Study of substructure of high transverse momentum jets produced in proton-antiproton collisions at $s=1.96\text{TeV}$

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

<table>
<thead>
<tr>
<th>Citation</th>
<th>Aaltonen, T. et al. “Study of Substructure of High Transverse Momentum Jets Produced in Proton-antiproton Collisions at $\sqrt{s}=1.96\text{TeV}$.” Physical Review D 85.9 (2012). © 2012 American Physical Society</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1103/PhysRevD.85.091101">http://dx.doi.org/10.1103/PhysRevD.85.091101</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Physical Society</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Apr 01 04:04:30 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/71909">http://hdl.handle.net/1721.1/71909</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Study of substructure of high transverse momentum jets produced in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV
STUDY OF SUBSTRUCTURE OF HIGH TRANSVERSE ...

PHYSICAL REVIEW D 85, 091101(R) (2012)

31 Institute of Particle Physics: McGill University, Montréal, Québec, Canada H3A 2T8; Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6; University of Toronto, Toronto, Ontario, Canada M5S 1A7; and TRIUMF, Vancouver, British Columbia, Canada V6T 2A3
32 University of Michigan, Ann Arbor, Michigan 48109, USA
33 Michigan State University, East Lansing, Michigan 48824, USA
34 Institution for Theoretical and Experimental Physics (ITEP), Moscow 117259, Russia
35 University of New Mexico, Albuquerque, New Mexico 87131, USA
36 Northwestern University, Evanston, Illinois 60208, USA
37 The Ohio State University, Columbus, Ohio 43210, USA
38 Okayama University, Okayama 700-8530, Japan
39 Osaka City University, Osaka 588, Japan
40 University of Oxford, Oxford OX1 3RH, United Kingdom
41 Istituto Nazionale di Fisica Nucleare, Sezione di Padova-Trento, I-35131 Padova, Italy
42 LPNHE, Universite Pierre et Marie Curie/IN2P3-CNRS, UMR7585, Paris, F-75252 France
43 University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
44 Istituto Nazionale di Fisica Nucleare Pisa, I-56127 Pisa, Italy
45 University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA
46 Purdue University, West Lafayette, Indiana 47907, USA
47 University of Rochester, Rochester, New York 14627, USA
48 The Rockefeller University, New York, New York 10065, USA
49 Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1, I-00185 Roma, Italy
4a Sapienza Università di Roma, I-00185 Roma, Italy
4b Texas A&M University, College Station, Texas 77843, USA
42a Istituto Nazionale di Fisica Nucleare Trieste/Udine, I-34100 Trieste, Italy
43a University of Udine, I-33100 Udine, Italy
44 University of Tsukuba, Tsukuba, Ibaraki 305, Japan
45 University of Tsukuba, Tsukuba, Ibaraki 305, Japan

a Deceased.

b Visitor from Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09042 Monserrato (Cagliari), Italy.

c Visitor from University of California Irvine, Irvine, CA 92697, USA.

d Visitor from University of California Santa Barbara, Santa Barbara, CA 93106, USA.

e Visitor from University of California Santa Cruz, Santa Cruz, CA 95064, USA.

f Visitor from CERN, CH-1211 Geneva, Switzerland.

g Visitor from Cornell University, Ithaca, NY 14853, USA.

h Visitor from University of Cyprus, Nicosia CY-1678, Cyprus.

i Visitor from Office of Science, U.S. Department of Energy, Washington, DC 20585, USA.

j Visitor from University College Dublin, Dublin 4, Ireland.

k Visitor from University of Fukui, Fukui City, Fukui Prefecture, Japan 910-0017.

l Visitor from Universidad Iberoamericana, Mexico D.F., Mexico.

m Visitor from Iowa State University, Ames, IA 50011, USA.

n Visitor from University of Iowa, IA City, IA 52242, USA.

o Visitor from Kinki University, Higashi-Osaka City, Japan 577-8502.

p Visitor from Kansas State University, Manhattan, KS 66506, USA.

q Visitor from University of Manchester, Manchester M13 9PL, United Kingdom.

r Visitor from Queen Mary, University of London, London, E1 4NS, United Kingdom.

s Visitor from University of Melbourne, Victoria 3010, Australia.

t Visitor from Muons, Inc., Batavia, IL 60510, USA.

u Visitor from Nagasaki Institute of Applied Science, Nagasaki, Japan.

v Visitor from National Research Nuclear University, Moscow, Russia.

w Visitor from University of Notre Dame, Notre Dame, IN 46556, USA.

x Visitor from Universidad de Oviedo, E-33007 Oviedo, Spain.

y Visitor from Texas Tech University, Lubbock, TX 79609, USA.

z Visitor from Universidad Tecnica Federico Santa Maria, 110v Valparaiso, Chile.

aa Visitor from Weizmann Institute of Science, Rehovot, Israel.

bb Visitor from Yarmouk University, Irbid 211-63, Jordan.

cc On leave from J. Stefan Institute, Ljubljana, Slovenia.
A study of the substructure of jets with transverse momentum greater than 400 GeV/c produced in proton-antiproton collisions at a center-of-mass energy of 1.96 TeV at the Fermilab Tevatron Collider and recorded by the CDF II detector is presented. The distributions of the jet mass, angularity, and planar flow are measured for the first time in a sample with an integrated luminosity of 5.95 fb$^{-1}$. The observed substructure for high mass jets is consistent with predictions from perturbative quantum chromodynamics.

The study of high transverse momentum ($p_T$) massive jets produced in proton-antiproton ($p\bar{p}$) interactions provides an important test of perturbative QCD (pQCD) and gives insight into the parton showering mechanism (see, e.g., [1,2] for recent reviews). Furthermore, massive boosted jets constitute an important background in searches for various new physics models [3–6], the Higgs boson [7], and highly boosted top quark production. Particularly relevant is the case where the decay of a heavy resonance produces high $p_T$ top quarks that decay hadronically. In all these cases, the hadronic decay products can be detected as a single jet with a large mass and internal substructure that differs on average from pQCD jets once the jet $p_T$ is greater than 400–500 GeV/c. However, experimental studies of the substructure of high $p_T$ jets at the Tevatron have been limited to jets with $p_T < 400$ GeV/c [8,9]; recently, results with higher $p_T$ jets produced at the Large Hadron Collider have been published [10].

Jets produced through QCD processes with large mass are expected to arise predominantly through a process of single hard gluon emission from a high $p_T$ top quark or gluon [11]. The probability of this process is given by the jet function, $J(m^{jet}, p_T, R)$, for which a simple next-to-leading order (NLO) approximation is

$$J(m^{jet}, p_T, R) \approx \alpha_s(p_T) \frac{4C_{q,g}}{\pi m^{jet}} \log \left( \frac{R \cdot p_T}{m^{jet}} \right),$$

where the sum is over the constituents in the jet cluster, $E_i$ is the energy, and $\theta_i$ is the angle of each constituent relative to the jet axis. It is sensitive to radiation near the edge of the cone and has a characteristic shape for QCD jets. Planar flow is defined as

$$P_f = \frac{4\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2},$$

where $\lambda_{1,2}$ are the eigenvalues of the two-dimensional moment matrix

$$I_{ij}^{\ell k} = \frac{1}{m^{jet}} \sum_{i \in \text{jet}} \frac{P_{i \ell k}}{E_i} \frac{P_{i k}}{E_i},$$

in which $P_{i \ell k}$ is the $k$th component of the jet constituent’s transverse energy relative to the jet axis, i.e. in one of the two directions that span the plane perpendicular to the jet direction. Jets with three or more energetic constituents, such as those arising from a boosted top quark, are more planar, with $P_f \sim 1$, compared with massive QCD jets where the energy flow is along the line defined by the two final-state partons, with $P_f \sim 0$. Both of these variables are perturbatively calculable.

We report in this paper the first measurement of the jet mass distribution for jets with $p_T > 400$ GeV/c produced in 1.96 TeV $p\bar{p}$ collisions at the Fermilab Tevatron Collider and recorded by the CDF II detector. We also measure for jets with masses greater than 90 GeV/$c^2$ their angularity and planar flow distributions. We use the Midpoint cone algorithm [17] to reconstruct jets using the FASTJET program [18] and the anti-$k_t$ algorithm [19], allowing for a direct comparison of cone and recombination algorithms.

The CDF II detector [20] consists of a solenoidal charged particle spectrometer surrounded by a calorimeter and muon system. Charged particle momenta are measured over $|\eta| < 1.1$. The calorimeter covers the region $|\eta| < 3.6$, with the region $|\eta| < 1.1$ segmented into towers of size $\Delta \eta \times \Delta \phi = 0.11 \times 0.26$ [21]. The calorimeter system is used to measure jets and missing transverse energy ($E_T$) defined as
\[ \vec{E}_T = - \sum_i E_T^i \hat{n}_i, \]  
where the sum is over the calorimeter towers with \( |\eta| < 3.6 \) and \( \hat{n}_i \) is a unit vector perpendicular to the beam axis and pointing at the \( i \)th calorimeter tower. We also define \( \vec{E}_T = \sum E_T^i \). The four-momentum of a jet is the sum over the calorimeter towers in the jet, where each calorimeter tower is treated as a massless four-vector, and the jet mass is obtained from the resulting four-vector.

We select events in a sample with 5.95 fb\(^{-1}\) integrated luminosity identified with an inclusive jet trigger requiring at least one jet with transverse energy \( (E_T) > 100 \) GeV, with the trigger becoming fully efficient for jets with \( E_T > 140 \) GeV. Jet candidates are constructed with a Midpoint cone algorithm with cone radii of \( R = 0.4 \) and 0.7 and with the \( \text{anti-}k_t \) algorithm with a distance parameter \( R = 0.7 \). Primary collision vertices are reconstructed using charged particle information. Events are required to have at least one high quality primary vertex with \( \sqrt{s} < 60 \) cm. Events also need to be well measured by requiring that they satisfy a missing transverse energy significance requirement of \( S_{\text{MET}} < 10 \) GeV\(^{1/2}\), defined as

\[ S_{\text{MET}} = \frac{\vec{E}_T}{\sqrt{\sum E_T^i}}, \]

where the sum is over all calorimeter towers. We calculate for each jet the scalar sum of the \( p_T \) of the tracks associated with the jet cluster. Each jet is required to either have more than 5\% of its energy registered in the electromagnetic calorimeter or to have its summed track momentum be at least 5\%. This criterion eliminates jet candidates arising from instrumental backgrounds. Furthermore, we restrict the jet candidates to have \( 0.1 < |\eta_j| < 0.7 \), where \( \eta_j \) is the jet pseudorapidity in the detector frame of reference, to ensure optimal calorimeter and charged particle tracking coverage. The minimum pseudorapidity requirement avoids a region of the calorimeter where the energy response is varying rapidly. We further require that the leading jet in the event have \( p_T > 400 \) GeV/c. We observe 2699 events.

The jet four-momentum is corrected to take into account calorimeter energy response, which is known to a precision of 3\% [22] for central calorimeter jets with \( p_T > 400 \) GeV/c. We have determined the uncertainty on the calibration of the jet mass measurement by comparing the momentum flux of charged particles into three concentric regions of the calorimeter around the jet centroid with the corresponding calorimeter response.

The number of interaction vertices \( (N_{\text{vtx}}) \) is a measure of the number of multiple interactions (MI), i.e. additional collisions in the same bunch crossing, and averages \(~3\) in this sample. We make a data-driven correction for MI effects on the jet substructure variables [23]. To calculate these corrections, we select a subset of events with a back-to-back dijet topology. We then define cones at right angles to the leading jet in azimuth of the same size as the jet cluster, and add the calorimeter towers in these cones to the jet four-vector after rotation by 90° into the jet cone. The resulting average mass shift upward as a function of \( m_{\text{jet}} \) is taken as the correction downward due to MI and the energy flow from the underlying event of the hard collision.

We separately measure the underlying event correction by using only events with \( N_{\text{vtx}} = 1 \). We correct the leading jet mass \( m_{\text{jet}} \) for events with \( N_{\text{vtx}} > 1 \) by the difference between the mass shift in multivertex events and the mass shift in single vertex events. The correction has an approximate \( 1/m_{\text{jet}}^2 \) behavior and averages \(~4 \) GeV/c\(^2\) for a jet cone size of \( R = 0.7 \). The jet mass correction for a cone size of \( R = 0.4 \) is \(~0.5 \) GeV/c\(^2\), consistent with the expected \( R^4 \) scaling [2]. In the following, we focus on results for \( R = 0.7 \) Midpoint jets.

To model the high \( p_T \) processes, we used a \textsc{pythia} 6.216 calculation [16] of QCD jet production generated with parton \( \vec{p}_T > 300 \) GeV/c, using the Tune A [24] parameters for the underlying event and the CTEQ5L parton distribution functions (PDFs), followed by a full detector simulation. Based on a \textsc{pythia} calculation, we estimate \( W \) and \( Z \) boson production to contribute \(~25\) jets with masses between 60 and 100 GeV/c\(^2\), which is less than 5\% of the number observed. However, top quark pair production can contribute to the jet mass region \( m_{\text{jet}} > 100 \) GeV/c\(^2\), where the expected QCD jet rate is much lower. We employ an approximate next-to-next-to-leading order calculation of the \( \bar{t}t \) differential cross section [25] updated with the MSTW 2008 PDFs [26] and a top quark mass of \( m_{\text{top}} = 173 \) GeV/c\(^2\) [27]. This yields a cross section for top quark jets with \( p_T > 400 \) GeV/c of 4.6 fb. We used the \textsc{pythia} 6.216 generator to create a \( \bar{t}t \) Monte Carlo (MC) sample and applied the same selection requirements used to define the event sample. The estimated \( \bar{t}t \) contribution to the data sample, normalized to the next-to-next-to-leading order cross section, is \(~13 \pm 4\) events.

Two-thirds of the \( \bar{t}t \) events with a leading high \( p_T \) jet would produce a recoil jet with a large jet mass \( (m_{\text{jet}}) \) arising from the fully hadronic decay of the recoil top quark. The remaining \( \bar{t}t \) events would have a recoil top quark that decays sem-leptonically, resulting in large \( \vec{E}_T \) and a recoil jet with lower \( p_T \) and \( m_{\text{jet}} \). We reduce these backgrounds by rejecting events with \( m_{\text{jet}} > 100 \) GeV/c\(^2\) or by making a more stringent \( \vec{E}_T \) requirement by rejecting events with \( S_{\text{MET}} > 4 \) GeV\(^{1/2}\). Approximately 25\% (80\%) of the \( \bar{t}t \) (QCD) MC events survive these requirements. We observe 30 jets with \( m_{\text{jet}} > 140 \) GeV/c\(^2\) and expect a \( \bar{t}t \) contribution of at most three jets.

In order to compare our results with QCD predictions, we correct the \( m_{\text{jet}} \) distributions for effects of selection and resolution by an unfolding procedure, where we correct, bin by bin, the observed \( m_{\text{jet}} \) distribution by the ratio of the QCD \textsc{pythia} MC \( m_{\text{jet}} \) distribution without detector effects and the same distribution after measurement and selection.
effects have been included. This jet mass unfolding correction was derived for each jet algorithm separately, and the correction factors vary from 1.6 to 2.0 over the jet mass range $>70 \text{ GeV}/c^2$. These corrections were verified through studies of the data and confirmed with MC calculations.

We summarize briefly our estimates of the systematic uncertainties that affect the substructure observables. The overall jet mass scale at these energies is known to 2(10) \text{ GeV}/c^2 for jet masses of 60(120) \text{ GeV}/c^2, based on the jet energy scale uncertainty and the comparison of the calorimeter energy and track momentum measurements within the jet mentioned above. We assign an uncertainty on the MI correction of 2 \text{ GeV}/c^2, which is half of the average correction. We assign a $\sim$15% uncertainty on the jet mass unfolding correction due to modeling of the jet hadronization, the uncertainty arising from the selection, and MC statistical uncertainties. The hadronization uncertainty is conservatively determined by comparing the change in the correction when hadronization is turned off in the MC samples. We estimate the PDF uncertainties on the PYTHIA predictions by reweighting the MC events using the $\pm1\sigma$ variations in the 20 eigenvectors describing the uncertainties in the PDFs [28]; the uncertainties on the jet mass, angularity, and planar flow distributions are 10% or less in all cases.

We show in Fig. 1 a comparison of the unfolded $m^{\text{jet}}$ distribution for a cone size $R = 0.7$ with the analytic predictions for the jet function. This comparison, made for jet masses above 70 \text{ GeV}/c^2, shows that the analytical prediction for quark jets describes approximately the shape of the distribution and fraction of jets but tends to overestimate the rate for jet masses from 130 to 200 \text{ GeV}/c^2. The better agreement of the quark jet function with data compared with that of the gluon is consistent with the pQCD prediction that $\sim$80% of these jets arise from quarks [29], though we emphasize that the uncertainties of the pQCD predictions are large. Furthermore, the data and the PYTHIA distributions are in reasonable agreement. We also compare in the inset figure the distributions obtained for the Midpoint and anti-$k_t$ algorithms. The anti-$k_t$ jets have a similar mass distribution to the Midpoint jets. The anti-$k_t$ algorithm, however, does not produce as large a tail of very massive jets, presumably due to the lack of an explicit merging mechanism. This difference in algorithm performance is reproduced by the PYTHIA calculation. We find that 1.4 $\pm$ 0.3% of the Midpoint jets with $p_T > 400 \text{ GeV}/c$ have $m^{\text{jet}} > 140 \text{ GeV}/c^2$. This is the first measurement of this rate, and it allows us to constrain QCD predictions of this fraction and provide the first measurement of the rate of backgrounds in a massive jet sample from QCD production of high $p_T$ light quarks and gluons.

A key prediction of the NLO QCD calculation is that the distribution of angularities [12,13] of high mass jets has relatively sharp kinematical edges, with minimum and maximum values given by

$$\tau^{\text{min}}_2 \sim (2/\zeta)^3, \quad \tau^{\text{max}}_2 \sim \zeta R^2/2^3,$$

with $\zeta \equiv m^{\text{jet}}/p_T$. We show in Fig. 2 the angularity distribution for the leading jet requiring that $m^{\text{jet}} \in (90, 120) \text{ GeV}/c^2$. The requirement of a relatively narrow $m^{\text{jet}}$ window allows us to compare the observed distribution

**FIG. 1** (color online). The normalized jet mass distribution for Midpoint jets with $p_T > 400 \text{ GeV}/c$ and $|\eta| \in (0.1, 0.7)$. The uncertainties shown are statistical (black lines) and systematic (yellow bars). The theory predictions for the jet function for quarks and gluons are shown as solid curves and have an estimated uncertainty of $\sim$30%. We also show the PYTHIA MC prediction (red dashed line). The inset compares Midpoint (full black circles) and anti-$k_t$ (open green squares) jets.

**FIG. 2** (color online). The angularity distribution for Midpoint jets with $p_T > 400 \text{ GeV}/c$ and $|\eta| \in (0.1, 0.7)$. We have applied cuts to reject $t\bar{t}$ events and required that $m^{\text{jet}} \in (90, 120) \text{ GeV}/c^2$. We also show the PYTHIA calculation (red dashed line) and the pQCD kinematic endpoints. The inset compares the distributions for Midpoint (full black circles) and anti-$k_t$ (open green squares) jets.
with the shape and kinematic endpoints predicted by pQCD. The PYTHIA and pQCD predictions are in good agreement with the data for Midpoint and anti-$k_T$ jets, although the small size of the jet sample after applying the mass criterion limits the statistical precision of the comparison. This further strengthens the interpretation that these massive jets arise from two-body configurations. The small number of jets with angularity below $\tau_{-2}$ arise from resolution effects. The PDF uncertainties on the PYTHIA predictions are 10%, and are shown in the figure. The results for jets with cone sizes of $R = 0.4$ are similar.

Figure 3 shows the planar flow distribution for jets where the jet mass is required to be in the range $130-210$ GeV/$c^2$, relevant for jets arising from top quark decays. Comparisons with the PYTHIA predictions are also shown for both QCD multijet and $t\bar{t}$ production. Although the data are in good agreement with the predictions from QCD, the comparison is statistically limited because of the small number of observed jets in this jet mass range. The PDF uncertainties on the PYTHIA QCD predictions are 10%. The results for jets reconstructed with the Midpoint and anti-$k_T$ algorithms are in good agreement with each other and are consistent with the general expectation based on MC calculations [11]. This study suggests that with higher statistics it will be possible to use the planar flow variable to discriminate high $p_T$ QCD and top quark jets independent of jet mass.

In summary, we have measured for the first time the mass, angularity, and planar flow distributions for jets with $p_T > 400$ GeV/$c$ using Midpoint and anti-$k_T$ jet algorithms. We find good agreement between PYTHIA Monte Carlo predictions, the NLO QCD jet function predictions, and the data for the jet mass distribution above 100 GeV/$c^2$ for Midpoint and anti-$k_T$ jets. The Midpoint and anti-$k_T$ algorithms have very similar jet substructure distributions for high mass jets. Our results show that the use of jet mass is an effective variable for separation of jets produced through QCD and through $t\bar{t}$ production, with a jet mass requirement of greater than 140 GeV/$c^2$ leaving only $1.4 \pm 0.3\%$ of the QCD jets. We have also shown that the high mass jets coming from light quark and gluon production are consistent with two-body final states from a study of the angularity variable, and that it may be possible to use the planar flow variable to further reject high mass QCD jets. These results provide the first experimental evidence that validates the MC calculations employing jet substructure to search for exotic heavy particles.

We acknowledge the contributions of I. Sung and G. Sterman for discussions involving nonperturbative effects in QCD jets, and thank N. Kidonakis for updated top quark differential cross section calculations. 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). This work was supported in part by a grant from the Shrum Foundation, and by the Weizmann Institute of Science.


[21] We use a coordinate system where $\phi$ and $\theta$ are the azimuthal and polar angles around the proton beam axis, which defines the $z$ axis. The origin of the coordinate system is the nominal interaction point in the detector. The pseudorapidity is $\eta = -\ln \tan(\theta/2)$ and $R = \sqrt{(\delta \eta)^2 + (\delta \phi)^2}$.


[27] N. Kidonakis (private communication).
