DØnote 4534-CONF

Angular correlations between b jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV

The DØ Collaboration
URL http://www-d0.fnal.gov
(Dated: August 19, 2004)

We present a preliminary measurement of angular correlations between b jets. The distribution of $\Delta\phi$ between a muon-tagged jet and a lifetime-tagged jet in data is compared to individual distributions for gluon splitting, flavor excitation and flavor creation $b\bar{b}$ processes, and background. The relative contributions of gluon splitting, flavor excitation and flavor creation to the data distribution are determined.

I. INTRODUCTION

We present a preliminary measurement of angular correlations in b jet production. Recent efforts (e.g. [1]) to describe the production of b jets in $p\bar{p}$ collisions should be compared to the data. In addition, b jet hadroproduction is an important background to many other interesting signals, including top and Higgs signals, and should therefore be well understood.

In $p\bar{p}$ collisions, b quarks are predominantly produced through the strong interaction. At leading order, b quark pairs are produced through the *flavor creation* (FCR) process. At next-to-leading order, additional processes have to be taken into account. A distinction can be made between FCR, *gluon splitting* (GSP) and *flavor excitation* (FEX) processes. The distinction between the processes, which is more appropriate for Monte Carlo event generation than to NLO QCD, is described in more detail in [1, 2]. In standard Monte Carlo event generators, only the FCR process is normally included in the matrix element; GSP and FEX processes are generated as final- and initial-state shower processes, respectively. Previous measurements at $\sqrt{s} = 1.8$ TeV [3–6] have shown that the contribution of higher-order processes is significant.

In leading order QCD, the b quarks are predominantly produced back-to-back in azimuth. The GSP process, however, favors b quarks that are very close in phase space. In FEX processes, the b quarks are not strongly correlated in azimuth, but one b will be produced at high momentum relative to the incoming (anti-) proton beam while the other b quark will emerge very close to the beam direction.

The difference in the azimuthal correlations between the three processes is used to determine their relative contribution to the total b jet cross section. A b jet is defined in Monte Carlo as a jet with a matched B hadron within a cone of size $\Delta R < \sqrt{\Delta\phi^2 + \Delta\eta^2}$. The angular correlations between jets in a data sample with a jet with a matched muon and an impact parameter tagged jet are compared with the separate contributions in simulated Monte Carlo events.

The relative contribution from each process and the contribution from lighter flavors is determined using a pseudo-2D maximum likelihood fit to the P_T^{Rel} distribution of the muon jet and the azimuthal correlation $\Delta\phi$ between the muon jet and the impact parameter jet.

II. DATA SAMPLE AND EVENT SELECTION

The data used were taken between August 22, 2002 and October 24, 2002 and selected with the `MU_JT20_L2M0` trigger, requiring the presence of the following signals:

- A 5 GeV calorimeter tower at Level 1.
- Muon scintillator hits in a single region of the detector at Level 1.
- A muon track at Level 2.
- A 20 GeV jet at Level 3.

For the selected samples, the trigger efficiency only depends on jet E_T. Runs with bad tracking, calorimetry or muon reconstruction were rejected. The data were reconstructed using the fall 2002 version of the DØ reconstruction code. The data set corresponds to 7.8 ± 0.8 pb^{-1} of collision data.

The events for the analysis were required to meet the following requirements:

- At least two reconstructed jets of good quality (i.e. passing cuts designed to remove noise jets and electrons faking jets) with $|\eta| < 1$ and $E_T^{CAL} > 15$ GeV, reconstructed with the improved legacy cone algorithm [7] with cone size $R < 0.5$.
- A tight muon in the central muon system with $p_T > 6$ GeV/c and matched to one of the jets within $\Delta R < 0.7$.
- A reconstructed primary vertex with $|z| < 22$ cm.

These events constitute the *analysis sample*. The jet with a matched muon is referred to as the *muon jet*. The second jet is referred to as the *away jet*. If more than two jets are present in the event, all possible combinations are considered.

Muons were reconstructed in the muon system only. The muon ϕ coordinate was extrapolated to the value at the beam position to account for bending in the magnetic field of the solenoid, based on an ad-hoc parametrisation of the shift in ϕ of the local muon with respect to a matched central track as a function of local muon p_T. The η coordinate of the local muons was re-calculated using the primary vertex to improve the η resolution.

The jet energy was corrected for the jet energy scale (typically an increase of 35-40%). If a matching muon was found, the muon 4-vector was added to the jet 4-vector and the energy of the jet was corrected for the mean energy deposit of the muon in the calorimeter and the energy carried away by the undetected neutrino. Throughout this note, E_T refers to the energy of the jet including the matched muon, if present. The calorimeter-only jet energy is referred to as E_T^{CAL}.

To determine the efficiency of the impact parameter tag, two additional samples were used: a *muon+jet* sample with at least one jet with a matched muon and a *background jet* sample with at least one jet and a photon at $\Delta\phi > 2.8$ from that jet.

III. MONTE CARLO SAMPLES

The Monte Carlo samples were generated with PYTHIA 6.158, followed by a GEANT simulation of the DØ detector response and reconstruction. The CLEO QQ [8] package was used for B decays. The $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ resonances and all anti-b mesons were forced to decay directly to μX or τX. Cascade decays ($b \to c \to \mu$) were not forced but included in the final sample when they occurred. Events are reweighted to reproduce the correct fraction of $b \to c \to \mu$ events. All samples were generated with a cutoff on the particles participating in the hard scatter of $p_T > 20$ GeV/c. Five different samples were generated:

1. A FCR sample, generated by explicitly requiring two b quarks to be produced in the hard scatter.
2. A GSP sample, generated by running the inclusive QCD process and selecting events with a $b\bar{b}$ quark pair originating from a common gluon parent.
3. A FEX sample, generated by running inclusive QCD and selecting events with exactly one b quark participating in the hard scatter.
4. A $c\bar{c}$ background sample, generated by running inclusive QCD and selecting events in which c quarks were present.
5. A udsg background sample, generated by forcing a random $\pi^{\pm}$ or $K^{\pm}$ to decay before reaching the calorimeter. The events were weighted with the probability that the decay would occur naturally.

IV. MUON P_T^{REL} TEMPLATE DEFINITIONS

The semileptonic decay $b \to \mu X$ leads to the production of a jet with a matched muon or *muon jet*. In addition to b production, muon jets can also arise from semileptonic decays of c quarks and in-flight decays of $\pi^{\pm}$ and $K^{\pm}$ mesons. The background from c quarks and lighter flavors can be distinguished using the transverse momentum of the muon relative to the associated (muon+jet) axis (P_T^{Rel}) as a discriminator [9]. The higher mass of the decaying b quark generates a different P_T^{Rel} distribution than that obtained from quark decays of lighter flavor.

Individual P_T^{Rel} distributions for muons from b, c and $\pi^{\pm}/K^{\pm}$ decays were extracted from the Monte Carlo samples described in section III. The obtained distributions were used as *templates* for a likelihood fit to the observed P_T^{Rel} distribution in data. Muons from $b \to c \to \mu$ decays generate a P_T^{Rel} distribution very similar to that obtained from direct $c \to \mu$ and $\pi^{\pm}/K^{\pm}$ decays but were included in the b template. To account for the distortion of the fraction of $b \to c \to \mu$ decays relative to direct $b \to \mu$ decays caused by the forced decay of anti-b mesons, an additional weight was assigned to $b \to c \to \mu$ decays.

The events were selected using the same selection criteria as in the muon+jet data sample (section II), with the following additional requirements:

- For the b and c templates, the reconstructed muon was required to match a Monte Carlo muon originating from a b or c decay within a cone of size $\Delta R < 0.3$.
- Because we will explicitly request the presence of two separate tagged jets in the analysis, muons associated with $b\bar{b}$ jets, defined as jets with two matched B hadrons or a $b\bar{b}$ resonance within a $\Delta R < 0.5$ cone, were not included in the templates.

The jets and muons were smeared to account for the difference between the resolutions in data and Monte Carlo. The events were weighted with the probability to trigger the event based on the E_T^{CAL} of the leading jet in the event. Events in the $\pi^{\pm}/K^{\pm}$ template were weighted with the $\pi^{\pm}/K^{\pm}$ decay probability, and events in the b template were weighted to account for the relative fraction of $b \to c \to \mu$ decays.

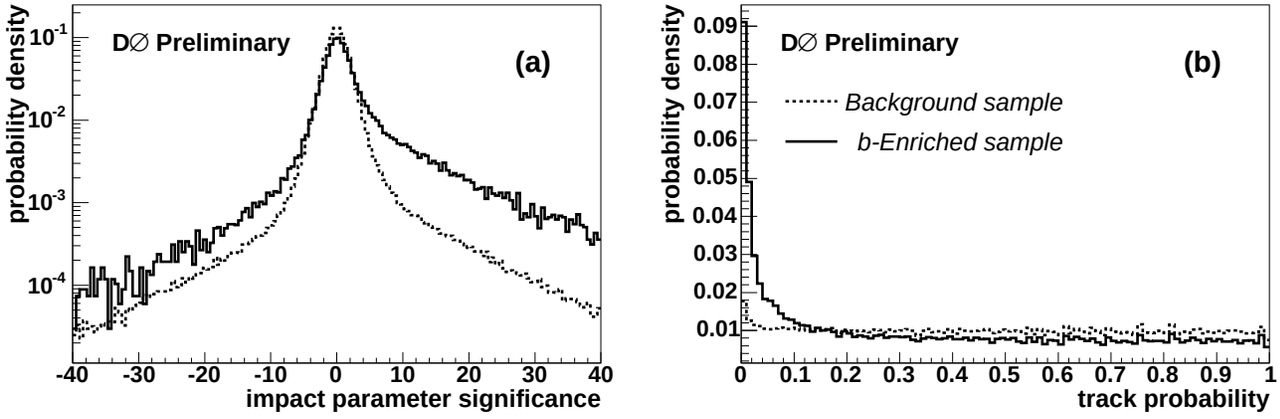

FIG. 1: Distributions of track impact parameter significance (a) and track background probability (b) for tracks matched to jets in the background sample and in a b-enriched sample.

V. IMPACT PARAMETER TAG

A. Method

To enhance the b fraction of the analysis sample, the away jet was tagged with an impact parameter tag. Because of the long lifetime of B hadrons, the trajectories of their decay particles will be displaced from the beam. The displacement of a particle trajectory or a reconstructed track is referred to as the impact parameter or distance of closest approach (dca). The impact parameter *significance* is defined as $s = \mathrm{dca}/\sigma_{\mathrm{dca}}$.

The dca of tracks is measured in the plane transverse to the beam direction. In this plane, the dca of a reconstructed track is assigned a positive sign if the track crosses the jet axis in the same hemisphere (defined by the jet and the beam) as the jet and negative otherwise. The signed significance distribution of tracks originating from zero-lifetime decays is symmetrically smeared around zero, while long-lived particle decays lead to a long tail at high significance. The significance distributions for tracks associated with jets in a background sample and tracks from a sample in which the b-content has been increased by requiring a matched muon with $P_T^{Rel} > 1$ GeV/c are shown in Fig. 1.

To determine the probability that a jet is consistent with the production of particles with zero lifetime, tracks are associated with the jet in a cone of size $\Delta R < 0.5$. For each track, the probability that the track originates from a zero-lifetime decay is computed by comparing the significance to a resolution function determined from data. Resolution functions are determined from the negative significance distribution in data, for several track quality categories. Track probability distributions are shown in Fig. 1. The jet probability is computed as a normalized product of the track probabilities:

$$P_{\mathrm{jet}} = \Pi \cdot \sum_{i=0}^{N-1} \frac{(-\ln \Pi)^i}{i!}, \tag{1}$$

where $N \geq 2$ is the number of positive significance tracks, $\Pi = \prod_{i=1}^{N} P_{track,i}$ and $P_{track,i}$ is the background probability of individual tracks. The tagging discriminant D is defined as

$$D = -\ln P_{\mathrm{jet}}. \tag{2}$$

Distributions of the jet background probability and the tagging discriminant for jets in the background sample and in the b-enriched sample are shown in Fig. 2. Figure 3 shows the P_T^{Rel} distributions in the analysis sample for jets tagged with $D > 5$ and for "anti-tagged" jets with $D < 5$.

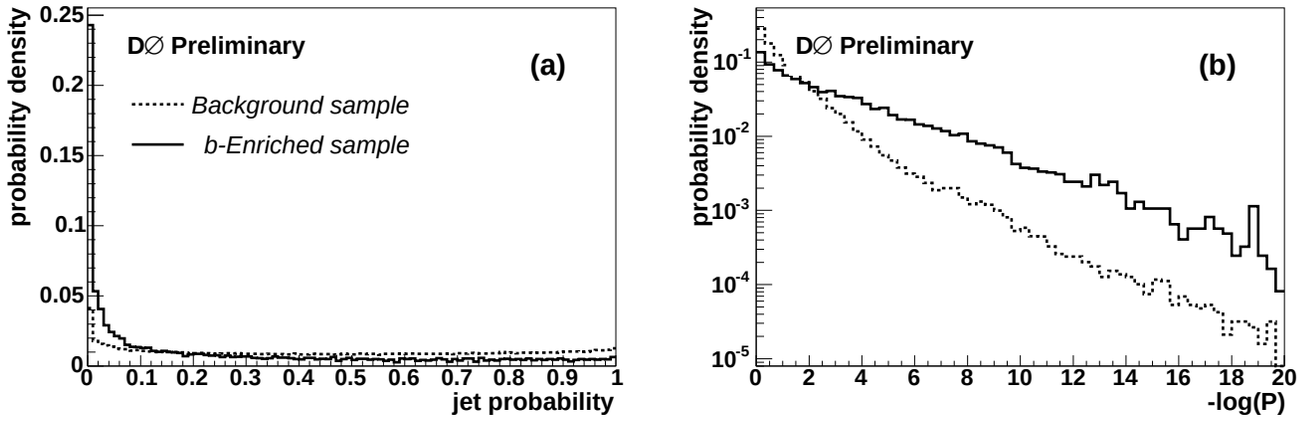

FIG. 2: Distributions of the jet background probability (a) and tagging discriminant (b) for jets in the background sample and in a b-enriched sample.

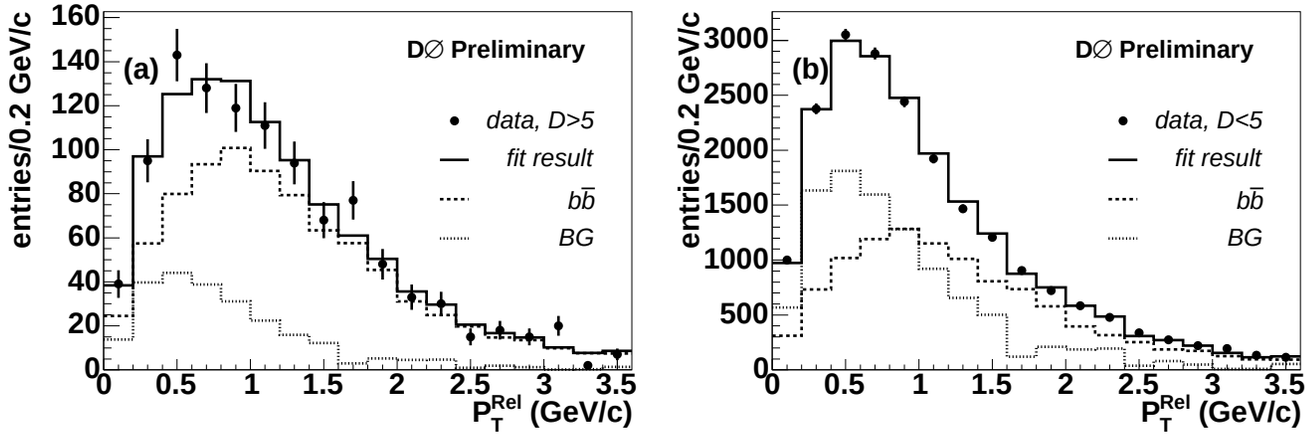

FIG. 3: Distributions of P_T^{Rel} in the analysis sample for jets tagged with $D > 5$ (a) and for "anti-tagged" jets with $D < 5$ (b).

B. Efficiency

The efficiency of the impact parameter tag for b jets is determined using the muon+jet sample. The number of b jets in the sample is estimated by fitting the P_T^{Rel} templates to the P_T^{Rel} distribution in data,

$$\epsilon_b = \frac{f_b^{\rm tag} \times N^{\rm tag}}{f_b^{\rm no\ tag} \times N^{\rm no\ tag}}, \tag{3}$$

where $N^{\rm no\ tag}$ and $N^{\rm tag}$ are the number of jets before and after replying the tag, and $f_b^{\rm no\ tag}$ and $f_b^{\rm tag}$ are the fractions of b-jets determined with the P_T^{Rel} template fit. The efficiency of the impact parameter tag is determined as a function of jet E_T, η and ϕ.

The efficiency for light jets is determined using the background jet sample. (The background from $c \to \mu$ decays is addressed as a systematic effect in section VIII.) The fraction of b jets in this sample is determined by counting the number of muon jets in the sample and determining the fraction of b jets in this subsample. The resulting number of b jets is divided by the branching ratio $BR(b \to \mu X)$ and the muon reconstruction efficiency to determine the real number of b jets in the background sample. The fraction of b jets in the total sample was estimated to be about 0.5%. The light jet efficiency is then determined as

$$\epsilon_{\rm background} = \frac{N^{\rm tagged}/N^{\rm tot} - \epsilon_b \times f_b}{1 - f_b}, \tag{4}$$

where ϵ_b is the b jet efficiency, f_b is the fraction of b jets in the total sample, and $N^{\rm tot}$ and $N^{\rm tagged}$ are the total number of jets and the number of tagged jets, respectively.

The fraction of jets with at least two matched tracks of sufficient quality (the *taggability*) is 0.57 ± 0.02 for muon jets and 0.526 ± 0.001 for light jets. The efficiency times taggability for the lifetime tag with $D > 5$ is 0.098 ± 0.004 for b jets and 0.0089 ± 0.0002 for light jets.

VI. $\Delta\phi$ TEMPLATE DEFINITIONS

The $\Delta\phi$ templates were generated using the samples described in section III. Reconstructed jets and muons were smeared to account for the resolution difference between the simulation and real data.

The event selection requirements were the same as for the analysis sample. Because of the limited size of the samples, a reconstructed muon and an impact parameter tag were not required in the $\Delta\phi$ templates. Instead, one of the jets was required to have $E_T > 21$ GeV to account for the additional energy carried by the muon. Each event was assigned a weight to account for the impact parameter tagging efficiency as a function of jet E_T, η and ϕ and for the probability to have a matched muon in the central region as a function of jet η and ϕ. The trigger efficiency as a function of the $E_T^{\rm CAL}$ of the leading jet in the event was taken into account with an additional weight.

VII. RESULTS

A. Signal and background sources

The possible configurations of the muon jet and the away jet are:

1. Both jets are b jets.
2. Both jets are not b jets.
3. The muon jet is a b jet but the away side is a falsely tagged light jet.
4. The muon jet is not a b jet but the away side jet is.

The first two components are correctly resolved by the 2D fit. Both components with exactly one fake tag are considered background, since no two b jets have been found. The third component occurs in GSP and especially FEX events where the second b jet may lie outside the acceptance of the analysis and the transverse energy in the event is balanced by a recoil jet of light flavor. These "fake IP tag" events are identified as $b\bar{b}$ events by the fit but have a different $\Delta\phi$ distribution. The fraction of these events is taken as an uncertainty on the result. The fourth component is identified as background by the fit but has a different $\Delta\phi$ distribution.

The contribution of fake IP tag events is estimated using a matrix method calculation.At different operating points (i) of the impact parameter tag, the number of $b \to \mu$ decays is given by:

$$N_{b\to\mu}^{(i)} = \epsilon_{\rm background}^{(i)} \times N_{\text{fake IP tag}} + \epsilon_b^{(i)} \times N_{b\bar{b}}, \tag{5}$$

where $N_{b\to\mu}^{(i)}$ is the estimated number of events remaining at operating point (i), $N_{\text{fake IP tag}}$ is the number of fake IP tag events and $N_{\rm signal}$ is the number of $b\bar{b}$ events in the untagged sample. Given the efficiencies for signal $\epsilon_{\rm signal}$ and background $\epsilon_{\rm background}$ at two operating points (1) and (2) the total number of fake IP tag events can be determined as:

$$N_{\text{fake IP tag}} = \frac{\epsilon_b^{(2)} \times N^{(1)} - \epsilon_b^{(1)} \times N^{(2)}}{\epsilon_b^{(2)} \times \epsilon_{\rm background}^{(1)} - \epsilon_b^{(1)} \times \epsilon_{\rm background}^{(2)}}. \tag{6}$$

As the first operating point we chose to use taggable jets. As the second operating point we chose a signal-enriched sample with $D > 5$. The efficiencies are given in section V B. $N_{b\to\mu}^{(1)}$ and $N_{b\to\mu}^{(2)}$ are given by the total number of muon jets at each operating point times the b fraction determined using the P_T^{Rel} fit. The resulting fraction of fake IP tag events is $(31 \pm 8)\%$ of the total number of events where the muon jet is a b jet. The distribution of these background events can be determined by performing the method in bins of $\Delta\phi$.

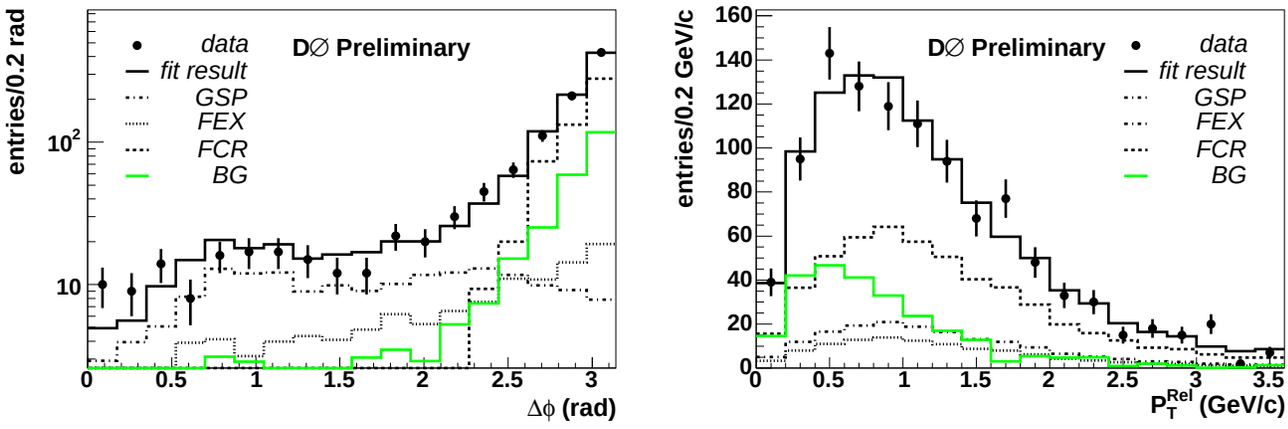

FIG. 4: Data $\Delta\phi$ and P_T^{Rel} distributions for jets with $E_T > 15$ GeV and $|\eta| < 1$. The weighted templates normalized to their total contribution to the fit result are also shown.

TABLE I: Relative contributions of different processes to the data distribution, after correction for fake IP tag background. The uncertainty shown is the statistical fit uncertainty.

FCR	0.45 ± 0.04
GSP	0.17 ± 0.03
FEX	0.10 ± 0.04
background	0.28 ± 0.04

B. Fit result

The P_T^{Rel} and $\Delta\phi$ templates were fit to the 2D data P_T^{Rel}-$\Delta\phi$ distribution. The P_T^{Rel}- and $\Delta\phi$ projections of the data with the fit result are shown in Fig. 4.

The fake IP tag background contribution was estimated from the P_T^{Rel}-only fit as described in section VII A. The background fraction was corrected to include the fake IP tag background events. The relative contributions of GSP, FEX and FCR to the total $b\bar{b}$ contribution were kept constant. The background correction is taken into account as a systematic uncertainty on the relative contributions.

VIII. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties on the result arise from the following sources:

- The fake IP tag background component is identified by the fit as signal but is in fact background. The fraction of fake IP tag events is small and is taken as an uncertainty on the fractions $f_{\rm FCR}$, $f_{\rm GSP}$ and $f_{\rm FEX}$.
- In the analysis, the background is assumed to be dominated by light jets. To evaluate the uncertainty due to the fraction of background from c-jets, the analysis is repeated using the $c\bar{c}$ template as the background template.
- The uncertainty due to the jet energy scale is evaluated by repeating the analysis with jets corrected for the energy scale $\pm 1\sigma$.
- The trigger efficiency is used to weight events in the Monte Carlo templates. The uncertainty due to the uncertainty on the trigger efficiency is evaluated by repeating the analysis with templates generated using the trigger efficiency $\pm 1\sigma$.
- The events in the Monte Carlo distributions are weighted with tag rate functions to account for the E_T, η and ϕ dependence of the impact parameter tag. The uncertainty due to the uncertainty on the tag rate functions is evaluated by repeating the analysis, changing the tag rate functions by $\pm 1\sigma$.
- The uncertainty due to the limited statistics in the Monte Carlo distributions can be taken into account by modifying the fit to take the Monte Carlo statistics into account [10].

TABLE II: Relative contributions of FCR, GSP and FEX to the production of two central b jets. Only statistical uncertainties are shown.

Process	PYTHIA	Data
FCR	0.47 ± 0.05	0.63 ± 0.04
GSP	0.43 ± 0.05	0.23 ± 0.04
FEX	0.10 ± 0.03	0.14 ± 0.05

- The uncertainty due to the modeling of the different distributions in Monte Carlo is taken into account by re-weighting the Monte Carlo to reproduce independent distributions in data.
- The uncertainty that may arise from the choice of the fragmentation function is still under study. Using jets and including the muon, if present, should decrease the dependence of the $\delta\phi$ distributions on the fragmentation function, so this uncertainty is not expected to be very large. The P_T^{Rel} distribution may be affected, however.

IX. CONCLUSIONS

We have presented a study of azimuthal correlations in b jet pair production in the central region. The relative contributions of FCR, GSP and FEX to the production of two central b-jets are shown in Table II. Only the fit uncertainties due to the statistics of the data sample are given. The relative contributions to an inclusive $b\bar{b}$ sample generated with PYTHIA 6.158 and corrected for the acceptance of the data analysis are also given along with a statistical uncertainty.

With the larger data sample and improved b-tagging in recent data, it will be possible to study the distribution of fake IP tag background effects which are a major source of uncertainty. The tighter tagging available in recent data will reduce the total fraction of background events, which in turn will reduce the uncertainty due to the choice of the background template. The uncertainty due to the jet energy scale correction can also be reduced with more recent data. The analysis can also be extended into the forward region ($|\eta| > 1$), increasing the acceptance for FEX events and making it possible to use $\Delta\eta$ or ΔR to better separate the GSP and FEX contributions.

Acknowledgments

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the Department of Energy and National Science Foundation (USA), Commissariat à L'Energie Atomique and CNRS/Institut National de Physique Nucléaire et de Physique des Particules (France), Ministry for Science and Technology and Ministry for Atomic Energy (Russia), CAPES, CNPq and FAPERJ (Brazil), Departments of Atomic Energy and Science and Education (India), Colciencias (Colombia), CONACyT (Mexico), Ministry of Education and KOSEF (Korea), CONICET and UBACyT (Argentina), The Foundation for Fundamental Research on Matter (The Netherlands), PPARC (United Kingdom), Ministry of Education (Czech Republic), Natural Sciences and Engineering Research Council and WestGrid Project (Canada), BMBF (Germany), A.P. Sloan Foundation, Civilian Research and Development Foundation, Research Corporation, Texas Advanced Research Program, and the Alexander von Humboldt Foundation.

[1] S. Frixione, P. Nason, and B. R. Webber, JHEP **08**, 007 (2003).
[2] R. Field, Phys. Rev. **D65**, 094006 (2002).
[3] F. Abe et al., Phys. Rev. **D53**, 1051 (1996).
[4] F. Abe et al., Phys. Rev. **D55** (1997).
[5] B. Abbott et al., Phys. Lett. **B487**, 264 (2000).
[6] D. Acosta et al., Phys. Rev. **D69**, 072004 (2004).
[7] G. Blazey et al., *Run II jet physics* (2000), hep-ex/0005012.
[8] The CLEO collaboration, http://www.lns.cornell.edu/public/CLEO/soft/QQ.
[9] O. Peters, Ph.D. thesis, University of Amsterdam (2003).
[10] R. Barlow and C. Beeston, Computer Physics Communications **77**, 219 (1993).