A Measurement of the Ratio of Inclusive Cross Sections

$$\sigma(p\bar{p} \rightarrow Z + b \text{ jet})/\sigma(p\bar{p} \rightarrow Z + j \text{ et})$$ at $$\sqrt{s} = 1.96 \text{ TeV}$$

Inclusive $Z+b$-jet production is expected to be a major background to Higgs production in the $p\bar{p} \rightarrow ZH$ channel, with subsequent Higgs-boson decays into $b\bar{b}$. The parton-level subprocesses expected to contribute to the $Z+b$-jet final state are $bg \rightarrow Zb$ (where $g$ stands for a gluon), and $q\bar{q} \rightarrow Zg$, with $g \rightarrow b\bar{b}$ [1]. The process $bg \rightarrow Zb$, where the initial $b$ is from the sea of the proton parton distribution, is predicted to account for approximately two thirds of the total inclusive cross section $\sigma(p\bar{p} \rightarrow Z + b$ jet) at $\sqrt{s} = 1.96$ TeV. The $b$-quark den-
sity of the proton influences the production rates of single top quarks and the final state $hb$, with $h$ representing a supersymmetric Higgs boson. Consequently, the measurement of $Z + b$ jet production is an important step in constraining the $b$-quark density of protons.

In this Letter, we describe a measurement of the ratio of production cross sections of inclusive $Z + b$ jets to $Z$+jets. The measurement of the ratio benefits from cancellations of many systematic uncertainties, such as the 6.5% uncertainty in the luminosity, and therefore allows a more precise comparison with theory.

We search for $Z$ bosons in association with hadronic jets in about 180 pb$^{-1}$ of data collected at the DØ experiment between August 2002 and September 2003. The DØ detector at the Fermilab Tevatron collider is a general-purpose detector comprising a magnetic central-tracking, preshower, calorimeter, and muon systems [2]. The central-tracking system consists of a silicon microstrip tracker (SMT) and a central fiber tracker, both located within a 2 T superconducting solenoidal magnet. The design was optimized for tracking and vertexing capabilities at pseudorapidities $|\eta| < 3$, where $\eta = -\ln(\tan(\theta/2))$ and $\theta$ is the polar angle with respect to the proton beam direction ($z$). Particle energies are measured in three liquid-argon/uranium calorimeters: A central calorimeter (CC) covers $|\eta| < 1.1$, and two end calorimeters (EC) extend coverage to $|\eta| < 4.2$, each calorimeter housed in a separate cryostat [3]. Central and forward preshower detectors are located just outside of the superconducting coil (in front of the calorimetry), and additional scintillators between the CC and EC cryostats provide sampling of developing showers for $1.1 < |\eta| < 1.4$. The muon detection system is outside the calorimetry and consists of a layer of tracking detectors and scintillation trigger counters before 1.8 T iron toroid magnets, followed by two similar layers after the toroids. The trigger and data acquisition systems are designed to accommodate high luminosities.

The dielectron sample is selected from the data by requiring two clusters of energy in the electromagnetic (EM) layers at the trigger level. In the offline selection, two EM clusters are each required to have transverse momentum $p_T > 15$ GeV/c and $|\eta| < 2.5$. In addition, the shower development in the calorimeter and isolation from hadronic activity must be consistent with that expected of an electron, and at least one of the EM clusters is required to have an associated track to maximize the possibility of having a $Z$ boson in the event. The electron candidates with matching tracks are to required to have a ratio of measured energy in the calorimeter to momentum measured with the tracking system consistent with that expected of an electron. The $Z$ candidates are selected by requiring a dielectron mass ($m_{ee}$) of 80 GeV/c$^2 < m_{ee} < 100$ GeV/c$^2$. The $Z$+jet sample is then selected by requiring the presence of at least one reconstructed hadronic jet with $p_T > 20$ GeV/c and $|\eta| < 2.5$.

Jets are reconstructed from calorimeter clusters using a cone algorithm of cone size $\Delta R = \sqrt{\Delta\eta^2 + (\Delta\phi)^2} = 0.5$ in pseudorapidity and azimuth ($\phi$). Hadronic jets are required to have an associated cluster of tracks (“track jets”). This requirement reduces background from noise in the calorimeter. Track jets are found by applying a cone track clustering algorithm of size $\Delta R = 0.5$ with a seed track of $p_T > 10$ GeV/c, to tracks of $p_T > 0.5$ GeV/c at that close to the primary interaction vertex (whose determination is discussed below). A track jet can consist of two or more tracks.

A “taggable” jet is a calorimeter jet with a matching track jet within $\Delta R < 0.5$. Applying the taggability criterion to 2,661 jets in 2,219 $Z(ee)$ candidate events in the mass $Z$ window yields 1,658 events. Based on side bands to the $Z$ mass window, 121±4 events are estimated to be from background sources. The main background is from multijet production where two jets mimic EM objects, with one of the objects having an overlapping track that passes the track-matching criteria. The taggability per jet is (75 ± 1)% after background subtraction.

The dimuon sample is defined by the detection of at least one muon candidate at the trigger level. In the offline selection, two isolated muons are required to be of opposite charge, and to have $p_T > 15$ GeV/c and $|\eta| < 2$ with trajectories in the muon spectrometer matched to tracks in the central-tracking detector. Muon isolation is based on the transverse component of the muon momentum relative to the combined momenta of muon and the closest calorimeter jet in $(\eta, \phi)$ space, and is $p_{T,rel} > 10$ GeV/c. The $Z$ candidates are selected by requiring a dimuon mass of 65 GeV/c$^2 < m_{\mu\mu} < 115$ GeV/c$^2$. The $Z$ mass window is larger than in the dielectron channel due to worse momentum resolution for high $p_T$ muons. The criteria for reconstructed hadronic jets are the same as in the dielectron channel. A total of 1,406 events remain after the requirement that there be at least one taggable jet. The main background in this channel is from $b\bar{b}$ production, where both b jets contain muons that satisfy the isolation criterion (referred to as $b\bar{b}$ background). The isolation efficiencies of muons from $Z$ and $b\bar{b}$ are expected to be different, since, for the latter, a hadronic jet would be expected to be close to the muon. By performing fits to dimuon mass spectra, where the background contributes to the continuum, samples with different number of isolated muons are analyzed to measure the isolation efficiencies and background rates. From such analyses, we estimate that the background contribution to the final sample with two isolated muons is 17.5 ± 4.1 events.

Figure 1 shows distributions in transverse momentum of taggable jets for both channels (points with error bars), compared to a $Z$+jet Monte Carlo (MC) generated with ALPGEN [4], using PYTHIA [5] for parton showering and hadronization. Also shown is a background estimation based on data obtained from samples that are in the
side band for the dielectron channel, or fail the isolation criterion for the dimuon channel. The background distribution is normalized to the number of background events estimated in the selected sample. The simulated signal is then normalized so that the total agrees with the measurement in Fig. 1. Within the uncertainty of the jet energy scale (JES), indicated by the darker shading about the expectation, the shape of the distribution is well-described by the simulation.

The $b$ quark fragments into a $B$ hadron, which is identified by a displaced secondary vertex that is separated from the primary vertex. The reconstruction of secondary vertices proceeds in two steps. First, the primary interaction vertex is identified, and then additional nearby vertices are reconstructed. In the high luminosity environment of the Fermilab Tevatron collider, there can be more than one interaction per beam crossing, one of which is likely to have triggered the recorded event. The interaction region in DØ has a root mean-square width of $\approx 25$ cm along $z$, with a transverse beam size of $\approx 30$ $\mu$m. It is possible to distinguish the main hard-interaction vertex from any additional soft interactions because the vertices are normally well-separated along $z$. Primary interaction vertices are reconstructed in two passes. In the first pass, all tracks present in an event are used to find seed vertices using an iterative method, where tracks that contribute to a fit to a common vertex with a $\chi^2$/d.o.f. greater than some chosen threshold are removed. The fit is repeated until a stable set of seeds is obtained. The seed vertices are then used in a second pass to fit all tracks within a certain distance-of-closest-approach to any seed. This improves the position resolution on the vertex, since the fit is less affected by poorly reconstructed tracks. The $p_T$ distribution of the associated tracks is then used to select the primary interaction vertex (PV).

A $b$-jet tagging algorithm for secondary vertices (SV) is used to identify heavy-quark jets in the analysis. Tracks that are displaced from the PV in the transverse plane are used as seeds to find secondary vertices. First, a fixed-cone jet algorithm of $\Delta R = 0.5$ is used to cluster the tracks to form track-jets. Tracks are required to have hits in at least two layers of the SMT, $p_T > 0.5$ GeV/$c$, and be within 0.15 cm in the plane transverse to $z$ and 0.40 cm in $z$ relative to the PV. Tracks identified as arising from $K^0_S$ and $\Lambda$ decays or photon conversions are not considered. Any pair of tracks within a track-jet with an impact parameter relative to the hard-interaction vertex (distance of closest approach $= dca$) of a track to a vertex in the plane transverse to the $z$ direction divided by their uncertainties $= \sigma_{dca}$, $dca/\sigma_{dca} > 3$ is used as a seed for secondary vertices. Additional tracks are attached iteratively to the seed vertices if their $\chi^2$-contribution to the vertex fit is consistent with originating from the vertex. A secondary vertex consists of two or more tracks. The momentum vector of the SV is defined as the vector sum of track momenta. Finally, good-quality secondary vertices are selected based on the decay length (distance between PV and SV), collinearity of the vertex momentum with the direction from PV to SV, and vertex-fit $\chi^2$. A jet is considered $b$ tagged when it is taggable and has at least one secondary vertex, with a decay-length transverse to the PV ($L_{xy}$) divided by its uncertainty $L_{xy}/\sigma_{xy} > 7$, associated with it. A secondary vertex is associated to a jet if the opening angle between the direction of the calorimeter-based jet axis and the momentum vector of the SV is $\Delta R < 0.5$.

The $b$-tagging efficiency ($\epsilon_b$) and the light-flavor tagging rate ($\epsilon_L$) of the $b$-tagging algorithm are parametrized as functions of jet $p_T$ and $\eta$. The parametrization of $\epsilon_b$ is derived from a different data sample using events with jets containing muons (muonic jets), which are dominated by $b$ jets, but also have contributions from light quark jets, gluon jets, and charm jets. The $b$-tagging efficiency is extracted from the heavy-flavor component in this muonic jet sample. The light-flavor tagging rate is also derived from data, after compensating for effects of displaced vertices that do not originate from heavy-flavor decay ($K^0_S$, $\Lambda$, and photon conversions). Different types of samples are used to determine $\epsilon_L$ and $\epsilon_b$, and the spreads are taken as systematic uncertainties.

A comparison of inclusive $Z$+jet events, generated with the ALPGEN leading-order matrix element and PYTHIA for showering, with inclusive $Z+b$ events generated with PYTHIA, shows good agreement for jet $p_T$ and $\eta$ distributions. We therefore use the shapes of $p_T$ and

![Graph showing $p_T$ distribution of taggable jets in dielectron and dimuon channels compared to $Z$+jet ALPGEN with PYTHIA showering and full detector simulation (open histogram), and background (multijet for ee channel and $b\bar{b}$ background for $\mu\mu$ channel) derived from data. The error bars on the data points are statistical. The prediction is normalized to the data, as described in the text.](image-url)
\( \eta \) derived from the \( Z+\text{jet} \) data sample to estimate the expected \( b \)-tagging efficiency and the light-flavor tagging ("mistag") rate. The average \( b \)-tagging efficiency and mistag rate per jet, averaged over \( p_T \) and \( \eta \), are found to be \((32.8\pm1.3)\% \) and \((0.25\pm0.02)\% \), respectively, for the dielectron channel. Corresponding values for the dimuon channel are \((33.1\pm1.1)\% \) and \((0.24\pm0.02)\% \). To obtain the event mistag rate, we take into consideration jet multiplicity, and measure the event mistag rate of \(0.28\%\) (0.27\%) for the dielectron (dimuon) channel.

Since \( \epsilon_b \) is derived from events with a muon embedded in a jet, whereas most of the \( b \)-tagged jets do not contain such muons, the difference in \( b \)-tagging efficiencies for hadronic \( b \) jets and muonic \( b \) jets is derived from MC, and the ratio is used to correct \( \epsilon_b \). We cannot at this point derive the charm tagging efficiency \( (\epsilon_c) \) from data, so we rely on PYTHIA MC to compare \( Z \to b\bar{b} \) and \( Z \to c\bar{c} \) samples. We assume that \((\epsilon_c/\epsilon_b)_{\text{MC}} = 0.266 \pm 0.003\). The jet taggability, \( t_L \), is measured using data to be \((75\pm1)\% \), while that for \( b \) jets, \( t_b \), is obtained from MC, and scaled such that \((t_b)_{\text{data}} = (t_L)_{\text{data}} \times (t_b)_{\text{MC}} \). The results \((t_b)_{\text{data}} = (79.2\pm1.3)\% \) for the dielectron channel and \((80.7\pm1.1)\% \) for the dimuon channel. We assume that the taggability of charm jets is same as \( t_b \).

After applying \( b \) tagging, \( 27 \) \( Z(\rightarrow ee)+b\)-jet candidate events are left, with an expected background from the Drell-Yan \( ee \) continuum and multi jet background of \( 4.2\pm1.4 \) events based on the side-band subtraction method. In the dimuon channel, \( 22 \) events are observed with \( 5.0\pm1.1 \) events from \( b\bar{b} \) background.

After subtracting the background contributions, two equations, one before and the other after the requirement of \( b \) tagging, determine the contributions from different flavors in the remaining events:

\[
N_{\text{before b-tag}} = t_b N_b + t_c N_c + t_L N_L
\]

\[
N_{b\rightarrow \text{tagged}} = \epsilon_b t_b N_b + \epsilon_c t_c N_c + \epsilon_L t_L N_L,
\]

where \( N_b, N_c \) and \( N_L \) are the numbers of events with \( b, c \) and light jets, respectively; \( t_i \) are the taggabilities per event for different jet types; and the \( \epsilon_i \) are the corresponding mean event-tagging efficiencies. We assume that the tagging efficiencies per jet, \( \epsilon_b \) and \( \epsilon_c \), are the same as the tagging efficiencies per event. Equations (1) and (2) have three unknowns. We take the theoretical prediction of \( N_c = 1.69 N_b \) [1], to provide a solution of Eqs. (1) and (2) for \( N_b, N_c \) and \( N_L \).

The ratio \( \sigma(p\bar{p} \to Z+b\text{ jet})/\sigma(p\bar{p} \to Z+\text{jet}) = N_b/(N_b+N_c+N_L) \) is found to be \(0.026\pm0.007 \) for the dielectron channel and \(0.020\pm0.005 \) for the dimuon channel, where the errors are purely statistical. The combined ratio, using the statistical weighting of the number of observed \( Z+\text{jet} \) candidates, is \(0.023\pm0.004 \). The shape of the \( p_T \) spectrum for \( b \)-tagged jets and the significance of decay lengths of secondary vertices are compared to the sum of background and \( Z+b \) MC in Fig. 2. The contribution of each component is given by the solution to Eqs. (1) and (2). The distribution of the decay-length significance for secondary vertices shows clear evidence for a heavy-flavor component in the \( b \)-tagged candidate events.

Sources of systematic uncertainty in the ratio include:

i) Jet energy scale. The JES is varied within its uncertainty. The JES for hadronic \( b \) jets is assumed to be the same as that for light-flavored jets, whereas in MC some differences are observed, and this effect is included as part of the JES uncertainty.

ii) Different methods of estimating background. The background is varied by its measured uncertainty and the ratio is recalculated.

iii) Jets that contain a \( b\bar{b} \) or \( c\bar{c} \) pair from gluon splitting. These jets have a higher tagging probability. The expected contribution is taken from theory [1], and the relative increase in \( b(c) \)-tagging efficiency is estimated from MC. This is labeled as \( Z+(Q\bar{Q}) \) in Table I.
iv) Mistag rate for light jets, which depends on the type of jet sample. Using events collected from hadronic jet triggers, the light-jet tagging efficiency is measured to be 0.23%, and for a sample of events with an enhanced EM fraction and small imbalance in overall $p_T$, this is 0.26%. A tagging efficiency of 0.25% per jet (or 0.28% per event) is obtained for the combined data.

v) Uncertainty in tagging efficiency for $b$ and $c$ jets is obtained by varying the efficiency by a ±1 standard deviation, assuming complete correlation in the ratio of extracted cross sections. Also, for $c$ jets, there is additional uncertainty from the $\epsilon_c/\epsilon_b$ ratio obtained from MC. $\bar{t}_L$ is varied, as above, to estimate this effect.

vi) A small difference observed in $t_b/t_L$ for different MC samples of $Z + b$ jet/monojet, and $Z \rightarrow b\bar{b}/Z \rightarrow q\bar{q}$ is taken into account.

vii) Differences in tagging efficiency between hadronic jets and those containing muons. The $b$-tagging efficiency is measured in data using muonic jets. The tagging efficiency for hadronic jets is estimated to be 86% of that of muonic jets, as derived from $Z \rightarrow b\bar{b}$ MC. The same ratio in $Z + b\bar{b}$ MC is measured to be 84%, and the difference of 2% is taken as a systematic uncertainty.

viii) Different $p_T$-dependence in jet reconstruction for light, $b$, and $c$ jets, measured using MC samples, is accounted for as a systematic uncertainty.

ix) Uncertainty from theory for the ratio $\sigma(Z + c$ jet)/$\sigma(Z + b$ jet) = $N_c/N_b$ is estimated as 9.5% [1].

The effects of systematic uncertainties on the combined measurement are listed in Table I. All these uncertainties are assumed to be completely correlated for the two channels, except for that due to background estimation. Folding these uncertainties together, yields a ratio of 0.023 ± 0.004(stat)$^{+0.003}_{-0.002}$(syst). This measurement is in good agreement with the next-to-leading order (NLO) prediction of 0.018 ± 0.004 [6].

In summary, we have presented the first inclusive measurement of $b$-jet production in association with $Z$ bosons at the Fermilab Tevatron collider, which is a background to the standard-model Higgs searches in the $ZH$ production channel. The measurement is in agreement with the NLO calculations and can be used to constrain the $b$-quark density of proton.

We thank the staff at Fermilab and collaborating institutions, and acknowledge support from the Department of Energy and National Science Foundation (USA), Commissariat à l’Énergie 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, FAPERJ, 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), Natural Sciences and Engineering Research Council and WestGrid Project (Canada), BMBF and DFG (Germany), A.P. Sloan Foundation, Research Corporation, Texas Advanced Research Program, and the Alexander von Humboldt Foundation.