CDF/PUB/JET/ PUBLIC/10199
June 19, 2011
Version 2.3

The Substructure of High Transverse Momentum Jets Observed by CDF II

CDF II Collaboration

Abstract

We present the results of a study of jets with transverse momentum (p_T) greater than 400 GeV/c with particular emphasis given to the mass of the jets and two other measures of substructure, angularity and planar flow. The jets are produced in 1.96 TeV proton-antiproton collisions at the Fermilab Tevatron Collider and recorded by the CDF II detector. The jets in each event are reconstructed using the Midpoint algorithm with cone sizes of $R = 0.4$, $R = 0.7$ and $R = 1.0$. We select events with at least one jet with transverse momentum > 400 GeV/c in a sample of 5.95 fb^{-1} and correct them for the effects of multiple interactions. We find good agreement of the observed jet mass distribution with both the QCD theoretical predictions and PYTHIA Monte Carlo calculations employing full detector simulation. We have also studied for the first time the distributions of two other jet substructure variables, angularity and planar flow. We find that the distributions of these variables for massive jets are in good agreement with the qualitative QCD theoretical expectation and PYTHIA Monte Carlo calculations. We have also investigated the jet algorithm dependence of these variables, comparing the results using the Midpoint, anti-k_T and Midpoint (with search cones) algorithms.

1 Introduction

1.1 The Importance of High p_T Massive QCD Jets

High p_T massive QCD jets are interesting to study from various aspects (see *e.g.* [1, 2, 3] for recent reviews). From the experimental perspective, the substructure of these objects have not been extensively studied at the Tevatron and the studies that have been performed [4, 5] have been limited in both the p_T and mass of the jets. Theoretically, the observation of massive collimated jets provides an important test of perturbative QCD, an opportunity to tune the various Monte Carlo (MC) event generators and gives insight into the parton showering mechanism. Furthermore, massive boosted jets comprise perhaps the most important background for various new physics models [6, 7, 8, 9, 10, 11, 12] and even Higgs searches [13, 14, 15]. Particularly relevant is the case where microscopic dynamics, for instance via decay of a heavy resonance, produces high-p_T, standard model (SM) massive particles that decay hadronically. Consequently, these new physics signals are expected to be of the form of highly boosted W, Z, H and top jets. Finally, the results as well as the techniques presented may be useful in follow-up studies at the LHC experiments.

We also note that there have been relatively few studies at the Tevatron that explore the production of very highly boosted top quarks. The first top quark p_T measurement was by CDF [16] using 0.1 fb^{-1} of Run I data. The DZero collaboration has recently published a new measurement using 1 fb^{-1} of Run II data [17].

In both cases, the statistics of these measurements have been limited by the integrated luminosity and the branching fraction and identification efficiency for $t\bar{t}$ lepton+jet final states, and no data is available for top quarks with $p_T > 400$ GeV/c. Theoretical predictions for the differential top quark production cross section as a function of p_T now exist up to next-to-next-to-leading order (NNLO) [18] so that a measurement or upper limit on the top quark production cross section for $p_T > 400$ GeV/c will add in a unique way to our understanding of top quark production.

1.2 Theoretical Background

Based on QCD factorization (see *e.g.* [19]), an analytic calculation of the QCD jet mass distribution has been derived where the mass, m_J, is dominantly due to a single gluon emission [20]. The jet function can be defined via the total differential rate

$$\frac{d\sigma(R)}{dp_T dm_J} = \sum_{q,G} J^{q,G}(m_J, p_T, R) \frac{d\hat{\sigma}^{q,G}(R)}{dp_T}, \tag{1}$$

where $\hat{\sigma}^{q,G}$ is the factorized Born cross section and R is the half-cone size used to associate final state particles with the jet. Corrections of $O\left(R^2\right)$ are neglected and the analysis is applied to the high mass tail, $m_J^{\text{peak}} \ll m_J \ll p_T R$ (m_J^{peak} corresponds to the peak of the jet mass distribution). A simple approximation for the full result [20] is

$$J(m_J, p_T, R) \simeq \alpha_s(p_T) \frac{4C_{q,G}}{\pi m_J} \log\left(\frac{R\,p_T}{m_J}\right), \tag{2}$$

where $\alpha_s(p_T)$ is the strong coupling constant at the appropriate scale and $C_{q,G} = 4/3, 3$ for quark and gluon jets, respectively.

We note that the above result contains two non-trivial pieces of information. The first is that the overall scale of the jet mass distributions (i.e., the probability of having a jet at a given mass) is a physical observable and has no arbitrary or unknown normalization. The second is that the shape of the distribution is well defined. Both predictions can be tested by the data. The former can be translated into a rejection power as a function of a mass cut, or a mass window, both useful in distinguishing between new physics signal and QCD background [21, 22] and the latter can improve sensitivity to new physics searches by reducing the uncertainty in a background subtraction using a sideband analysis.

The above result is the leading log approximation to the full expression where the next to leading order is not fully known at present [2, 24, 25]. The corrections to the above expression is expected to be very roughly of order of $1/\log\left(R^2\,p_T^2/m_J^2\right) \sim 30\%$ for the relevant parameter space. Thus, while the above theoretical expression is not very precise they still provide a simple and powerful description for the qualitative behavior of the data, which is presented below.

Various substructure variables have been suggested to provide further discrimination of massive jets arising from QCD production and other sources such as top quark production. The effort in the literature (see [2] for a recent review) can be categorized into three broad classes: techniques geared towards specifically two-pronged kinematics [6, 13, 14, 23], techniques employing three-pronged kinematics [12, 20, 23, 26, 27, 28, 29, 30] (*e.g.* $h \to b\bar{b}$ for two-body and $t \to bq\bar{q}$ for three-body kinematics) and methods that are structured towards removing soft particle contamination [31].

In this study, we focus on measuring angularity and planar flow jet shape variables, which are representative of two of the three clases. At small cone sizes and large jet mass, these variables are expected to be quite robust against soft radiation and allow in principle a comparison with theoretical predictions in

addition to comparison with MC results. Both variables are also less dependent on the particular jet finding algorithm used.

Angularity corresponds to a class of jet shapes [32, 23]

$$\tau_a(R,p_T) = \frac{1}{m_J}\sum_{i\in jet}\omega_i \sin^a\theta_i \left[1-\cos\theta_i\right]^{1-a} \sim \frac{2^{a-1}}{m_J}\sum_{i\in jet}\omega_i\theta_i^{2-a}, \tag{3}$$

where ω_i is the energy of a component inside the jet (such as a calorimeter tower). Limiting the parameter $a \leq 2$ ensures IR safety, as can be directly seen from the expression on the right hand side of the equation which is valid for small angle radiation $\theta_i \ll 1$. [1]

The angularity distribution, $d\sigma/d\tau_a$, is similar over a large class of jet definitions (for instance the k_T and anti-k_T variety [33]) in the limit of $R \ll 1$ and high jet mass [23]. It is particularly sensitive to the degree of symmetry in the energy deposition. It therefore can distinguish jets originating from QCD production of light quarks and gluons from boosted heavy particle decay. The key point here is that for high mass jets, the distribution of the leading parton and the emitted gluon is expected to be peaked around a symmetric p_T configuration where both partons are at the same distance, θ_i, from the jet axis, $\theta_{1,2} = z \equiv m/p_T$ [23]. This implies that there is a minimum value for angularity

$$\tau_a^{\min}(z) \sim \left(\frac{z}{2}\right)^{1-a}. \tag{4}$$

This is true both for the signal and background events and hence Eq. (4) provides an important test for the energy distribution of massive jets. In particular, in our case this is a test for the two pronged description of the QCD energy distribution of boosted massive jets and the corresponding showering mechanism.

We now move to discuss the leading differential angularity distribution for the massive QCD jets. The jet leading-order (LO) energy distribution was derived in [23] and can be characterized by the angular distribution of the softer particle in the two-pronged description. It can be further simplified in the limit of

$$z, \left|\theta_s\right|, R \ll 1 \tag{5}$$

to give

$$\frac{d^3\sigma}{dp_T\, dm\, d\theta_s} \simeq \frac{dJ^{\rm QCD}}{d\theta_s}(m,p_T,R) \sim \frac{4\alpha_s C_F}{\pi m\theta_s}, \tag{6}$$

where θ_s is the angle between the softer particle and the jet axis. Through suitable approximations, one finds that the jet angularity distribution is given by [34]

$$\frac{dJ^{\rm QCD}}{d\tau_a}(m,p_T,R) \sim \frac{4\alpha_s C_F}{\pi am\tau}. \tag{7}$$

We can also get an estimate for the maximal value of τ_a that arises due to radiation near the edge of the cone:

$$\tau_a^{\max}(R,p_T) \sim 2^{a-1}R^{-a}z, . \tag{8}$$

We can evaluate the relevant range for angularity to check for the validity of our approximation by dividing $\tau_a^{\max}$ by $\tau_a^{\min}$

$$R_\tau \equiv \frac{\tau_a^{\max}}{\tau_a^{\min}} \sim \left(\frac{z}{R}\right)^a = \left(\frac{m}{p_T R}\right)^a. \tag{9}$$

[1] Note that in the original definition of angularity within a jet [23] the argument of the sin and cos functions was defined as $\frac{\pi\theta_i}{2R}$. However, for a generic jet algorithm configuration, $\theta_i \sim 2R$ are sometimes obtained and this results in singular behavior for angularity. Hence, we present a slightly improved expression where these singularities are avoided in the narrow cone case [35].

For the typical parameters measured below and for cones of $R = 0.4$ and 0.7, $z \sim 0.25$ and we therefore find $R_\tau \sim 2.6$ and 7.8 for the two cone sizes.

Planar flow (Pf) is another IR-safe jet shape that can be used to distinguish planar from linear configurations [20, 23, 27]. To define planar flow, we first construct for a given jet a matrix I_w

$$I_w^{kl} = \frac{1}{m_J} \sum_i w_i \frac{p_{i,k}}{w_i} \frac{p_{i,l}}{w_i}, \tag{10}$$

where m_J is the jet mass, w_i is the energy of particle i in the jet, and $p_{i,k}$ is the k^{th} component of its transverse momentum relative to the jet momentum axis. Given I_w, we define Pf for that jet as

$$Pf = 4 \frac{\det(\mathrm{I_w})}{\mathrm{tr}(\mathrm{I_w})^2} = \frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}, \tag{11}$$

where $\lambda_{1,2}$ are the eigenvalues of I_w. Pf vanishes for linear shapes and approaches unity for isotropic depositions of energy.

2 Event Selection

We select events identified with an inclusive jet trigger requiring at least one jet with > 100 GeV of transverse energy (E_T). The integrated luminosity of the data sample was 5.95 fb^{-1} and consisted of 75,764,270 events, corresponding to an effective triggered cross section of 12.7 nb. Jets in these events were subsequently reconstructed with a Midpoint cone algorithm [37] with cone sizes of $R = 0.4$, 0.7 and 1.0 using the Fastjet software package [38]. We compare the results obtained with the Midpoint algorithm with the results obtained using the anti-k_T algorithm and the Midpoint algorithm with search cones.

Events were required to have a high quality primary vertex with $|z_{vtx}| < 60$ cm, and to satisfy a relatively loose missing transverse energy ($\not\!\!E_T$) requirement of $S_{MET} < 10$ where the missing transverse energy significance variable, S_{MET}, is defined as

$$S_{MET} \equiv \frac{\not\!\!E_T}{\sqrt{\sum E_T}}, \tag{12}$$

where the sum is over the transverse energy observed in all calorimeter towers. Events were also required to satisfy quality cuts to reject cosmic ray backgrounds, and the individual jet candidates were required to satisfy requirements that ensured that there was confirmation of the jet in the tracking chamber by requiring that the summed transverse momentum of the charged tracks associated with the jet be at least 0.05 of the calorimeter E_T. Furthermore, we restricted the jet candidates to have their centroid in the interval $0.1 < |\eta| < 0.7$. The lower value was chosen to avoid a region of significant inhomogeneity in the calorimeter and the upper region was chosen to ensure that the the jets were contained in the well-understand central calorimeter of the CDF II detector.

The four-momentum of each jet was corrected to take into account calorimetry energy response, inhomogeneities in the detector as a function of pseudorapidity, and the non-linearity of the calorimeter response for lower energy deposition. The average correction factor was 1.12 for jets with $p_T > 400$ GeV/c and the uncertainty in the overall correction was $\pm 3\%$ for these momenta [39]. The mass of the jet was calculated using the "E-scheme" method, where each calorimeter tower in a jet was considered a massless 4-vector, and the jet's 4-momentum was the sum of the tower 4-vectors. The resulting mass was corrected by the jet energy correction mentioned above.

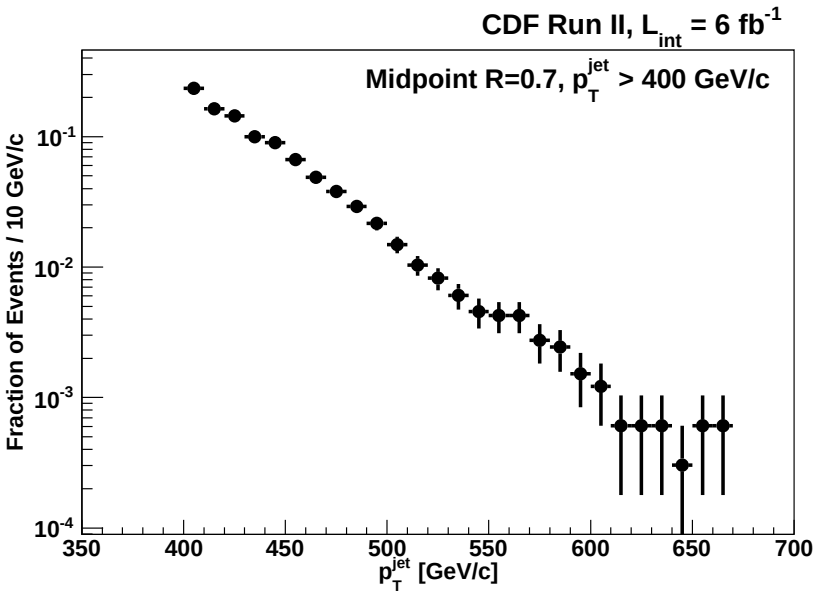

Figure 1: The p_T distribution for the all jet candidates from events that satisfy the event selection described in the text, as well as the pseudorapidity requirement $|\eta| \in (0.1, 0.7)$.

We required events in the sample to have at least one jet with $p_T > 400$ GeV/c. This selection resulted in 2699 events with jets with a cone size of $R = 0.7$.

The p_T distribution of the jets in this sample for $R = 0.7$ is shown in Fig. 1.

3 Monte Carlo Calculations

With the selection described in Section 2, we anticipate that the event sample is dominated by jets produced by QCD scattering. We use a PYTHIA 6.216 calculation of QCD jet production generated with $\hat{p}_T > 300$ GeV/c and the CTEQ5L parton distribution functions (PDFs). We generated 4.893×10^6 events, corresponding to an integrated luminosity of 795 fb^{-1}, giving us a high-statistics sample to compare with data.

W and Z boson production is expected to contribute to this sample. Based on a PYTHIA 6.4.1 calculation, we estimate these processes to contribute 2-4 fb of observed jets, which is negligible given that the total QCD cross section of jets is predicted to be ~ 500 fb.

The only other significant source of events to this sample is top quark pair production. The most recent NNLO calculation of the $t\bar{t}$ differential cross section [18] has been updated with the MSTW 2008 parton distribution functions and a top quark mass of $m_{top} = 173$ GeV/c^2 [40]. The results of this calculation are shown in Fig. 2. The calculation itself includes next-to-leading-order (NLO) corrections to the leading-order diagrams along with next-to-next-to-leading-order (NNLO) soft-gluon corrections. No rapidity cut was placed on this cross section though the author believes this would have a neglible effect on the overall rate. The scale used is $\mu^2 = p_T^2 + M_{top}^2$.

This calculation yields a total cross section of 8.15 pb and a cross section for $p_T > 400$ GeV/c of $4.55^{+0.50}_{-0.41}$ fb. Said another way, the fraction of top quarks produced with $p_T > 400$ GeV/c is 5.58×10^{-4} of the total top quark cross section.

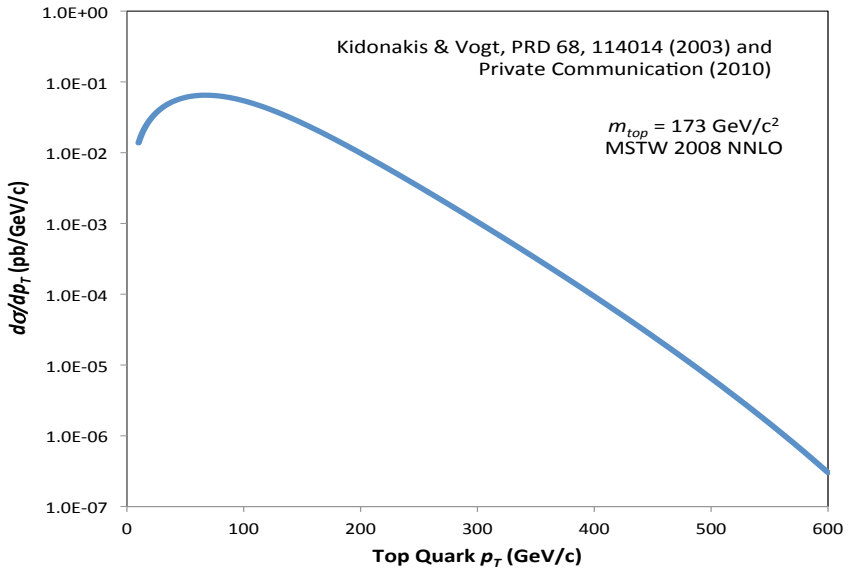

Figure 2: The top quark differential cross section as calculated by Kidonakis and Vogt [18, 40] as a function of the top quark p_T.

We can also calculate this cross section by using the PYTHIA 6.216 calculation to predict the fraction of high p_T top quarks produced and then scale this to the measured cross section. We observe 4041 events in a 4.75M inclusive $t\bar{t}$ MC sample with at least one top quark with $p_T > 400$ GeV/c. This yields a fractional rate of 8.51×10^{-4}. The measured total $t\bar{t}$ cross section is 7.50 ± 0.48 pb [41], which results in a cross section for top quarks with $p_T > 400$ GeV/c of 6.38 fb, or about 40% higher than the Kidonakis and Vogt prediction. In the calculations of $t\bar{t}$ rates below, we use the estimate provided by Kidonakis and Vogt.

4 Jet Corrections and Non-QCD Backgrounds

4.1 Calibration of Jets

The standard jet energy corrections are known to relatively high precision [39] for a variety of cone sizes. For jets with $p_T > 400$ GeV/c and measured in the central calorimeter, the systematic uncertainty in the overall jet energy scale is $\pm 3\%$ and is dominated by our understanding of the response of the calorimeter to individual particle energies. Other sources of uncertainty such as out-of-cone effects, underlying event energy flow and multiple interactions are modest at these high energies.

We verified that the energy calibration was also consistent internally in the jet by comparing the momentum flux into three concentric regions of the calorimeter around the jet centroid with the corresponding calorimeter response. This "ring" study allowed us to constrain the extent to which the calorimeter energy response modeling is uncertain independently of the studies done to set the uncertainties on the overall jet energy scale. Our studies showed that the mass of jets around 60 GeV/c^2 were uncertain to approximately 2 GeV/c^2. Jets with masses around 120 GeV/c^2 had a much larger sensitivity to this effect, with a jet mass scale uncertainty of 9.6 GeV/c^2 at jet masses of 120 GeV/c^2.

Multiple interactions, i.e. collisions that occured in the same bunch crossing that created the high p_T

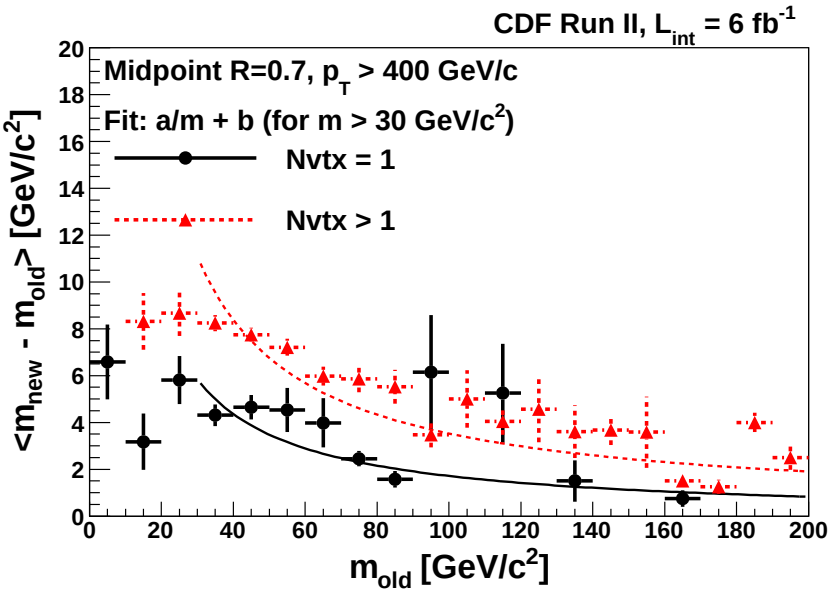

Figure 3: The correction to the jet mass from additional energy deposition due to multiple interactions and underlying event ($N_{vtx} > 1$ events) compared with the jet mass corrections for underlying event ($N_{vtx} = 1$ events) for jets with a cone size of $R = 0.7$. The average number of collisions per bunch crossing is ~ 3 for this data sample.

jets, occur at a significant rate given the average instantaneous luminosity during which these data were collected. By counting the number of reconstructed primary vertices, we estimate the average number of $p\bar{p}$ interactions to be 3 per bunch crossing (including the primary interaction that created the massive jet system). We developed a new technique to measure the average mass shift arising from this effect in a data-driven manner [42]. We selected a subset of the high p_T events that had a clear dijet topology, and defined cones in the calorimeter at right angles to the leading jet in azimuth of the same size as the jet cones. The energy observed in these "complementary" cones were then rotated under the jet and the resulting towers added to the jet 4-vector. The resulting mass shift was taken as the estimate for the shift that the jet itself experienced due to the multiple interactions.

In considering the theoretical implications of this correction, we believe that it is inappropriate to correct the jet mass for both the multiple interactions as well as the underlying event (UE) in each collision because some parts of the UE are coherent with the jet final state and other parts are not. We determine the average UE contribution by performing this calculation on the events with a single primary vertex, and substract that from the total MI/UE corrections. The average correction as a function of jet mass is shown in Fig. 3 for events with multiple interactions ($N_{vtx} > 1$) and events with a single collision in the bunch crossing ($N_{vtx} = 1$).

The correction has a $1/m^{jet}$ behaviour and averages approximately 4 GeV/c^2 for a jet cone size of $R = 0.7$. Below an observed jet mass of 30 GeV/c^2, we parametrize the correction to linearly fall to zero as the jet mass approaches zero because the actual correction upward for such jets must be negligible. The jet mass correction for a cone size of $R = 0.4$ is approximately 0.5 GeV/c^2. This is consistent with an R^4 scaling when compared with the corrections for $R = 0.7$ cones, expected from geometry and the integration of this contribution across the face of a jet. We use the $R = 0.7$ correction scaled by a factor of $(0.4/0.7)^4$ and

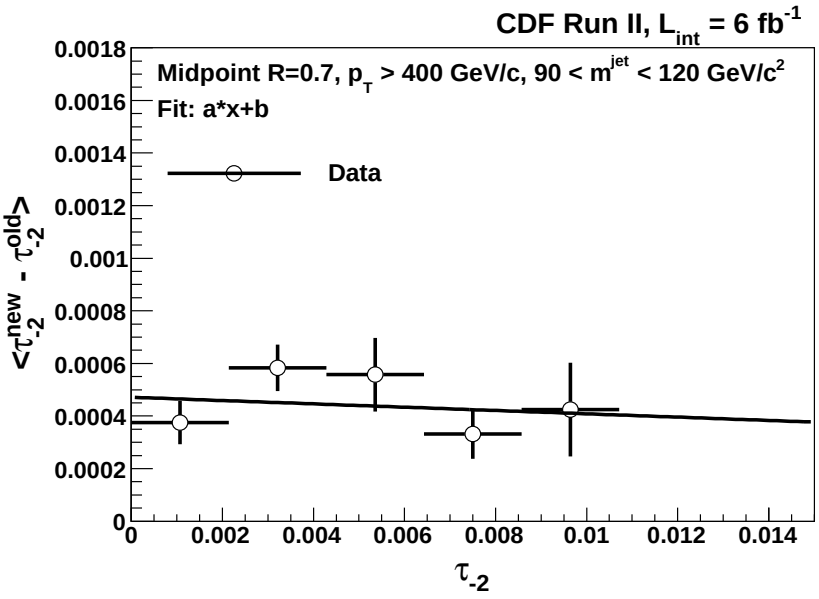

Figure 4: Estimated shift to angularity from multiple interactions and underlying event to angularity in data using Midpoint jets with $R = 0.7$.

$(1.0/0.7)^4$ for jet cones with $R = 0.4$ and $R = 1.0$, respectively, given that we have limited data for larger jet masses with the smaller cone and it is not possible to measure this correction directly for the largest cones.

We employ a similar approach to determining MI and UE corrections to angularity and planar flow, though the number of events do not allow us to separate out contributions from the two sources. The resulting shifts in angularity and planar flow are shown in Figs. 4 and 5. We find that these shifts are on average small compared to the range of each variable, as shown below. Given their size, we do not perform an MI correction on the observed angularity and planar flow distributions.

4.2 Top Quarks as a Background

Top quark production is dominantly a pair-production process ($t\bar{t}$) with the transverse momentum of the top quark being approximately half the mass of the quark, but with a long tail to higher transverse momentum (see Fig. 2). It is this tail that in principle contributes to any analysis looking at very boosted objects.

We apply to the top quark MC sample the same jet selection requirements used to define the event sample, i.e. requiring events to have at least one jet with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$. We estimate the observed $t\bar{t}$ event rate in our sample to be approximately 1.6 fb. We show in Fig. 6 the jet mass distribution, m^{jet1}, for the leading jet in events with at least one jet with $p_T > 400$ GeV/c using Midpoint jets with $R = 1.0$. We have also required that $S_{MET} < 4$ to eliminate $t\bar{t}$ candidates where one top quark has decayed semileptonically. A clear broad peak around a mass of 160 to 190 GeV/c^2 is visible, along with a smaller shoulder around 80 GeV/c^2. We also show the leading jet mass distribution from our QCD MC sample.

We also investigated the properties of the jet in the event that has the second highest p_T (we call this the second leading jet). We show in Figs 7, 8 and 9 the p_T distributions, η distributions and m^{jet2} distributions, respectively, for this second jet for $t\bar{t}$ MC events and for QCD MC events. The jet mass distributions were calculated with a cone size of $R = 1.0$.

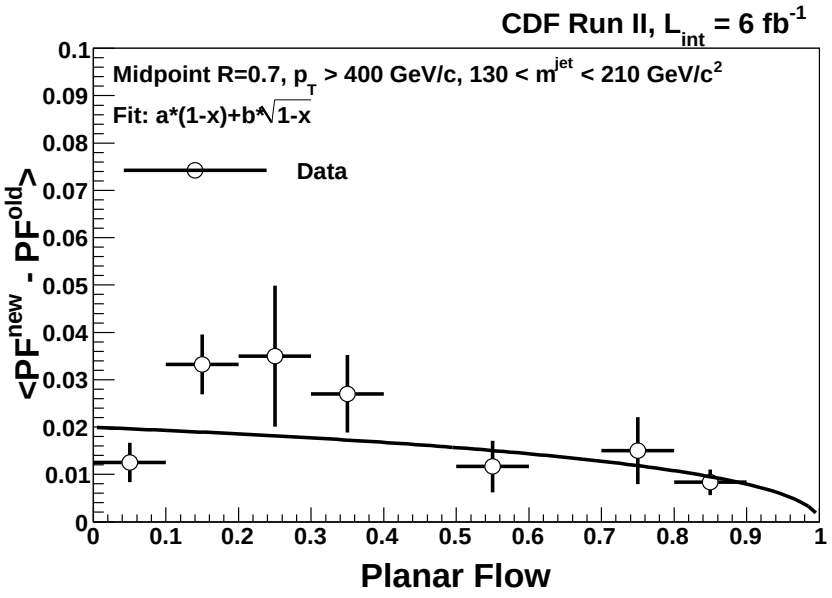

Figure 5: Estimated shift to planar flow from multiple interactions and underlying event in data events using Midpoint jets with $R = 0.7$.

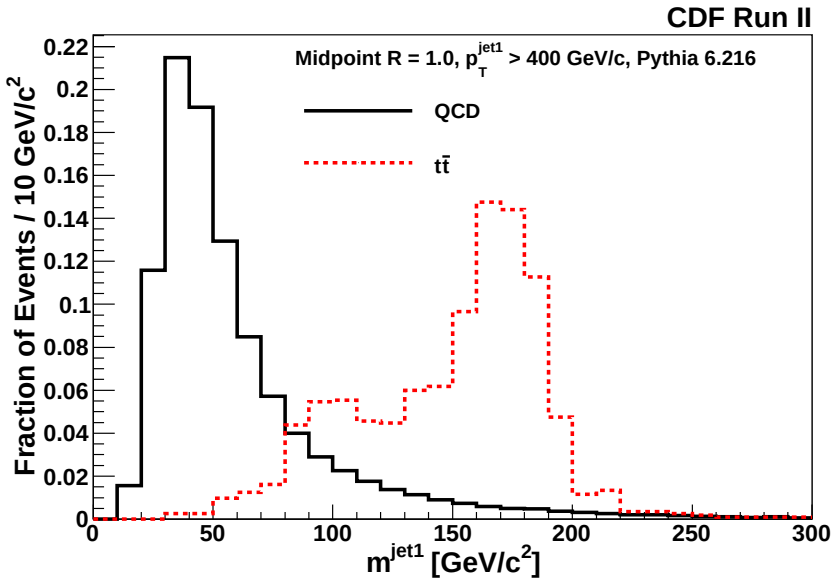

Figure 6: The jet mass distribution of the leading jet ($R = 1.0$) in $t\bar{t}$ MC events (red dashed line) and for QCD MC events (solid black line), where we have required each jet to satisfy $p_T > 400$ GeV/c and $|\eta| < 0.7$. We have also required that $S_{MET} < 4$ to reject events where one of the top quarks has decayed semileptonically.

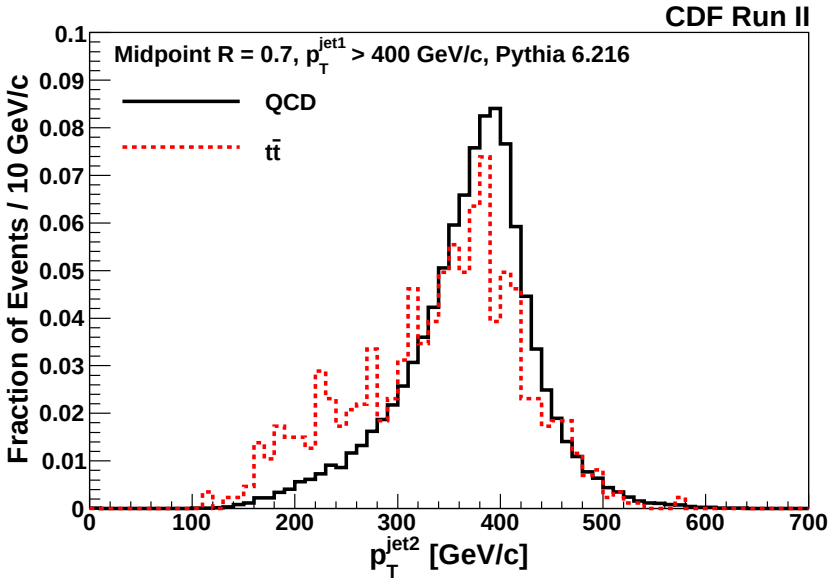

Figure 7: The jet p_T distribution of the second-leading jet ($R = 0.7$) in $t\bar{t}$ (red dashed line) and QCD (solid black line) MC events, requiring that the leading jet satisfy $p_T > 400$ GeV/c and $|\eta| < 0.7$ with a cone size of $R = 0.7$.

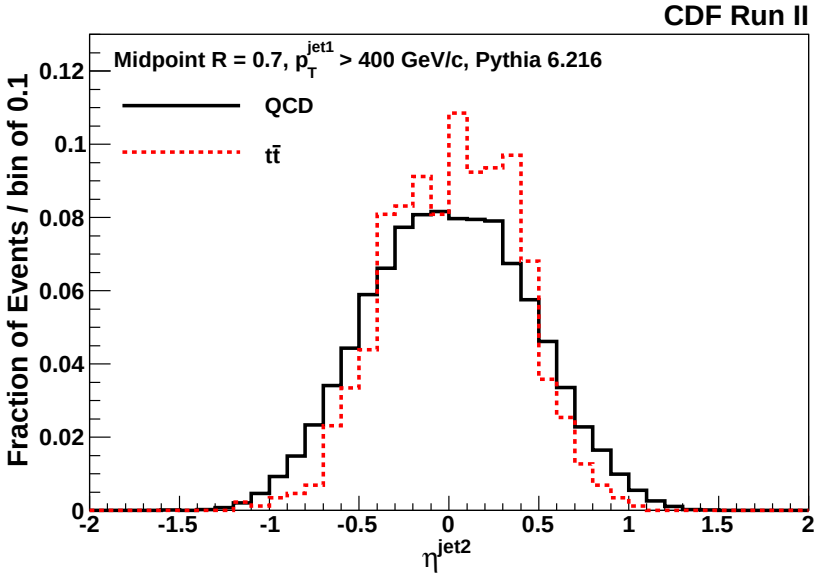

Figure 8: The jet η distribution of the second-leading jet ($R = 0.7$) in $t\bar{t}$ (red dashed line) and QCD (solid black line) MC events, requiring that the leading jet satisfy $p_T > 400$ GeV/c and $|\eta| < 0.7$.

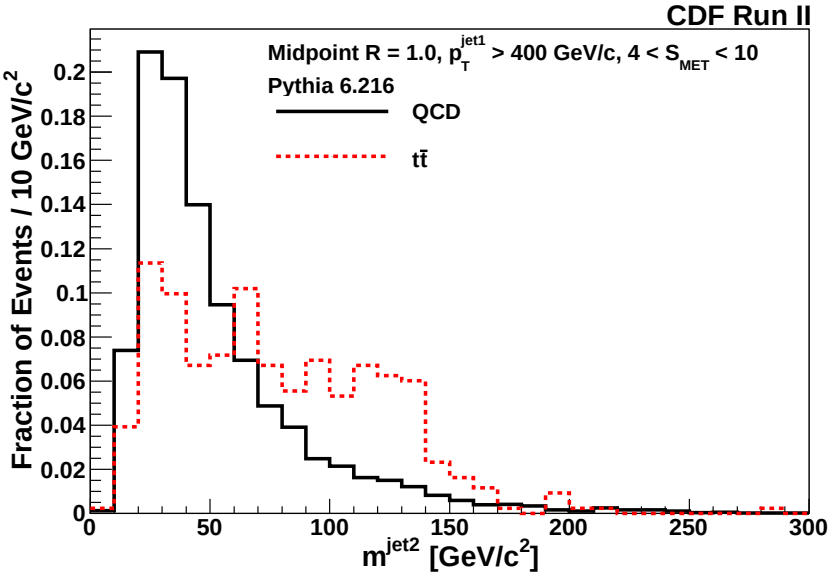

Figure 9: The jet mass distributions of the second-leading jet ($R = 1.0$) in $t\bar{t}$ (dotted lines) and QCD (solid black line) MC events, requiring that the leading jet satisfy $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$.

We expect that a fraction – approximately one-third – of $t\bar{t}$ events where we have observed one hadronically decaying top quark as the leading jet would have a recoil top quark that decays semileptonically, resulting in missing transverse energy and a second-leading jet with lower p_T and jet mass. We show in Fig. 10 the distribution of the missing E_T significance, S_{MET}, after requiring a leading jet meeting our standard requirements of $p_T > 400$ GeV/c and $|\eta| < 0.7$.

4.3 Rejection of Top Quark Events

The primary goal of this study was to compare the jet substructure associated with highly boosted QCD jets. Thus, any top quark contributions are a background that could distort these substructure distributions. We therefore used the the correlations predicted by the MC to reject $t\bar{t}$ backgrounds.

We applied the following relatively simple cuts to reject $t\bar{t}$ events:

- Reject any event with a second-leading jet with $m^{jet2} > 100$ GeV/c^2.
- Reject any event with $S_{MET} > 4$.

We also require that the second-leading jet have a $p_T > 100$ GeV/c so that we have a well-defined second-leading jet.

With these requirements, only 35% of the remaining $t\bar{t}$ MC events survive, while 82% of the QCD MC events satisfy this requirement. This rejection strategy reduces any remaining contamination of $t\bar{t}$ events to ~ 0.3 fb.

4.4 W and Z Boson Contamination

Based on the PYTHIA 6.4.1 calculation, we expect a total production rate of W and Z bosons with $p_T >$ 400 GeV/c of 4.5 fb and 3.0 fb, respectively. Factoring in the hadronic branching fractions of 0.68 and 0.70

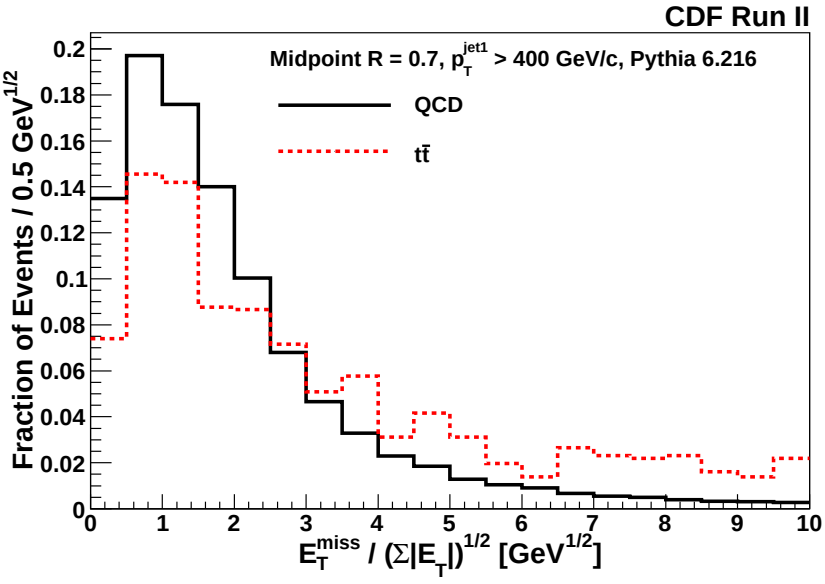

Figure 10: The missing transverse energy significance distribution for $t\bar{t}$ and QCD MC events requiring that the leading jet satisfy $p_T > 400$ GeV/c and $|\eta| < 0.7$.

and reconstruction efficiencies, we estimate that these result in a cross section of observed jets of ~ 3.6 fb. In our data sample, this would yield approximately 20 events with a jet in the 70-110 GeV/c^2 mass region where we see 197 events.

5 Systematic Uncertainties

The measurements that we are making in this analysis are fundamentally comparing distributions of a number of jet substructure variables with Monte Carlo predictions and analytical calculations. As such, much of the analysis revolves around how well we understand the measurement of jet masses. We detail here our methodology for estimating these systematic uncertainties.

5.1 Calorimeter Energy Scales

The ring study reported earlier and the overall jet energy scale uncertainty of 3% for these high-energy jets allows us to constrain the size of the possible shifts in jet mass scale to be 1 GeV/c^2 for jet masses around 60 GeV/c^2 and 10 GeV/c^2 for jet masses around 120 GeV/c^2.

We therefore set a systematic uncertainty of 10 GeV/c^2 on the jet mass arising from a possible energy-dependent miscalibration of the calorimeter for jets with masses above 70 GeV/c^2 where we compare the distribution of jets with theoretical predictions.

5.2 Jet Mass Unfolding Factor

We found that the p_T resolution is correlated with the jet mass, with higher mass jets have a narrower p_T resolution, based on MC studies and studies of the recoil jets in these events. This means that a correction

has to be made to the differential jet cross section vs jet mass, discussed below. The uncertainties in that unfolding factor come from the hadronization model used in the MC calculation ($\pm$10%), the p_T energy scale uncertainty ($\pm$10%) and the form of the parametrization ($\pm$5%), resulting in an overall uncertainty of $\pm$15% in the unfolding factor.

5.3 Energy Flow from Multiple Interactions

The studies of the energy flow in these events, both on average and as a function of the number of primary vertex candidates, showed that multiple interaction cause a shift in the jet mass scale. We estimated this shift to be $3-4$ GeV/c^2 for jets with masses above 70 GeV/c^2 and a cone size of $R = 0.7$. We conservatively assign an uncertainty on this correction of order half the size of the shift, 2 GeV/c^2.

5.4 Uncertainties on the PYTHIA Predictions

We estimated the uncertainty on the PYTHIA prediction for the jet mass distribution by reweighting the MC events using the $\pm 1\sigma$ variations in the 20 eigenvectors describing the uncertainties in the PDFs and choices of scale, as has become standard practice [43]. Specifically, we used this approach to determine the changes in the shape of the normalized shape distributions. In all cases, these uncertainties were 10% or less for jet mass and the other substructure variables we have studied. These uncertainties, determined bin-by-bin, are shown where we compare data with the PYTHIA MC distributions.

6 Properties of Observed Jets

The mass distribution of all jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ in our sample, after the event selection and corrections for multiple interactions, is shown in Fig. 11 for jets with cone sizes of $R = 0.4$ and $R = 0.7$. We observe 1836 and 2108 events with cone sizes of $R = 0.4$ and $R = 0.7$, respectively, where the reduction in the number of events has come from the top-rejection requirements.

We also show in Fig. 12 the jet mass distribution for jets reconstructed with the three Midpoint cone sizes of $R = 0.4$, 0.7 and 1.0. We see that the cone size plays a clear role in limiting the high mass behaviour.

6.1 Jet Mass and Jet Function

We noted in the introduction that the jet mass distribution for highly boosted objects is characterized theoretically by a "jet function" that over a relatively wide range of large jet masses not only predicts the shape of the distribution but also its normalization (i.e., the fraction of jets that would have masses in the range relative to all the jets in the sample).

We find that we have to correct the observed jet mass distribution for the effect of the p_T cut on the jet, as the p_T resolution and jet mass are correlated. This unfolding factor for the jet mass distribution above 70 GeV/c^2 ranges from 1.5 to 2, and has an approximately 15% uncertainty arising from the hadronization model and the uncertainty on the jet p_T scale (as discussed above in Section 5).

We show in Fig. 13 a comparison of the observed mass distribution of the leading jet, corrected for multiple interactions (MI), with the analytic predictions for the jet function, plotting the function for both quark and gluon jets, for a cone size of $R = 0.4$. This comparison, made for jet masses above 70 GeV/c^2, shows that the theoretical prediction describes both the total rate and the shape of the mass distribution correctly over the jet mass range of 70 to 160 GeV/c^2. It is also consistent with the expectation that over 80% of these jets would arise from quark showering, given that a gluon jet would tend to have a larger high

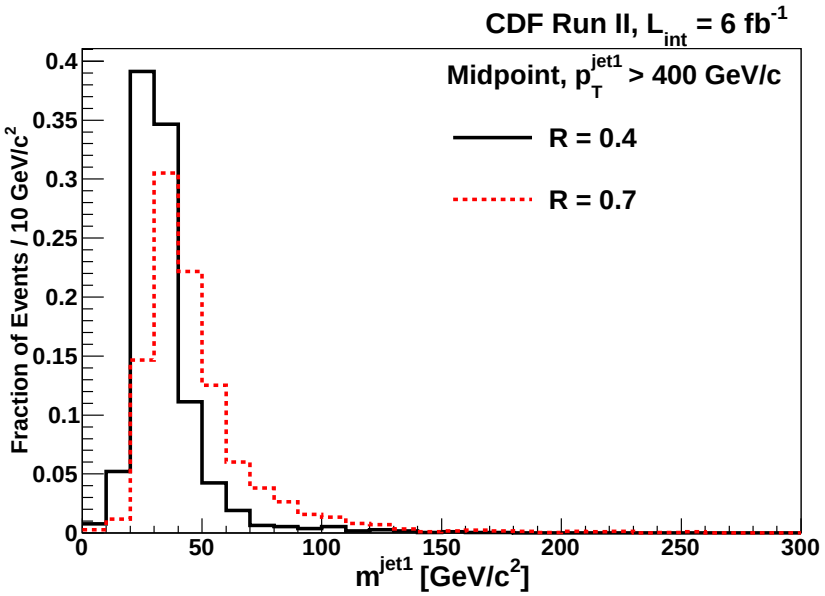

Figure 11: The mass distribution for the leading jet in events with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ for cone sizes of $R = 0.4$ and $R = 0.7$. The top rejection cuts have been applied, and the jet mass has been corrected for multiple interactions.

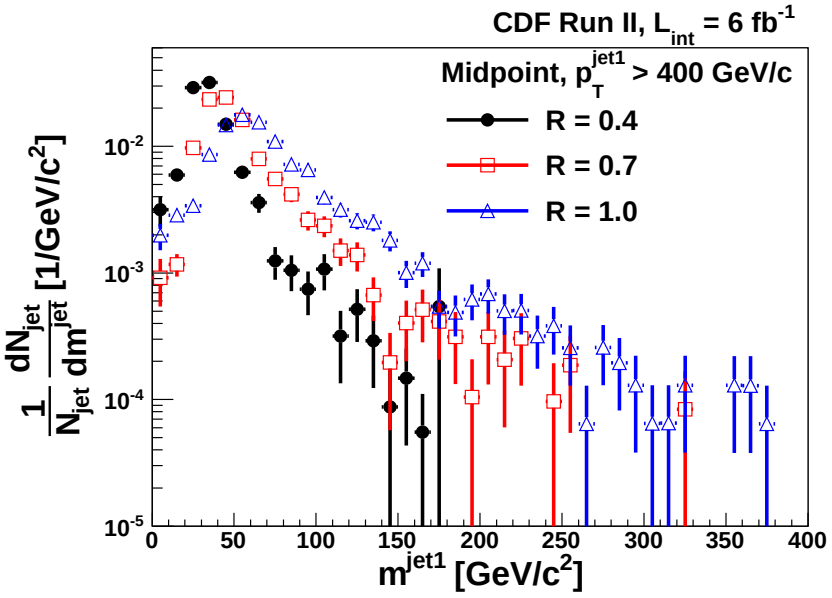

Figure 12: The jet mass distributions for the leading jet with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ for the three Midpoint cone sizes $R = 0.4$, 0.7 and 1.0. The top rejection cuts have been applied and the jet mass has been corrected for multiple interactions and the effect of the p_T selection on the jet mass distribution.

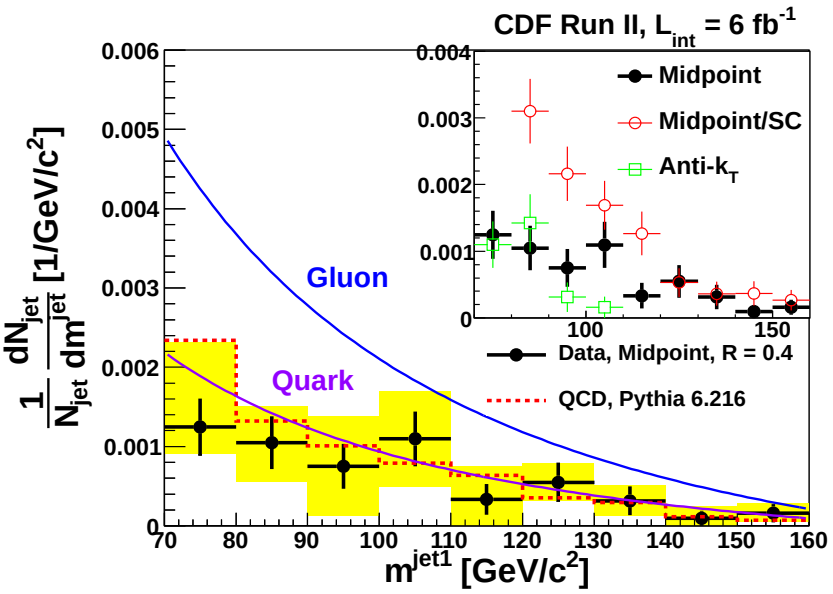

Figure 13: The jet mass distribution for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 0.4$ Midpoint cone algorithm and corrected for MI. We have rejected $t\bar{t}$ events, and show comparisons to the analytic expression for the jet function for quarks and gluons. We also show the PYTHIA prediction (red dashed line). The inset shows the comparison between the results of the three clustering algorithms.

mass tail. We note that the theoretical prediction is an absolute prediction, in that it gives the probability distribution for producing a jet with a given mass.

We expect that the perturbative NLO calculation for the jet mass would be sensitive to the cone size and so we show in Fig. 14 the corresponding mass distribution for the leading jet in the selected events constructed using a cone size of $R = 0.7$. We see good agreement between the data distribution and the theoretical prediction.

6.2 Angularity

The angularity of boosted jets have been considered one substructure variable that helps discriminate QCD jets from jets produced by other processes [32, 23]. We show in Fig. 15 the distribution of angularity for the leading jet in our sample requiring that $m^{jet1} \in (90, 120)$ GeV/c^2, comparing the observed angularity distribution with the prediction from PYTHIA and the limits arising from the analytical QCD prediction. We use the MI-corrected jet mass in selecting events for this distribution and following distributions. We show in Fig. 16 the angularity distribution for jets formed with a cone size of $R = 0.7$.

The PYTHIA distributions are in agreement with the data with jets reconstructed with the Midpoint algorithm. The analytic QCD predictions for τ_{min} and τ_{max} also agree well with the data for both cone sizes.

6.3 Planar Flow

As discussed in the theory introduction, the planar flow variable provides another IR-safe variable that characterizes top quark jets from QCD jets [20, 23, 27]. We note that this substructure variable is independent

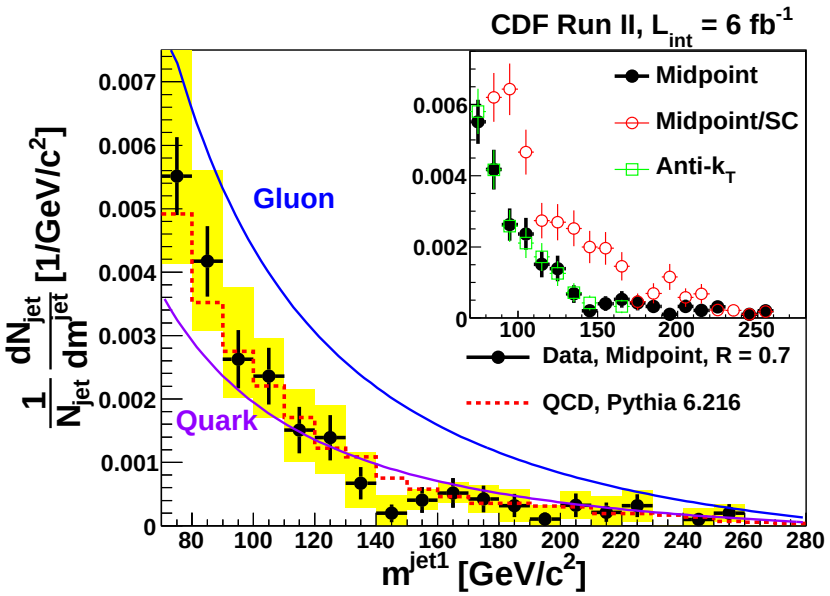

Figure 14: The jet mass distribution for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 0.7$ Midpoint cone algorithm and corrected for MI. We have rejected the $t\bar{t}$ events, and show comparisons to the analytic expression for the jet function for quarks and gluons. We also show the PYTHIA MC prediction (red dashed line). The inset shows the comparison between the results of the three clustering algorithms.

of jet mass and therefore provides significant additional separation between QCD jets and boosted objects with multi-prong decay kinematics.

We show in Figs. 17 and 18 the planar flow distribution for the data compared with the PYTHIA MC prediction, where in the first case we do not make a jet mass cut and in the second we require that the jets be in the mass window $m^{jet} \in (130, 210)$ GeV/c^2. As illustrated, the planar flow changes significantly for QCD jets as a function of the jet mass and provides additional separation power between QCD jets and those arising from top quark production . These results are in reasonably good agreement with theoretical predictions. In particular, we see that when no jet mass cut is applied, the planar flow of QCD jets is monotonically increasing, as predicted by the PYTHIA calculation though the prediction has a lower slope than the data.

In order to be sure that we are looking at QCD candidates, we apply the top rejection cuts to the events in the high mass window of 130 and 210 GeV/c^2. The results are shown in Fig. 19. The good agreement of the data with the PYTHIA QCD prediction supports the premise that this sample is dominated by QCD scattering, though the statistical power of this comparison is limited given the relatively small number of jets in this sample.

For completeness, we show in Figs. 20 and 21 the planar flow distributions for jets reconstructed with a cone size of $R = 1.0$, making the same requirements on the leading jet and before and after making top rejection cuts, respectively. The larger cone size has greater efficiency for capturing all the decay products of the leading top quark jet, making it the preferred algorithm for separating QCD and top quark candidates. In this case, we have performed an unfolding correction that corrects the measured distribution to parton level jets. We note that the observed distribution prior to making top rejection cuts is in reasonable agreement

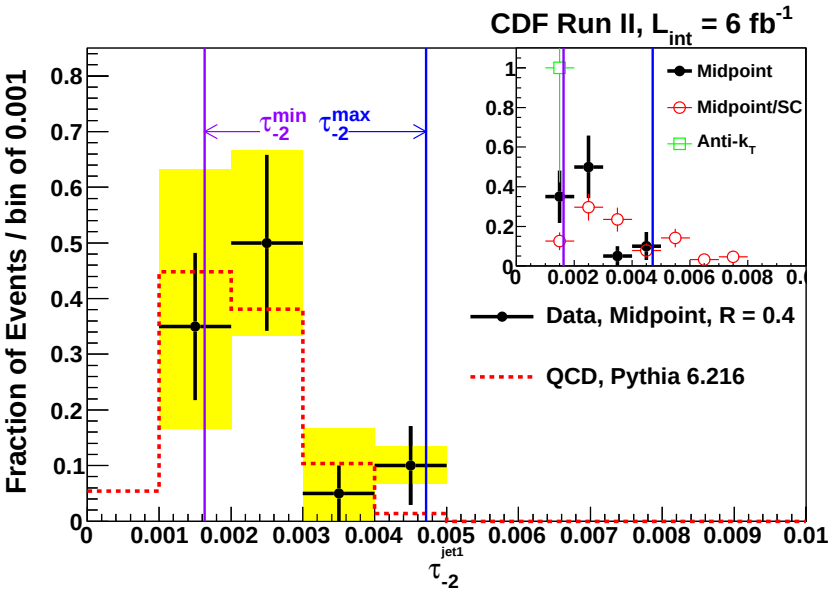

Figure 15: The angularity distribution for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 0.4$ Midpoint cone algorithm. We have rejected the $t\bar{t}$ events and required that $m^{jet1} \in (90, 120)$ GeV/c^2. We also show the results from the PYTHIA calculation, as well as the limits predicted by NLO QCD predictions. The inset shows the comparison between the results of the three clustering algorithms.

with the QCD prediction with an excess at larger planar flow consistent with a small admixture in the sample of top quark jets. The excess at larger planar flow is reduced once we apply the top rejection cuts. We also note that the Midpoint and anti-k_T algorithms give comparable results, while the Midpoint/SC algorithm results in jets with larger planar flows. Hence the algorithm differences seen are largely independent of jet cone size.

7 Conclusion

We have made significant progress on understanding the nature of very high p_T jets at CDF, especially their substructure. We have developed a technique to correct for the effects of multiple interactions to jet mass, angularity and planar flow. There is good agreement between the jet mass distributions and the PYTHIA 6.216 Monte Carlo calculations, and also in agreement with NLO QCD calculations. The agreement between the analytical theoretical calculations and the observed data for jet masses above 70 GeV/c^2 indicates that these theoretical models can be used to extrapolate to searches for new phenomena at the LHC. The measurements of the angularity of QCD jets produced with masses > 90 GeV/c^2 show that these are consistent with the NLO prediction of two-body structure, and the planar flow distribution for jets with masses between 130 and 210 GeV/c^2 show similar consistency with QCD predictions.

We have compared several different jet-finding algorithms, and find very similar results between the anti-k_T and Midpoint algorithms. We find that that Midpoint algorithm employing seach cones generates more massive jets on average, as well as jets with larger angularity and planar flows.

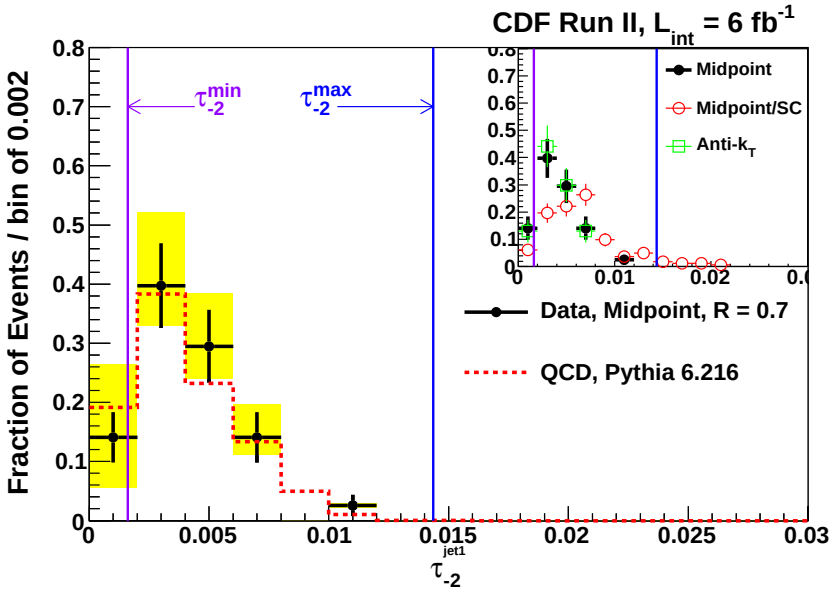

Figure 16: The angularity distribution for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 0.7$ Midpoint cone algorithm. We have rejected the $t\bar{t}$ events and required that $m^{jet1} \in (90, 120)$ GeV/c^2. We also show the results from the PYTHIA calculation (red dashed line), as well as the analytic QCD prediction (dotted). The inset shows the comparison between the results with the three clustering algorithms.

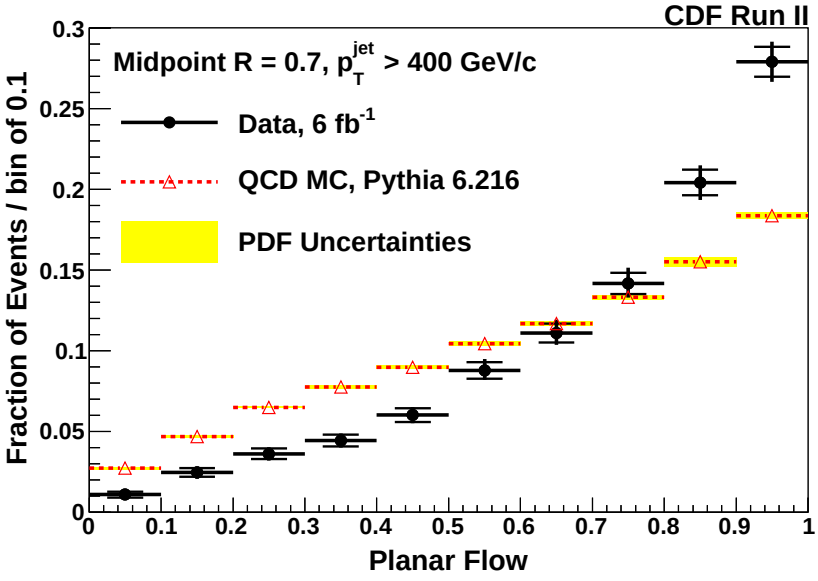

Figure 17: The planar flow distribution for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 0.7$ Midpoint cone algorithm. We have not rejected $t\bar{t}$ events and have made no requirements on the mass of the jet. We also show the PYTHIA prediction for QCD jets (red dashed line).

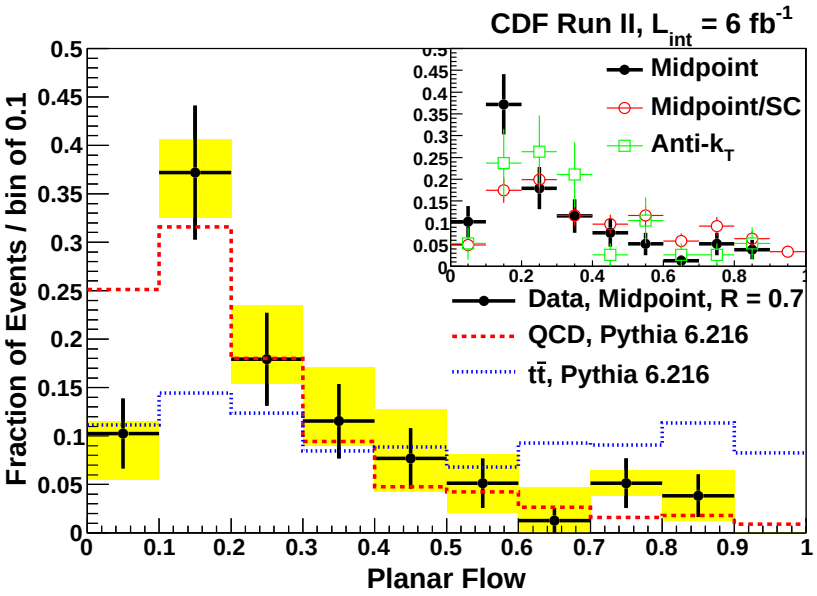

Figure 18: The planar flow distribution for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 0.7$ Midpoint cone algorithm. We have not rejected $t\bar{t}$ events and have required the jet mass to be between 130 and 210 GeV/c^2. We also show the PYTHIA prediction for QCD jets and for $t\bar{t}$ events (red and blue dashed histograms, respectively). Each plot is normalized to unity and the $t\bar{t}$ rate is expected to be significantly smaller than that from QCD. The inset shows the comparison between the results with the three clustering algorithms.

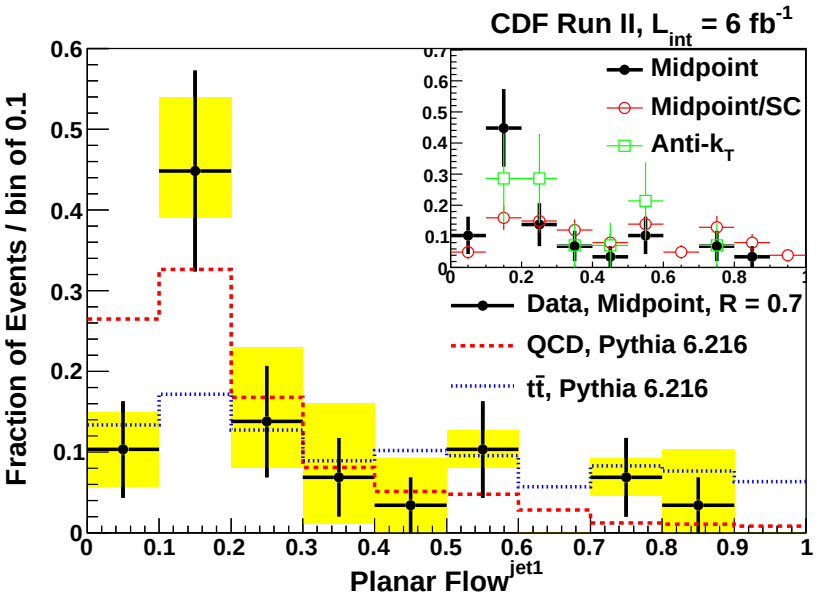

Figure 19: The planar flow distributions for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 0.7$ Midpoint cone algorithm after applying the top rejection cuts and requiring that $m^{jet1} \in (130, 210)$ GeV/c^2 (after MI corrections). An unfolding correction has been applied. We also show the PYTHIA predictions for QCD and $t\bar{t}$ events (red and blue dashed histograms, respectively). Each plot is normalized to unity and the $t\bar{t}$ rate is expected to be signifcantly smaller than that from QCD. The inset shows the comparison between the results with the three clustering algorithms.

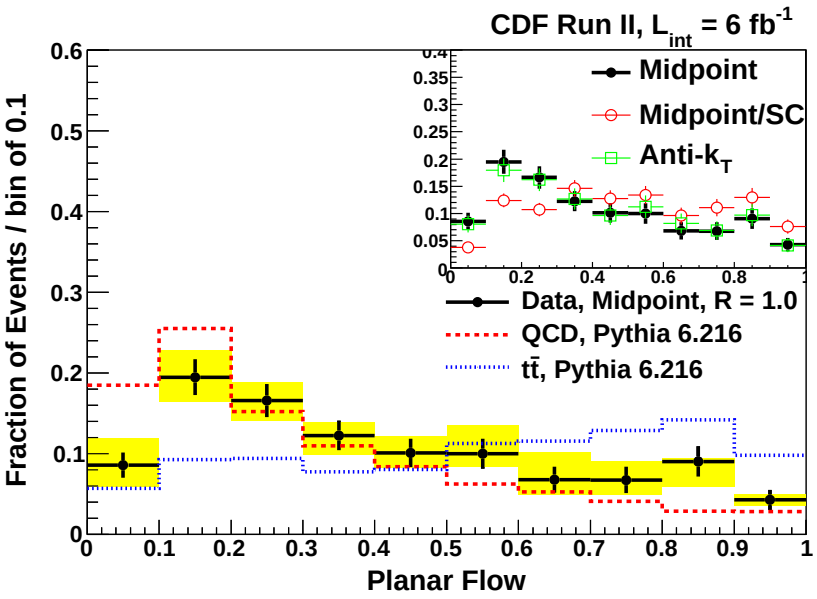

Figure 20: The planar flow distributions for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 1.0$ Midpoint cone algorithm requiring that $m^{jet1} \in (130, 210)$ GeV/c^2 (after MI corrections). We also show the PYTHIA predictions for the distributions associated with QCD and $t\bar{t}$ events (red and blue dashed histograms, respectively). The inset shows the comparison between the results with the three clustering algorithms. No top rejection has been applied.

Acknowledgements

We acknowledge the contributions of numerous theorists for insights and calculations. Special thanks go to Ilmo Sung and George Sterman, for their elucidation of the "soft function" that determines the behaviour of QCD jets at low jet masses, to Gavin P. Salam for clarifying aspects of Fastjet and to Nikolaos Kidonakis for performing an updated calculation of the top quark p_T distribution at NNLO and providing us with an estimate of the uncertainties in the cross section for top quarks with $p_T > 400$ GeV/c.

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 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; and the Academy of Finland.

This work was supported in part by a grant from the Shrum Foundation. One of us (P. Sinervo) gratefully acknowledges the Rosi and Max Varon Visiting Professorship of the Weizmann Institute of Science for supporting the early stages of this study. G. Perez, E. Duchovni and R. Alon were supported by the Weizmann Institute of Science and one of them (G.Perez) is supported by the Isreal Science Foundation

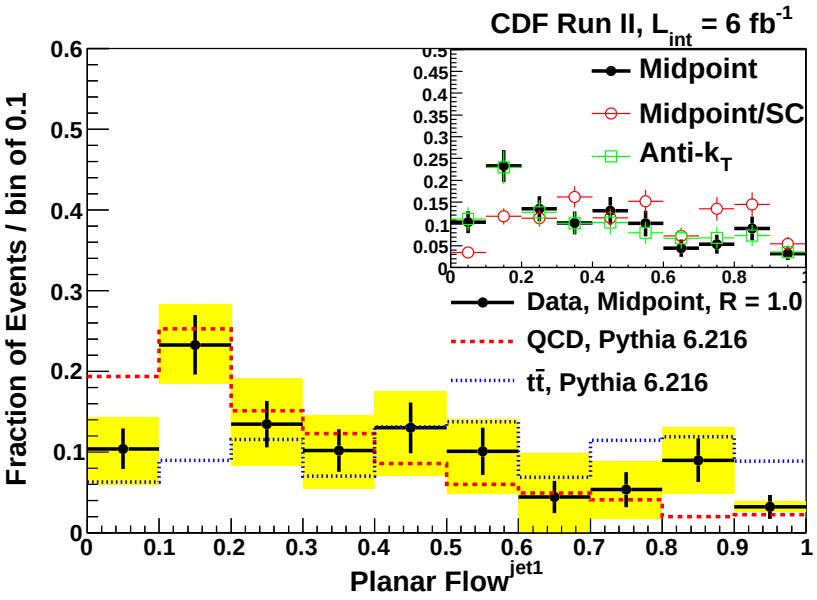

Figure 21: The planar flow distributions for jets with $p_T > 400$ GeV/c and $|\eta| \in (0.1, 0.7)$ reconstructed with an $R = 1.0$ Midpoint cone algorithm after applying the top rejection cuts and requiring that $m^{jet1} \in (130, 210)$ GeV/c^2 (after MI corrections). We also show the PYTHIA predictions for QCD and $t\bar{t}$ events (red and blue dashed histograms, respectively). Each plot is normalized to unity and the $t\bar{t}$ rate is expected to be significantly smaller than that from QCD. The inset shows the comparison between the results with the three clustering algorithms.

(grant #1087/09), Marie Curie, IRG fellowship and the Peter & Patricia Gruber Award.

References

[1] S. D. Ellis, J. Huston, K. Hatakeyama, P. Loch and M. Tonnesmann, Prog. Part. Nucl. Phys. **60**, 484 (2008) [arXiv:0712.2447 [hep-ph]].

[2] G. P. Salam, arXiv:0906.1833 [hep-ph].

[3] T. Han, S.J. Lee, F. Maltoni, G. Perez, Z. Sullivan, T.M.P. Tait and L. T. Wang, part of Nucl. Phys. Proc. Suppl. **200-202**, 185 (2010) [arXiv:1001.2693 [Unknown]].

[4] V. M. Abazov *et al.* [DZero Collaboration], Phys. Rev. D **65**, 052008 (2002) [arXiv:hep-ex/0108054].

[5] D. E. Acosta *et al.* [CDF Collaboration], Phys. Rev. D **71**, 112002 (2005) [arXiv:hep-ex/0505013].

[6] M. H. Seymour, Z. Phys. C **62** (1994) 127; D. Benchekroun, C. Driouichi, A. Hoummada, SN-ATLAS-2001-001, ATL-COM-PHYS-2000-020, EPJ Direct 3, 1 (2001); J. M. Butterworth, B. E. Cox and J. R. Forshaw, Phys. Rev. D **65**, 096014 (2002) [arXiv:hep-ph/0201098].

[7] K. Agashe, A. Belyaev, T. Krupovnickas, G. Perez and J. Virzi, Phys. Rev. D **77**, 015003 (2008) [arXiv:hep-ph/0612015].

[8] A. L. Fitzpatrick, J. Kaplan, L. Randall and L. T. Wang, JHEP **0709**, 013 (2007) [arXiv:hep-ph/0701150].

[9] B. Lillie, L. Randall and L. T. Wang, JHEP **0709**, 074 (2007) [arXiv:hep-ph/0701166].

[10] K. Agashe, H. Davoudiasl, G. Perez and A. Soni, Phys. Rev. D **76**, 036006 (2007) [arXiv:hep-ph/0701186].

[11] K. Agashe *et al.*, Phys. Rev. D **76**, 115015 (2007) [arXiv:0709.0007 [hep-ph]].

[12] J. M. Butterworth, J. R. Ellis, A. R. Raklev and G. P. Salam, Phys. Rev. Lett. **103**, 241803 (2009) [arXiv:0906.0728 [hep-ph]].

[13] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008) [arXiv:0802.2470 [hep-ph]].

[14] G. D. Kribs, A. Martin, T. S. Roy and M. Spannowsky, arXiv:0912.4731 [hep-ph].

[15] T. Plehn, G. P. Salam and M. Spannowsky, Phys. Rev. Lett. **104**, 111801 (2010) [arXiv:0910.5472 [hep-ph]].

[16] T. Affolder *et al.* [CDF Collaboration], Phys. Rev. Lett. **87**, 102001 (2001).

[17] V. M. Abozov *et al.* [DZero Collaboration], arXiv:1001.1900 (January 2010).

[18] N. Kidonakis and R. Vogt, Phys. Rev. D **68**, 114014 (2003) [hep-ph/0308222].

[19] J. C. Collins, D. E. Soper and G. Sterman, Adv. Ser. Direct. High Energy Phys. **5**, 1 (1988) [arXiv:hep-ph/0409313].

[20] L. G. Almeida, S. J. Lee, G. Perez, I. Sung and J. Virzi, Phys. Rev. D **79**, 074012 (2009) [arXiv:0810.0934 [hep-ph]].

[21] W. Skiba and D. Tucker-Smith, Phys. Rev. D **75**, 115010 (2007) [arXiv:hep-ph/0701247].

[22] B. Holdom, JHEP **0708**, 069 (2007) [arXiv:0705.1736 [hep-ph]].

[23] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, I. Sung and J. Virzi, Phys. Rev. D **79**, 074017 (2009) [arXiv:0807.0234 [hep-ph]].

[24] H. Contopanagos, E. Laenen and G. Sterman, Nucl. Phys. B **484**, 303 (1997) [arXiv:hep-ph/9604313].

[25] M. Dasgupta and G. P. Salam, J. Phys. G **30**, R143 (2004) [arXiv:hep-ph/0312283]; M. Dasgupta, arXiv:1006.5679 [hep-ph]; A. Banfi, M. Dasgupta, K. Khelifa-Kerfa and S. Marzani, arXiv:1004.3483 [hep-ph]; S. D. Ellis, C. K. Vermilion, J. R. Walsh, A. Hornig and C. Lee, arXiv:1001.0014 [hep-ph].

[26] J. M. Butterworth, J. R. Ellis and A. R. Raklev, JHEP **0705** (2007) 033 [arXiv:hep-ph/0702150].

[27] J. Thaler and L. T. Wang, JHEP **0807**, 092 (2008) [arXiv:0806.0023 [hep-ph]].

[28] D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie, Phys. Rev. Lett. **101** (2008) 142001 [arXiv:0806.0848 [hep-ph]].

[29] D. Krohn, J. Shelton and L. T. Wang, arXiv:0909.3855 [hep-ph].

[30] G. Brooijmans, ATLAS note, ATL-PHYS-CONF-2008-008; ATL-COM-PHYS-2008-001;ATL-PHYS-PUB-2009-081; J. Conway, *et. al.*, LPC Workshop on Early Physics at CMS, UC Davis (2007); G. Brooijmans *et al.*, arXiv:0802.3715 [hep-ph].

[31] D. Krohn, J. Thaler and L. T. Wang, arXiv:0912.1342 [hep-ph]; S. D. Ellis, A. Hornig, C. Lee, C. K. Vermilion and J. R. Walsh, arXiv:1001.0014 [hep-ph]; S. D. Ellis, C. K. Vermilion and J. R. Walsh, arXiv:0912.0033 [hep-ph]; S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **80**, 051501 (2009) [arXiv:0903.5081 [hep-ph]].

[32] C. F. Berger, T. Kucs and G. Sterman, Phys. Rev. D **68**, 014012 (2003) [arXiv:hep-ph/0303051];

[33] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804** (2008) 063 [arXiv:0802.1189 [hep-ph]].

[34] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman and I. Sung, arXiv:1006.2035 [Unknown].

[35] R. Alon, O. Gedalia, G. Perez, G. Sterman and I. Sung, unpublished.

[36] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. D **78**, 052006 (2008) [Erratum-ibid. D **79**, 119902 (2009)] [arXiv:0807.2204 [hep-ex]].

[37] G. C. Blazey *et al.*, arXiv:hep-ex/0005012; Published in "Proceedings of the Physics at RUN II: QCD and Weak Boson Physics Workshop," Batavia, Illinois (1999).

[38] M. Cacciari, G.P. Salam and G. Soyez, Phys. Lett. **B641**, 57 (2006) [hep-ph/0512210]. We use Fastjet Version 2.4.2.

[39] A. Bhatti *et al.*, Nucl. Instrum. Meth. A **566**, 375 (2006) [arXiv:hep-ex/0510047].

[40] N. Kidonakis, personal communication.

[41] T. Aaltonen et al. [CDF Collaboration], CDF Conference Note 9913 (October 2009).

[42] R. Alon, E. Duchovni, G. Perez, A. Pranko and P. Sinervo, arXiv:1101.3002 [hep-ph] (January 2011).

[43] J. Pumplin *et al.*, Phys. Rev. D **65**, 014013 (2001).