A Search for Anomalous Heavy-Flavor Quark Production in Association with W Bosons

We present a search for anomalous production of heavy-flavor quark jets in association with a W boson at the Fermilab Tevatron pp Collider. This search is conducted through an examination of the exclusive jet spectrum of W +jets final states in which the heavy-flavor quark content has been enhanced by requiring at least one tagged jet in an event. Jets are tagged by the combined use of two algorithms, one based on semileptonic decays of $b/c$ hadrons, and the other on their lifetimes. We compare data in $e^+\text{jets}$ (164 pb$^{-1}$) and $\mu^+\text{jets}$ (145 pb$^{-1}$) channels, collected with the D0 detector at $\sqrt{s} = 1.96$ TeV, to expectations from the standard model, and set upper limits on anomalous production of such events.

PACS numbers: 13.85.Qk, 13.85.Ni

The heavy-flavor (HF) content of jets produced in association with a W boson in $p\bar{p}$ collisions provides a test of the standard model (SM), and an excess would suggest a non-SM source of physics. The CDF collaboration recently reported just such an excess in the exclusive $W+$HF-jet spectrum in which one jet was tagged using both
secondary-vertex (SVT) and soft-lepton (SLT) tagging algorithms [1]. To check for the presence of this anomaly in our data, we also select jets tagged with both algorithms. In addition, we use two benchmark SM processes as models for new physics and derive upper limits on such processes.

At the Tevatron, the primary SM contributions to a W boson associated with HF quarks in the final state are expected to be from $t\bar{t}$, $Wb\bar{b}/cc$ (where the $b\bar{b}$ or $cc$ pairs arise from gluon splitting), and $Wc$ final states, with additional contributions arising from single top quark or $WZ$ (with $Z \rightarrow b\bar{b}/cc$) production. The production of W bosons accompanied by light quarks or gluons (referred to as $W$+jets in this Letter) contributes to the background when the light-quark or gluon jets are misidentified as jets from HF quarks. Since $W$ bosons are identified through their $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ decays, background can arise from $Zb\bar{b}$, $ZZ$ (with one $Z \rightarrow b\bar{b}/cc$), and $Z$+jets production when one of the leptons from the $Z \rightarrow \ell^+\ell^-$ decay is not observed in the detector. The main instrumental background arises from multijet processes in which a jet is misidentified as a lepton, and an imbalance in transverse momentum ($p_T$) is generated through a mis-measurement of the jets or a lepton. To be selected, these kinds of events must also contain tagged HF jets or misidentified non-HF jets.

The data were collected with the DØ detector [2] during Run II of the Fermilab Tevatron Collider in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. The components used in this analysis include the central tracker, calorimeter, and muon detectors. The central tracker consists of a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located within a 2 T superconducting solenoidal magnet. The uranium/liquid-argon calorimeter consists of three sections, each housed in a separate cryostat [3]. The central calorimeter (CC) covers pseudorapidity $|\eta| \leq 1.1$, while the two end calorimeters (EC) extend the coverage to $|\eta| \approx 4.0$. The muon system is located outside the calorimeters, and consists of a layer of tracking detectors and scintillation trigger points, clustering calorimeter energy within $R = 0.5$.

The average trigger efficiency for electrons with $p_T > 20$ GeV/c and $|\eta| < 1.1$, defined relative to the geometrical center of the detector. Lepton isolation requires a separation in $\eta$ and azimuth ($\phi$) of $R = \sqrt{(|\Delta\eta|)^2 + (|\Delta\phi|)^2} > 0.5$ from all jets in the event. Electrons are defined using a cone algorithm, and by the energies deposited in calorimeter towers within a radius of $R = 0.2$ of the electron axis, with at least 90% required to be within the electromagnetic portion of the calorimeter, and by the total energy in a cone of $R = 0.4$ centered on the same axis, which must not exceed by more than 15% the reconstructed electron’s energy. In addition, the longitudinal and transverse shower shape must be compatible with that expected from an electron.

Candidate events for $W \rightarrow \mu\nu$ decays must contain exactly one isolated muon with $p_T > 20$ GeV/c and $|\eta| < 1.6$, also defined relative to the geometrical center of the detector. Muons are required to satisfy two additional isolation criteria: the transverse energy deposited in the calorimeter in the annular region of $0.1 < R < 0.4$ around the muon’s path must be smaller than 2.5 GeV; and the vector sum of the $p_T$ values of all tracks within $R = 0.5$ of the muon’s trajectory must be less than 2.5 GeV/c (excluding the track matched to the muon’s trajectory).

Lepton identification is refined by requiring the trajectory of a track reconstructed in the SMT and CFT to match either the position of the electron energy cluster in the calorimeter or the position of hits in the muon detector. To complete the selection, all events are also required to have $E_T > 20$ GeV, and the azimuthal angle between the lepton and the direction of the $E_T$ must be greater than $\pi/8$. To eliminate poorly reconstructed events, the primary vertex (PV) of the event must contain at least three tracks, and its z-position (along the beam) has to be closer than 60 cm from the center of the detector. Finally, to reject multijet background, we require a reconstructed transverse mass consistent with that of the $W$ boson, $40 < M_{W_T} < 120$ GeV/c². In calculating $M_{W_T}$, we assume that the $E_T$ corresponds to the transverse energy of the neutrino.

Upon selection of $W$-boson candidates, we evaluate the HF-quark content of each event. Jets are defined using an iterative seed-based cone algorithm (including midpoints), clustering calorimeter energy within $R = 0.5$. This is subsequently corrected for jet energy scale, based on momentum balance in photon+jet events [4]. We consider only jets with $E_T > 25$ GeV/c and $|\eta| < 2.5$. These jets are then evaluated using two HF-tagging algorithms, as described below.

The soft-lepton tagging (SLT) algorithm is based on low-$p_T$ muons arising from semileptonic decays of $W$ quarks (via virtual $W$ bosons) that are produced near a jet in $(\eta, \phi)$ space. Only muons with $p_T > 4$ GeV/c and $|\eta| < 2.0$ are considered. To reject $Z \rightarrow \mu\nu$ background, we require $p_T < 15$ GeV/c for the muon. Jets with a muon within $R = 0.5$ of the jet axis are deemed tagged. Typical SLT efficiencies for $b$-quark jets are approximately 11%, and 0.4% for light-quark jets. The additional muon present in SLT events causes an increase in the average single-muon trigger efficiency from $(62.1 \pm 3.4)\%$ to $(68.4 \pm 3.5)\%$. 


Secondary-vertex tagging (SVT) is used to identify displaced decay vertices of long-lived particles. To form secondary vertices (SV), charged tracks are selected on the basis of the significance of their distance-of-closest-approach (dca) to the PV. Tracks are first grouped in $R = 0.5$ cones around a seed track with $p_T > 1$ GeV/c and $\text{dca}/\sigma_{\text{dca}} > 3.5$, where $\sigma_{\text{dca}}$ is the uncertainty on the track’s dca. Proto-vertices are formed by adding tracks to the initial grouping, provided their contribution to the $\chi^2$ of the vertex fit is small. Secondary vertices are selected by requiring the transverse distance from the SV to the beam direction, $L_{xy}$, to be less than 2.6 cm, and the decay-length significance, $L_{xy}/\sigma_{L_{xy}}$, to be greater than 7, where $\sigma_{L_{xy}}$ is the estimated uncertainty on $L_{xy}$ calculated from the error matrices of the tracks in the vertex. Jets are considered tagged by this algorithm when a SV lies within $R = 0.5$ of the original jet axis. This SVT algorithm exhibits a typical tagging rate for $b$-quark jets of $\sim 32\%$, and 0.25% for light-quark jets.

To predict SM rates, Monte Carlo (MC) events are generated for the processes mentioned above, with the exception of multijet production, which is estimated from data as described below. $W/Z + \text{jets}$ (both HF and light-quark jets), $t\bar{t}$, and diboson processes are simulated with ALPGEN [5]. Single top quark processes are simulated using COMPHEP [6]. All events are generated with $m_{t\bar{t}} = 175$ GeV/$c^2$. Hadronization and showering of these events is based on PYTHIA [7]. The exceptions are $W/Z + b$ processes, where $Zb$ is simulated using PYTHIA, and the contribution from $Wb$ is estimated from the parameterized MCFM MC [8] and used to calculate a cross section relative to $Wbb$ production assuming the jet-$p_T$ spectrum from PYTHIA for inclusive $W$ boson production. All MC events are generated at $\sqrt{s} = 1.96$ TeV, using CTEQ5L [9] parton distribution functions and a detailed detector simulation based on GEANT [10]. To simulate the effect of multiple interactions in beam crossings, a Poisson-distributed minimum-bias event overlay, with an average of 0.8 events, is included for all events. To avoid an incorrect combination of cross sections among simulated $W/Z + \text{jets}$ samples, only events with the same number of reconstructed jets as the number of initial partons are retained. The background from multijet events, in which a jet is misidentified as a lepton, is evaluated using the “matrix method” as follows. Two samples of $W + \text{jets}$ event candidates are used: a “tight” sample, for which the lepton identification criteria are as described above, and a “loose” sample in which some of these identification criteria are relaxed. The probabilities for true leptons to be identified as loose, and jets as tight leptons (as determined from independent studies of samples of pure leptons and pure jets) yields the fractions of true leptons and of misidentified jets in the tight and loose samples.

After the $W$-boson and jet selections, we apply the two HF-tagging algorithms to the jets. The MC samples are normalized to the appropriate luminosity, and corrected for differences in HF-tagging and lepton-identification efficiencies relative to data. Also, discrepancies in trigger efficiency between MC and data are corrected for each set of selections. In the following, the $e + \text{jets}$ and $\mu + \text{jets}$ samples are combined. Figures 1 and 2 show the exclusive number of jets in events with at least one SLT-tagged jet and at least one SVT-tagged jet, respectively. The transverse mass for $W$-boson candidate events containing at least one SLT or SVT-tagged jet, shown in Fig. 3, agrees well with SM expectation. The distribution for events with at least one jet tagged with both algorithms is shown in Fig. 4.

The dominant sources of experimental uncertainty are common to both the $e + \text{jets}$ and $\mu + \text{jets}$ selections: (i) a 6.5% uncertainty on the integrated luminosity, (ii) a 6% per jet uncertainty arising from jet-energy-scale correc-
FIG. 3: Transverse W-boson mass for events containing at least one SLT- or SVT-tagged jet.

FIG. 4: Exclusive jet multiplicity of W-boson candidate events with at least one jet tagged with both the SVT and SLT algorithms. The fourth bin represents the sum of events containing four or more jets.

tions and jet identification, (iii) a 10% per jet uncertainty arising from the HF-tagging algorithms, and (iv) a 10–18% uncertainty on the predicted MC cross sections (depending on sample). The total systematic uncertainty on the \( \bar{t}\bar{t} \) background is 16%, 21% on single top-quark backgrounds, and 22% on the W/Z+jets backgrounds.

No excess is observed in the “doubly-tagged” jet sample. We therefore proceed to set a limit on the rate anomalous HF-quark production in association with a W boson. Because we do not propose a model for such production, we do not base this limit on any specific efficiency or jet spectrum. We quote limits on the number of events beyond SM expectation per exclusive jet bin. The 95% confidence level (C.L.) upper limits for additional event production in each bin are shown in Table I. These limits are calculated using a modified Frequentist (CLs) method [11].

Assuming that anomalous HF production has the same event topology as certain SM process, the above limits can be translated into limits on cross sections. To this end, we consider two benchmark scenarios:

1. “Wbb-like” production in which two b quarks are produced in association with a W boson. In this scenario, additional light quarks or gluons can be produced, and thereby shift the event topology to more than two jets. Jets not within the acceptance of the detector can also cause the event topology to drop to less than two jets. We model this production using efficiencies for SM W/Z +bb production.

2. “Top-like” production in which a heavy particle is produced and decays to a W boson and a b quark. An event can contain two such heavy particles (“tt-like”) or one heavy quark (“single-top-like”), with additional light or heavy quarks and gluons possible for both cases. We model this scenario using the cross-section weighted efficiencies for SM tt and single top-quark production combined.

We calculate a limit on exclusive jet production for each scenario, but first ignore the probability for reconstructing the predicted number of jets, providing a model-independent comparison of processes with specific jet topologies. The remaining efficiency represents the effect of W-boson selection and HF tagging, and limits for a specific model can be extracted by multiplying this value by the efficiency to reconstruct the number of jets found in each exclusive jet bin. These results are shown in Table II. To evaluate an upper cross-section limit on inclusive jet production for each scenario, we reintroduce the efficiency for reconstructing the predicted jets. For inclusive Wbb-like anomalous production, we sum the first two W+jets bins, as the contribution from the remaining bins is negligible. For top-like anomalous production, we sum all W+jets bins, except the \( n = 1 \) bin, where the contribution is again negligible. Table III shows the 95% C.L. event limits for the combinations of jet bins for these two hypotheses, and also the corresponding anomalous HF production cross-section limits. The jet reconstruction efficiency is included in the calculations, and the limits contain the expected efficiencies for the specified SM processes.

In summary, we observe no excess beyond the SM prediction for heavy-flavor quark production in association

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Source & 1 jet & 2 jets & 3 jets & \( \geq 4 \) jets \\
\hline
Data observation & 5 & 1 & 1 & 0 \\
\hline
SM prediction & 5.0±1.2 & 2.0±0.5 & 1.0±0.2 & 0.4±0.06 \\
\hline
95% C.L. Limit (events) & 6.7 & 3.9 & 4.1 & 3.0 \\
\hline
\end{tabular}
\caption{The numbers of observed and predicted W-boson events with at least one jet tagged by both the SLT and SVT algorithms, as a function of exclusive jet multiplicity. Also shown are the 95% C.L. limits in the form of additional events.}
\end{table}
TABLE II: Cross-section upper limits in pb, based on the hypotheses of “Wbb-like” and “top-like” anomalous production of exclusive number of jets. Each value must still be corrected for the efficiency of reconstructing the predicted number of jets.

<table>
<thead>
<tr>
<th>Model</th>
<th>1 jet</th>
<th>2 jets</th>
<th>3 jets</th>
<th>&gt;4 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wbb-like</td>
<td>35.0</td>
<td>9.2</td>
<td>6.0</td>
<td>4.5</td>
</tr>
<tr>
<td>Top-like</td>
<td>12.6</td>
<td>8.0</td>
<td>11.3</td>
<td>15.4</td>
</tr>
</tbody>
</table>

TABLE III: 95% C.L. limits for the number of events summed over the indicated jet bins. Also shown are cross-section limits based on the hypotheses of “Wbb-like” and “top-like” anomalous production for the selected number of jets.

<table>
<thead>
<tr>
<th>Source</th>
<th>1,2 jets</th>
<th>≥2 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data observation</td>
<td>6 ± 0.2</td>
<td>6.6 ± 0.4</td>
</tr>
<tr>
<td>SM prediction</td>
<td>6.9±1.2</td>
<td>3.3±0.5</td>
</tr>
<tr>
<td>95% C.L. Limit (events)</td>
<td>6.6</td>
<td>4.4</td>
</tr>
</tbody>
</table>

with W bosons in 164 pb⁻¹ of data in the e + jets channel and 145 pb⁻¹ in the µ + jets channel. Using a sample of events containing at least one jet tagged with both the SLT and SVT algorithms, we derive 95% C.L. limits on anomalous heavy-flavor production (Table I). Using benchmark SM processes, we also derive anomalous cross-section limits of 26.4 pb in a Wbb-like scenario and 14.9 pb in a top-like scenario. For comparison, the DØ collaboration has recently published a similar study in the form of a search for Wbb production [12]. Based on the two-jet topology, with both jets HF-tagged, that study sets a 95% C.L. upper cross-section limit of 6.6 pb on Wbb production.

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 of Education and Science, Agency for Atomic Energy and RF President Grants Program (Russia), CAPES, CNPq, FAPEPI, FAPESP and FUNDUNESP (Brazil), Departments of Atomic Energy and Science and Technology (India), Colciencias (Colombia), CONACyT (Mexico), KRF (Korea), CONICET and UBACyT (Argentina), The Foundation for Fundamental Research on Matter (The Netherlands), PPARC (United Kingdom), Ministry of Education (Czech Republic), Canada Research Chairs Program, CFI, Natural Sciences and Engineering Research Council and WestGrid Project (Canada), BMFW and DFG (Germany), A.P. Sloan Foundation, Research Corporation, Texas Advanced Research Program, and the Alexander von Humboldt Foundation.

[1] Visitor from University of Zurich, Zurich, Switzerland.

[1] Visitor from University of Zurich, Zurich, Switzerland.