin basis [18,19] in the parent top-quark rest frame. One characteristic feature of this basis is that the number of \( q \bar{q} \to t \bar{t} \) events that have parallel top-quark spins as evaluated in this basis vanishes. Top pair events with parallel top-quark spins come exclusively from gg production. The decay variables are the cosines of the angles between the direction of the off-diagonal basis and the lepton, neutrino, leptonically decaying \( W \), down-type quark, up-type quark, and hadronically decaying \( W \). The distribution of data events for these variables shows very good agreement with the distributions from simulated background and \( t \bar{t} \) events with \( G_f = 0.15 \).

To obtain a single discriminating quantity the eight kinematic variables are fed into a NN. The NN used in this analysis [20] has an architecture of eight inputs, two hidden layers individually with ten and five nodes, and a single output. We train the NN to distinguish between \( q \bar{q} \) and gg simulated \( t \bar{t} \) events, with separate training for the one b-tag events and for two or more b-tags events. Reducing or increasing the numbers of layers and nodes does not significantly change the discriminating power. Approximately one-third of the discriminating power comes from the production variable \( \beta \), one-third from \( \cos \theta^* \), and one-third from the remaining six decay variables. Figure 1 shows the distribution of \( \cos \theta^* \), one of our more sensitive variables, for events with one b-tag in data, expected \( t \bar{t} \), and background.

In each b-tag category we obtain three template distributions of the NN output, \( T^{qg}, T^{gg} \), and \( T^{bbg} \), by running the NN over \( q \bar{q} \) produced \( t \bar{t} \), gg produced \( t \bar{t} \), and background events with.

![Fig. 1](image_url)
ground MC events, respectively. These templates represent the probability for an event to have the NN output obtained assuming it was a $t\bar{t}$ produced by $gg$ fusion, a $t\bar{t}$ produced by $q\bar{q}$ annihilation, or a background event.

An estimator of the $gg$ fraction in the sample $G_f^S$ is obtained by maximizing a likelihood function. We calculate the likelihood of the full event sample for a given $G_f^S$ as

$$L(G_f^S) = L^1(G_f^S)L^2(G_f^S),$$

where $L^1$ and $L^2$ are the one $b$-tag and two or more $b$-tags likelihoods, respectively. The individual likelihoods are defined as

$$L^i(G_f^S) = \exp\left[\frac{- (S_f^i - \bar{S}_f^i)^2}{2\sigma_{\bar{S}_f^i}^2}\right] \prod_{j\neq i} \left\{ (1 - S_f^j) T_i^{bg} + S_f^j [G_f^S T_i^{gg} + (1 - G_f^S) T_i^{tq}] \right\},$$

where $i = \{1, 2\}$, the product is over the events in the $i$th $b$-tag category, the values of the signal fraction $\bar{S}_f^i$ and its uncertainty $\sigma_{\bar{S}_f^i}$ are taken from Table I, and the variables $S_f^i$ represent the observed signal fractions in the sample. The overall multiplicative term represents a Gaussian weight centered at the expected signal fraction. Scanning over values of $G_f^S$ we find the maximum likelihood solution by varying the fractions $S_1^f$ and $S_2^f$. In a given sample the $G_f^S$ value for which the likelihood is maximum is called $G_f^{\text{fit}}$. The fitted fraction $G_f^{\text{fit}}$ is related to the true production fraction $G_f$ by the acceptance ratio of $gg \rightarrow t\bar{t}$ to $q\bar{q} \rightarrow t\bar{t}$. Using HERWIG MC [15] calculations we estimate the acceptance ratio to be $1.29 \pm 0.02$ and $1.25 \pm 0.02$ for the one $b$-tag and two or more $b$-tags categories, respectively. The value of $G_f^{\text{fit}}$ is not constrained to the physically allowable range between zero and unity, and neither would an estimate of $G_f$ obtained from taking into account the acceptance ratio. To ensure a result for $G_f$ in the physical range we use the Feldman-Cousins (FC) prescription [21], which maps any result of $G_f^{\text{fit}}$ to a range of the true fraction $G_f$. We generate this map by fitting for $G_f^{\text{fit}}$ in simulated experiments with a known $G_f$ and $\bar{S}_f$. These simulated experiments are a mixture of $gg$, $q\bar{q}$, and background events keeping the total number in each experiment fixed to that observed in data for that $b$-tag category. We fit the distribution of $G_f^{\text{fit}}$ for the simulated experiments to a Gaussian shape for each value of $G_f$. The FC likelihood-ratio ordering principle [21] is applied to the Gaussian obtained from the simulated experiments to construct the confidence level (C.L.) bands.

To incorporate systematic effects into the FC prescription we generate auxiliary sets of simulated experiments chosen from signal and/or background samples designed to study various sources of systematic uncertainty. The difference of the mean of the $G_f^{\text{fit}}$ distributions generated with the standard and the auxiliary simulated experiments is added in quadrature to the original width of the Gaussian distribution obtained with the standard sample. We repeat the procedure for each value of $G_f$ and for significant sources of systematic uncertainties. The dominant systematic uncertainties result from uncertainties in the background shape and composition, and differences between LO and next-to-leading-order (NLO) predictions estimated by comparing to a $t\bar{t}$ MC sample generated with the NLO generator MC@NLO [22]. We evaluate the systematic uncertainty on this measurement due to parton distribution functions by using MC samples generated with Martin-Roberts-Stirling-Thorne PDFs [23] and the full set of eigenvectors known as CTEQ6M from the CTEQ Collaboration [24]. We also include sources of systematic effects arising from the jet energy scale [12] and initial- and final-state radiation [25].

Finally, we evaluate the log likelihood for the data sample and find the minimum of the negative log likelihood to be $G_f^{\text{fit}} = -0.075$. The variables $S_1^f$ and $S_2^f$ for this value of $G_f^{\text{fit}}$ match those given in Table I within the uncertainties. The $\chi^2$ goodness-of-fit test between the observed data values and the expected values at $G_f^{\text{fit}}$ results in $\chi^2/ndf = 0.9$, indicating a good agreement between the observed and fitted distributions. For this value of $G_f^{\text{fit}}$ the FC construction results in $G_f < 0.33$ at the 68% C.L. and $G_f < 0.61$ at the 95% C.L., respectively.

This measurement is combined with the one performed in [9] also using 955 pb$^{-1}$ of CDF data, in which the $G_f^{\text{fit}}$ fraction is estimated from a fit to the distribution of the number of low transverse momentum charged particles in the event by comparing the data distribution to those from gluon-originated and quark-originated processes. This analysis alone results in $G_f^{\text{fit}} = 0.09 \pm 0.18$, where the uncertainty includes the statistical and systematic uncertainties.

We perform the combination using the FC prescription by including the track multiplicity information in the simulated experiments used for the evaluation of the FC bands. The statistical correlation between the event kinematics analysis and the track multiplicity analysis is found to be negligible. For each $gg$ ($q\bar{q}$) produced $t\bar{t}$ event in a simulated experiment, the value of the track multiplicity for that event is chosen randomly from the gluon-(quark-) originated track distribution. Primary gluons are estimated to contribute to background processes in $54 \pm 9%$ of the cases [9]. Therefore, for background events in a given simulated experiment the value of the track multiplicity is obtained from the gluon-originated distribution $54\%$ of the time and from the quark distribution the remaining times. For each simulated experiment we evaluate the likelihood as a function of $G_f^S$ for the track multiplicity analysis using the goodness-of-fit to data for that fraction. We construct the combined likelihood by multiplying the
Gf = 0.07 ± 0.15 -0.07, and we find the 95% C.L. limit to be Gf < 0.38 [26].

To conclude, we report on the first limit of the fraction of gg produced $\bar{t}t$ events relative to the total by differentiating between the kinematic properties and their correlations of both production processes. Using this technique we limit the fraction $G_f < 0.33$ at 68% C.L. and find it to be consistent with SM expectations. The combination with the measurement described in [9] results in $G_f = 0.07 ± 0.15 +0.07$, yielding the most stringent measurement by CDF of this quantity to date.

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 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.

[10] CDF uses a ($z, \phi, \theta$) coordinate system with the $z$ axis in the direction of the proton beam; $\phi$ and $\theta$ are the azimuthal and polar angle, respectively. The pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$, and the transverse momentum and energy as $p_T = p \sin \theta$ and $E_T = E \sin \theta$, respectively. Missing transverse energy ($E_T = |E_T|$) is defined as $E_T = -\sum E_T \hat{n}_i$, where $\hat{n}_i$ is a unit vector in the transverse plane that points from the beam line to the $i$th calorimeter tower.
[13] A jet is determined to be tagged if it shows a displaced secondary vertex. These jets typically originate from the decay of long-lived mesons such as those resulting after the hadronization process of the $b$ quark.
The result presented in [9], which uses a Bayesian method with a prior that includes the physical boundaries, finds the 95% C.L. to be $G_f < 0.33$. In this article we use the Feldman-Cousins approach finding a combined 95% C.L. of $G_f < 0.38$, where the seemingly lower discriminating power stems only from the difference in the statistical treatment of both analyses. The result presented in [9] evaluated with the Feldman-Cousins approach yields a 95% C.L. of $G_f < 0.42$. 