Study of Substructure of High Transverse Momentum Jets Produced in Proton-Antiproton Collisions at $\sqrt{s} = 1.96$ TeV

that differs from pQCD jets once the jet products can be detected as a single jet with substructure. In all these cases, the hadronic decay of a heavy resonance produces high-\( p_T \) jets. Particularly relevant is the case where the decay of a boson [7], and highly boosted top quark pair production, searches for various new physics models [3–6], the Higgs boson [7, 8], and gives insight into the parton showering mechanism and provides an important test of perturbative QCD (pQCD) [8], which has been limited to jets with \( p_T < 400 \text{ GeV}/c \) [8, 9]; recently results with higher \( p_T \) jets produced at the Large Hadron Collider have been published [10].

In the leading-log approximation, a QCD jet acquires a large mass through single hard gluon emission. The probability of this process is given by the jet function, \( J(m_{\text{jet}}, p_T, R) \), for which a simple next-to-leading-order (NLO) approximation is [11]

\[
J(m_{\text{jet}}, p_T, R) \approx \alpha_s(p_T) \frac{4 C_{q,g}}{\pi m_{\text{jet}}} \log \left( \frac{R \cdot p_T}{m_{\text{jet}}} \right),
\]

where \( m_{\text{jet}} \) is the jet mass, \( \alpha_s(p_T) \) is the strong coupling, \( C_{q,g} = 4/3 \) and 3 for quark and gluon jets, respectively, and \( R \) is the cone radius used to define the jet. The approximation holds for \( m_{\text{jet}} \ll R \cdot p_T \). Although uncertain-
ties from higher-order corrections are \( \sim 30\% \), it predicts both the shape of the spectrum and the fraction of jets with masses greater than about 100 GeV/c\(^2\). Two other jet substructure variables insensitive to soft radiation at high jet mass are angularity and planar flow \([12][16]\). The angularity is defined as

\[
\tau = \frac{1}{m^{jet}} \sum_{i\in jet} E_i \sin^{-2} \theta_i \left[ 1 - \cos \theta_i \right]^3, \quad (2)
\]

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. Planar flow is defined as

\[
P_f = \frac{4 \lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2}, \quad (3)
\]

where \( \lambda_1,2 \) are the eigenvalues of the two-dimensional momentum matrix

\[
I^{kl} = \frac{1}{m^{jet}} \sum_{i\in jet} E_i P^{i,k} P^{i,l}, \quad (4)
\]

in which \( P^{i,k} \) 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 perpendicular to the jet direction.

We report in this Letter the first measurement of the jet mass distribution for jets with \( p_T > 400 \) GeV/c produced in 1.96 TeV \( pp \) 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 \( \text{anti}-k_t \) 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

\[
\vec{E}_T = -\sum_i E_i \hat{n}_i, \quad (5)
\]

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 \( E_T = |\vec{E}_T| \). The 4-momentum of a jet is the sum over the calorimeter towers in the jet, where calorimeter tower is treated as a massless 4-vector, and the jet mass is obtained from the resulting 4-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 \(|z_{vtx}| < 60 \text{ cm}\). Events are also required to be well-measured by requiring that they satisfy a missing transverse energy significance requirement of \( S_{MET} < 10 \) GeV\(^{1/2}\), defined as

\[
S_{MET} = \frac{E_T}{\sqrt{\sum E_T}}, \quad (6)
\]

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 \( 0.1 < |\eta_d| < 0.7 \), where \( \eta_d \) is the jet pseudorapidity in the detector frame of reference, to ensure optimal calorimeter and charged particle tracking coverage. We further require that the leading jet in the event have \( p_T > 400 \) GeV/c. We observe 2699 events.

The jet 4-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 jet mass calibration 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 (\( N_{vtx} \)) is a measure of the number of multiple interactions (MI), \( i.e. \) additional collisions in the same bunch crossing, and averages \( \sim 3 \) in this sample. We make a data-driven correction for MI effects on the jet substructure variables \([23]\). 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 4-vector after rotation by 90° into the jet cone. The resulting average mass shift as a function of \( m^{jet} \) is taken as the correction due to MI and the energy flow from the underlying event (UE) of the hard collision. We separately measure the UE correction by using only events with \( N_{vtx} = 1 \). We correct the leading jet mass, \( m^{jet1} \), for events with \( N_{vtx} > 1 \) by the difference between the mass shift in multi-vertex events and the mass shift in single vertex events. The correction has an approximate \( 1/m^{jet1} \) behaviour and averages \( \sim 4 \) 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 $\sim 0.5$ GeV/c$^2$, consistent with the expected $R^2$ 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 PYTHIA 6.216 calculation [10] of QCD jet production generated with parton $p_T > 300$ GeV/c, using the Tune A [24] parameters for the underlying event and the CTEQ5L parton distribution functions (PDFs). Based on a PYTHIA calculation, we estimate $W$ and $Z$ boson production to contribute $\sim 25$ 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 $m_{jet} > 100$ GeV/c$^2$ where the expected QCD jet rate is much lower. We employ an approximate next-to-next-to-leading order (NNLO) calculation of the $t \bar{t}$ differential cross section [25] updated with the MSTW 2008 PDFs [26] and a top quark mass of $m_{top} = 173$ 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 PYTHIA 6.216 generator to create a $t \bar{t}$ MC sample and applied the same selection requirements used to define the event sample. The estimated $t \bar{t}$ contribution to the data sample, normalized to the NNLO cross section, is $13 \pm 4$ events.

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

In order to compare our results with QCD predictions, we correct the $m_{jet}^2$ distributions for effects of selection and resolution by an unfolding procedure, where we correct bin-by-bin the observed $m_{jet}^2$ distribution by the ratio of the QCD PYTHIA MC $m_{jet}^2$ 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 $> 70$ GeV/c$^2$.

We summarize briefly our estimates of the systematic uncertainties. The overall jet mass scale at these energies is known to 2 (10) GeV/c$^2$ for jet masses of 60 (120) 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 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 $\pm 1\sigma$ 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 $m_{jet}^2$ distribution for a cone size $R = 0.7$ with the analytic predictions for the jet function. This comparison, made for jet masses above 70 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 over-estimate the rate for jet masses from 130 to 200 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]. 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 very similar mass distribution to the Midpoint jets. We find that $1.4 \pm 0.3\%$ of the Midpoint jets with $p_T > 400$ GeV/c have $m_{jet}^2 > 140$ GeV/c$^2$.

A key prediction of the NLO QCD calculation is that the distribution of angularities [12] [13] of high mass jets has sharp kinematical edges, with minimum and maxi-
antik}{\textsc{perturbations, and the data for the jet mass distribution above Monte Carlo predictions, the NLO QCD jet function prediction. We find good agreement between \textsc{pythia} and QCD predictions are in good agreement with each other and consistent with the pQCD kinematic endpoints. The inset compares the distributions for Midpoint (full black circles) and antik jets.

\begin{equation}
\tau_{\text{min}} \sim (2/z)^{-3}, \ \tau_{\text{max}} \sim z R^2/2^3,
\end{equation}

with \( z \equiv m_{\text{jet}}/p_T \). 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 with the shape and kinematic endpoints predicted by pQCD. The \textsc{pythia} and QCD predictions are in good agreement with the data for Midpoint and antik jets. The PDF uncertainties on the \textsc{pythia} predictions are 10%. The results for jets with cone sizes of 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 \textendash} 210 \text{ GeV}/c^2, relevant for jets arising from top quark decays. Comparisons with the \textsc{pythia} predictions are also shown for both QCD multi-jet and tt\bar{t} production. The PDF uncertainties on the \textsc{pythia} QCD predictions are 10%. The results for Midpoint and anti-k jets are in good agreement with each other and consistent with the general expectation based on MC calculations \textsuperscript{[11]}. In particular, these results confirm that planar flow provides an effective tool to separate high \( p_T \) jets arising from QCD and tt\bar{t} production.

In summary, we have measured for the first time the mass, angularity and planar flow distributions for jets with \( p_T > 400 \text{ GeV}/c \) using Midpoint and anti-k jet algorithms. We find good agreement between \textsc{pythia} Monte Carlo predictions, the NLO QCD jet function predictions, and the data for the jet mass distribution above 100 \text{ GeV}/c^2 for Midpoint and antik jets. The Midpoint and antik 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 tt\bar{t} production, with a jet mass requirement of greater than 140 \text{ 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 further rejection against high mass QCD jets can be obtained using the planar flow variable. These results provide the first experimental evidence that validate 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 non-perturbative effects in QCD jets, and to 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 Partic-
ules/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 Israel Science Foundation, the Shrum Foundation, and by the Weizmann Institute of Science.

---

* With visitors from "Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09142 Cagliari, Italy, "University of CA Irvine, Irvine, CA 92697, USA, "University of CA Santa Barbara, Santa Barbara, CA 93106, USA, "University of CA Santa Cruz, Santa Cruz, CA 95064, USA, "CERN, CH-1211 Geneva, Switzerland, "Cornell University, Ithaca, NY 14853, USA, "University of Cyprus, Nicosia CY-1678, Cyprus, "Office of Science, U.S. Department of Energy, Washington, DC 20585, USA, "University College Dublin, Dublin 4, Ireland, "University of Fukui, Fukui City, Fukui Prefecture, Japan 910-0017, "Universidad Iberoamericana, Mexico D.F., Mexico, "Iowa State University, Ames, IA 50011, USA, "University of Iowa, Iowa City, IA 52242, USA, "Kinki University, Higashi-Osaka City, Japan 577-8502, "Kansas State University, Manhattan, KS 66506, USA, "University of Manchester, Manchester M13 9PL, United Kingdom, "Queen Mary, University of London, London, E1 4NS, United Kingdom, "University of Melbourne, Victoria 3010, Australia, "Muons, Inc., Batavia, IL 60510, USA, "Nagasaki Institute of Applied Science, Nagasaki, Japan, "National Research Nuclear University, Moscow, Russia, "University of Notre Dame, Notre Dame, IN 46556, USA, "Universidad de Oviedo, E-33007 Oviedo, Spain, "Texas Tech University, Lubbock, TX 79409, USA, "Universidad Tecnica Federico Santa Maria, Valparaiso, Chile, "Weizmann Institute of Science, Rehovot, Israel, "Yarmouk University, Irbid 211-63, Jordan, hh On leave from J. Stefan Institute, Ljubljana, Slovenia,

[21] We use a coordinate system where $\phi$ and $\theta$ are the azimuthal and polar angles around the proton beam axis. The pseudorapidity is $\eta = -\ln \tan(\theta/2)$ and $R = \sqrt{(\delta \eta)^2 + (\delta \phi)^2}$.