The following full text is a preprint version which may differ from the publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/141221

Please be advised that this information was generated on 2017-12-13 and may be subject to change.
Measurement of the ratio of inclusive cross sections 

$$\sigma(pp \to Z + 2b \text{jets})/\sigma(pp \to Z + 2 \text{jets})$$ in $pp$ collisions at $\sqrt{s} = 1.96$ TeV


(D0 Collaboration)

1LAFFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil
2Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
3Universidade Federal do ABC, Santo André, Brazil
4University of Science and Technology of China, Hefei, People’s Republic of China
5University of the Andes, Bogotá, Colombia
Studies of Z boson production in association with a bottom and an antibottom quark provide important tests of the predictions of perturbative quantum chromodynamics (pQCD) [1,2]. A good theoretical description of this process is essential since it forms a major background for a variety of physics processes, including standard model (SM) Higgs boson production in association with a Z boson, \( ZH(H \rightarrow b\bar{b}) \) [3], and searches for supersymmetric partners of the \( b \) quark [4].

The ratio of \( Z + b \) jet to \( Z + \) jet production cross sections, for events with at least one jet, has been previously measured by the CDF [5,6] and D0 [7–9] Collaborations using Run II data. The ATLAS [10] and CMS [11] Collaborations have also studied \( Z + b \) jet production at \( \sqrt{s} = 7 \text{ TeV} \).

This article presents the ratio of \( Z + 2b \) jets to \( Z + 2 \) jets inclusive production cross sections and is an extension of the previous D0 measurements utilizing similar event selections. The measurement of the ratio benefits from the cancellation of many systematic uncertainties, such as the uncertainty in luminosity and those related to lepton and jet identification, allowing a more precise comparison with theory. The remaining systematic uncertainties arise from the differences between \( b \) jets and light jets. In the following, light-quark flavor (\( u, d, s \)) and gluon jets are referred to as “light jets.” The \( Z + 2b \) jet production cross sections have been measured at CMS [12] and ATLAS [13] at \( \sqrt{s} = 7 \text{ TeV} \). The current measurement is based on the complete Run II data sample collected by the D0 experiment at the Fermilab Tevatron \( p\bar{p} \) collider at a center-of-mass energy of 1.96 TeV. This measurement uses data corresponding to an integrated luminosity of 9.7 fb\(^{-1}\) collected by the D0 experiment in Run II of Fermilab Tevatron. The measured integrated ratio of 0.0236 ± 0.0032(stat) ± 0.0035(syst) is in agreement with predictions from next-to-leading-order perturbative QCD and the Monte Carlo event generators PYTHIA and ALPGEN.

DOI: 10.1103/PhysRevD.91.052010

PACS numbers: 12.38.Qk, 13.85.Qk, 14.65.Fy, 14.70.Hp
has a central section (CC) covering pseudorapidities $|\eta_{\text{det}}| < 1.1$, and two end calorimeters (EC) that extend coverage to $|\eta_{\text{det}}| < 4.2$, with all three housed in separate cryostats [18]. An outer muon system, at $|\eta_{\text{det}}| < 2$, consists of a layer of tracking detectors and scintillation counters in front of 1.8 T toroids, followed by two similar layers after the toroids. Luminosity is measured using plastic scintillator arrays located in front of the EC cryostats. The trigger and data acquisition systems are designed to accommodate the high instantaneous luminosities of Run II.

This analysis relies on all components of the D0 detector: tracking systems, the liquid-argon sampling calorimeter, muon system, and the ability to identify secondary vertices [14]. The SMT allows for precise reconstruction of the primary $p\bar{p}$ interaction vertex and secondary vertices [17,19]. It also enables an accurate determination of the impact parameter, defined as the distance of closest approach of a track to the primary interaction vertex in the $x$-$y$ plane. The impact parameter measurements of tracks, along with reconstructed secondary vertices, are important inputs to the $b$-jet tagging algorithm.

Events containing $Z$ bosons decaying to $\mu\mu$ or $ee$ are collected using triggers based on single electrons or muons. For the off-line selection requirements discussed below, the triggers have an efficiency of approximately 100% for $Z \to ee$ and more than 78% for $Z \to \mu\mu$ decays depending on the transverse momentum of the muon. The $Z + 2$ jet sample requires the presence of at least two jets in the event, while the $Z + 2b$ jet sample requires at least two $b$-jet candidates, selected using a $b$-tagging algorithm [20].

An event is selected if it contains a $p\bar{p}$ interaction vertex, reconstructed from at least three tracks, located within 60 cm of the center of the D0 detector along the beam axis. The selected events must also contain a $Z$ boson candidate with a dilepton invariant mass $70 < M_{\ell\ell} < 110$ GeV.

Dielectron ($ee$) events are required to have two electrons of transverse momentum ($p_T$) greater than 15 GeV identified through electromagnetic (EM) showers in the calorimeter. The showers must have more than 97% of their energy deposited in the EM calorimeter, be isolated from other energy depositions, and have transverse and longitudinal energy profiles consistent with that expected for electrons. At least one electron must be identified in the CC with $|\eta_{\text{det}}| < 1.1$, and a second electron either in the CC or the EC, $1.5 < |\eta_{\text{det}}| < 2.5$. Electron candidates in the CC are required to match central tracks or have a pattern of hits consistent with the passage of an electron through the central tracker. Electrons in the ECs are not required to have a track matched to them due to deteriorating tracking coverage for $|\eta_{\text{det}}| > 2$. Due to the lack of track requirement in EC regions we do not apply any opposite sign requirement for the dielectron events.

The dimuon ($\mu\mu$) event selection requires two oppositely charged muons detected in the muon system that are matched to reconstructed tracks in the central tracker with $p_T > 15$ GeV and $|\eta_{\text{det}}| < 2$. These muons must pass a combined tracking and calorimeter isolation requirement discussed in detail in Ref. [3]. Muons originating from cosmic rays are rejected by applying timing criteria using the hits in the scintillation counters and by limiting the measured displacement of the muon track with respect to the $p\bar{p}$ interaction vertex [21].

A total of about 1.2 million $Z$ boson candidate events are retained in the combined $ee$ and $\mu\mu$ channels with the above lepton selection criteria. The $Z + 2$ jet sample is then selected by requiring at least two jets in the event with $p_T^{\text{jet}} > 20$ GeV and $|\eta_{\text{jet}}| < 2.5$. Jets are reconstructed from energy deposits in the calorimeter using an iterative midpoint cone algorithm [22] with a cone of radius $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta y)^2} = 0.5$ where $\phi$ is the azimuthal angle and $y$ is the rapidity. Jet energy is corrected for detector response, the presence of noise and multiple $p\bar{p}$ interactions. We also correct the jet energy for the energy of those particles within the reconstruction cone that is deposited in the calorimeter outside the cone (and vice versa) [23].

To suppress background from top-antitop quark ($t\bar{t}$) production, events are rejected if the missing transverse energy is larger than 60 GeV, reducing the $t\bar{t}$ contamination by a factor of two. These selection criteria retain an inclusive sample of 20,950 $Z + 2$ jet event candidates in the combined $ee$ and $\mu\mu$ channels.

Processes such as diboson ($WW$, $WZ$, $ZZ$) production can contribute to the background when two leptons are reconstructed in the final state. Inclusive diboson production is simulated with the PYTHIA [24] Monte Carlo (MC) event generator. The $Z + jet$, including heavy flavor jets, and $t\bar{t}$ events are modeled by ALPGEN [25], which generates hard subprocesses including higher-order QCD tree-level matrix elements, interfaced with PYTHIA for parton showering and hadronization. The CTEQ6L1 [26] parton distribution functions (PDFs) are used in all simulations. The cross sections of the simulated samples are then scaled to the corresponding higher-order theoretical calculations. For the diboson and $Z + 2$ jet processes, including the $Z + b\bar{b}$ signal process and $Z + c\bar{c}$ production, next-to-leading order (NLO) cross section predictions are taken from MCFM [27]. The $t\bar{t}$ cross section is determined from NLO + NNLL (next-to-next-leading log) calculations [28].

To improve the modeling of the $p_T$ distribution of the $Z$ boson, simulated $Z + 2$ jet events are also reweighted to be consistent with the measured $p_T$ spectrum of $Z$ bosons observed in data [29].

These generated samples are processed through a detailed detector simulation based on GEANT [30]. To model the effects of detector noise and pile-up events, collider data from random beam crossings with the same instantaneous luminosity distribution as for data are superimposed on simulated events. These events are then
reconstructed using the same algorithms as used for data. Scale factors, determined from data using independent samples, are applied to account for differences in reconstruction efficiency between data and simulation. The energies of simulated jets are corrected, based on their flavor, to reproduce the resolution and energy scale observed in data [23].

The background contribution from multijet events, in which jets are misidentified as leptons, is evaluated from data. This is performed using a multijet-enriched sample of events that pass all selection criteria except for some of the lepton quality requirements. In the case of electrons, the multijet sample is obtained by inverting the shower shape requirement and relaxing other electron identification criteria, while for the muon channel, the multijet sample consists of events with muon candidates that fail the isolation requirements. The normalization of the multijet background is determined from a simultaneous fit to the dilepton invariant mass distributions in different jet multiplicity bins.

Figures 1 and 2 show the dilepton invariant mass and leading jet $p_T$ distributions in data compared to the expectations from various processes. The dominant contribution comes from $Z +$ light jet production. The non-$Z +$ jet background fraction in the $ee$ channel is about 15%, and is dominated by multijet production. The muon channel has a higher purity with a background fraction of about 7%.

This analysis employs a two-step procedure to determine the $b$-quark content of jets in the selected events. First, a $b$-tagging algorithm is applied to jets to select a sample of $Z + 2$ jet events that is enriched in heavy flavor jets. After $b$ tagging, the relative light-, $c$-, and $b$-quark content is extracted by fitting templates built from a dedicated background is determined from a simultaneous fit to the dilepton invariant mass distributions in different jet multiplicity bins.

Figures 1 and 2 show the dilepton invariant mass and leading jet $p_T$ distributions in data compared to the expectations from various processes. The dominant contribution comes from $Z +$ light jet production. The non-$Z +$ jet background fraction in the $ee$ channel is about 15%, and is dominated by multijet production. The muon channel has a higher purity with a background fraction of about 7%.

This analysis employs a two-step procedure to determine the $b$-quark content of jets in the selected events. First, a $b$-tagging algorithm is applied to jets to select a sample of $Z + 2$ jet events that is enriched in heavy flavor jets. After $b$ tagging, the relative light-, $c$-, and $b$-quark content is extracted by fitting templates built from a dedicated...
discriminant that provides an optimized separation between the three components.

Jets considered for $b$-jet tagging are subject to a preselection requirement, called taggability, to decouple the intrinsic performance of the $b$-jet tagging algorithm from effects related to the track reconstruction efficiency. For this purpose, the jet is required to have at least two associated tracks with $p_T > 0.5$ GeV, the leading track must have $p_T > 1$ GeV, and each track must have at least one SMT hit. This requirement has a typical efficiency of 90% per jet.

The $b$-jet tagging algorithm is based on a multivariate analysis (MVA) technique [31]. This algorithm, MVA$_{bl}$, discriminates $b$ jets from light flavor jets utilizing the relatively long lifetime of the $b$ hadrons when compared to their lighter counterparts [20]. Events with at least two jets tagged by this algorithm are considered.

The MVA$_{bl}$ discriminant combines various properties of the jet and associated tracks to create a continuous output that tends towards unity for $b$ jets and zero for light jets. Inputs include the number of secondary vertices and the charge track multiplicity, invariant mass of the secondary vertex ($M_{SV}$), decay length and impact parameter of secondary vertices, the multiplicity of charged tracks associated with them, and the jet lifetime probability (JLIP), which is the probability that tracks associated with the jet originate from the $p\bar{p}$ interaction vertex [20]. Events are retained for further analysis if they contain at least two jets with an MVA$_{bl}$ output greater than 0.15. After these requirements, 241 $Z + 2$ jet events are selected with at least two $b$-tagged jets, where only the two highest $p_T$ tagged jets are examined in the analysis and the electron and muon channels are combined. The efficiency for tagging two $b$ jets in data is 33%. In the MC correction, factors are applied to account for differences with data [20]. The background contamination from diboson, multijet, and top production after $b$-tagging, for the electron and muon channels combined are 8%, 2% and 15%, respectively.

To determine the fraction of events with 2$b$ jets, a dedicated discriminant, $D_{MIL}$, is employed [8,32]. It is a combination of the two most discriminating MVA$_{bl}$ inputs, $M_{SV}$ and JLIP: $D_{MIL} = 0.5 \times (M_{SV}/(5 \text{ GeV}) - \ln(\text{JLIP}))/20$.

To measure the fraction of events with different jet flavors in the selected sample, we count the number of events as a function of the $D_{MIL}$ of the two leading jets $N(D_{MIL}, D_{MIL})$ and then perform a two-dimensional binned maximum likelihood fit to that distribution. The data sample with two heavy flavor tagged jets is fitted to templates consisting mainly of 2$b$ jet, 2$c$ jet, and light flavor jet events, as obtained from ALPGEN+PYTHIA simulated samples. We also compared the shapes of the templates from SHERPA simulated samples and found the templates to be consistent for the two models. Before the fit, all non-$Z +$ jet background contributions, estimated from simulated samples after the MVA$_{bl}$ requirement, are subtracted from the data leaving 180 $Z + 2$ jet events in the combined $ee$ and $\mu\mu$ channel. Next, we measure the jet flavor fractions in the dielectron and dimuon samples combined, yielding the $2b$ jet flavor fraction ($f_{bb}$) of 0.64 $\pm$ 0.08(stat) and the $2c$ jet flavor fraction of 0.32 $\pm$ 0.08(stat). Figure 3 shows the one-dimensional projection onto the highest-$p_T$ jet and the second-highest-$p_T$ jet $D_{MIL}$ axis of the two-dimensional fit.

The fraction of 2$b$ jets measured in the heavy flavor enriched sample is combined with the corresponding event acceptances to determine the ratio, $R$, of the cross sections,

$$R = \frac{\sigma(Z + 2b \text{ jets})}{\sigma(Z + 2 \text{ jets})} = \frac{N_{bb} f_{bb}}{N_{incl\text{tag}}} A_{incl} A_{bb},$$

\[ (1) \]

FIG. 3 (color online). The one dimensional projection onto (a) the highest-$p_T$ jet and (b) the second highest-$p_T$ jet $D_{MIL}$ axis of the two-dimensional fit. The distributions of the $b$, $c$, and light jets are normalized by the fractions found from the fit.
where $N_{\text{incl}}$ is the total number of $Z + 2$ jet events before the tagging requirements, $N_{bb}$ is the number of $Z + 2$ jet events used in the $D_{\text{MIL}}$ fit, $f_{bb}$ is the extracted 2b jet fraction, and $\epsilon_{\text{tag}}^{bb}$ is the overall selection efficiency of $D_{\text{MIL}}$ for 2b jets that combines the efficiencies for taggability and MVA$_{bl}$ discriminant. Both $N_{\text{incl}}$ and $N_{bb}$ correspond to the number of events that remain after the contributions from non-$Z + \text{jets}$ processes have been subtracted from the data.

The pseudorapidity acceptance for electrons and muons is different. In order to quote a combined ratio for the two channels, we correct to a common lepton acceptance as follows. The detector acceptances for the inclusive jet sample and 2b jets are determined from MC simulations in the kinematic region that satisfies the $p_T$ and $\eta$ requirements for leptons and jets. For the $A_{bb}$ and $A_{\text{incl}}$ calculations, we apply selections for both the electron and muon channels for the fiducial region for the events with two jets and two leptons defined as

$$p_T^{\text{jet}} > 20 \text{ GeV and } |\eta^{\text{jet}}| < 2.5,$$

$$p_T^{\mu} > 15 \text{ GeV and } |\eta^{\mu}| < 2. \quad (2)$$

The resulting ratio of the two acceptances is measured to be $A_{\text{incl}}/A_{bb} = 1.09 \pm 0.02$ (stat).

Using Eq. (1), we obtain the ratio of the $Z + 2$ jet cross section to the inclusive $Z + 2$-jet cross section in the combined $\mu\mu$ and $ee$ channel to be $0.0236 \pm 0.0032$ (stat).

Several systematic uncertainties cancel when the ratio $\sigma(Z + 2b\text{jets})/\sigma(Z + 2\text{jets})$ is measured. These include uncertainties on the luminosity measurement, lepton trigger efficiency, and lepton and jet reconstruction efficiencies. The remaining uncertainties are estimated separately for the integrated result. The largest systematic uncertainty of 13.7% comes from the uncertainty on the shape of the $D_{\text{MIL}}$ templates used in the fit including that due to MC statistics of the samples used to construct the templates. The shape of the templates may be affected by the choice of the $b$-quark fragmentation function [33], the background estimation, the difference in the shape of the light jet MC template and a template derived from a light jet enriched dijet data sample, and the composition of the charm states used to determine the charm template shape [8]. It also includes uncertainties on production rates of different hadrons and uncertainties on branching fractions. These effects are evaluated by varying the central values by the corresponding uncertainties, one at a time. The entire analysis chain is checked for possible biases using a MC closure test and no significant deviations are observed. The next largest systematic uncertainty of 5.5% is due to the $b$-jet identification efficiency. The uncertainty on $b$-jet energy calibration is 2.6%; it comprises the uncertainties on the jet energy resolution and the jet energy scale. For the integrated ratio measurement, these uncertainties, when summed in quadrature, result in a total systematic uncertainty of 14.9%. For the integrated ratio, we obtain

$$R = 0.0236 \pm 0.0032 \text{(stat)} \pm 0.0035 \text{(syst)}. \quad (3)$$

To check the stability of the result, the ratio is remeasured using a looser(tighter) MVA$_{bl}$ selection with the lower limit on the MVA$_{bl}$ output of $> 0.10 (> 0.225)$. The looser selection provides increased data statistics and the tighter selection yields a 2b enriched sample. The new and default ratios are found to be in agreement within uncertainties of about 4%.

To validate the $t\bar{t}$ background estimation, we reduce the contribution of $t\bar{t}$ events by rejecting events where the scalar sum of all jet $p_T$ values is more than 130 GeV. This selection reduces the $t\bar{t}$ fraction by an additional factor of 2 with a signal efficiency of 80%. The new and default ratios are found to be in agreement within systematic uncertainties.

In Table I, we present the ratio of integrated cross sections, $\sigma(p\bar{p} \rightarrow Z + 2b\text{jet})/\sigma(p\bar{p} \rightarrow Z + 2\text{jet})$, in the fiducial region defined in Eq. (2). The ratio is compared to predictions from NLO QCD calculations and two MC generators, PYTHIA and ALPGEN. The NLO predictions use the MSTW2008 PDF set [34] using MCFM with central values of renormalization and fragmentation scales $\mu_r = \mu_f = M_Z$. Uncertainties are estimated by varying $\mu_r$ and $\mu_f$ together by a factor of 2, and are about 15%. ALPGEN generates multiparton final states using tree-level matrix elements. When interfaced with PYTHIA, it employs an MLM scheme [35] to match matrix element partons with those after showering in PYTHIA, resulting in an improvement over leading-logarithmic accuracy. The measured ratio is in reasonable agreement with MCFM NLO calculations considering the uncertainties on the data and theory.

In summary, we report the measurement at the Tevatron of the ratio of integrated cross sections, $\sigma(p\bar{p} \rightarrow Z + 2b\text{jet})/\sigma(p\bar{p} \rightarrow Z + 2\text{jet})$, for events with $Z \rightarrow \ell\ell$

\begin{table}[h]
\centering
\begin{tabular}{lllll}
\hline
$\sigma(p\bar{p} \rightarrow Z + 2b\text{jet})/\sigma(p\bar{p} \rightarrow Z + 2\text{jet})$ & $\delta_{\text{stat}} \pm \delta_{\text{syst}} $ & $\delta_{\text{tot}}$ & NLO QCD(MSTW) & PYTHIA & ALPGEN \\
\hline
(2.36 ± 0.32 ± 0.35) × 10^{-2} & 0.47 × 10^{-2} & (1.76 ± 0.26) × 10^{-2} & 2.42 × 10^{-2} & 2.21 × 10^{-2} \\
\hline
\end{tabular}
\end{table}

TABLE I. The ratio of integrated cross sections, $\sigma(p\bar{p} \rightarrow Z + 2b\text{jet})/\sigma(p\bar{p} \rightarrow Z + 2\text{jet})$ and total systematic uncertainties ($\delta_{\text{tot}}$) and and total systematic uncertainties ($\delta_{\text{syst}}$). The column $\delta_{\text{sys}}$ shows the total experimental uncertainty obtained by adding $\delta_{\text{stat}}$ and $\delta_{\text{syst}}$ in quadrature. The last three columns show theoretical predictions obtained using NLO QCD with scale uncertainties and two MC event generators, PYTHIA and ALPGEN.

052010-7
in a restricted phase space of leptons with $p_T^j > 15$ GeV, $|\eta^j| < 2.0$ and with two jets limited to $p_T^{jets} > 20$ GeV and $|\eta^{jets}| < 2.5$. Measurements are based on the full data sample collected by the D0 experiment in Run II of the Tevatron, corresponding to an integrated luminosity of 9.7 fb$^{-1}$ at a center-of-mass energy of 1.96 TeV. The measured integrated ratio of 0.0236 ± 0.0032(stat) ± 0.0035(syst) is in agreement with the theoretical predictions within uncertainties.

We thank John Campbell and Doreen Wackeroth for valuable discussions, and the staffs at Fermilab and collaborating institutions, and acknowledge support from the Department of Energy and National Science Foundation (United States of America); Alternative Energies and Atomic Energy Commission and National Institute of Nuclear Energies and Atomic Energy Commission and National Foundation (United States of America); Alternative from the Department of Energy and National Science collaborating institutions, and acknowledge support.

[17] We use a standard right-handed coordinate system. The nominal collision point is the center of the detector with coordinates (0, 0, 0). The direction of the proton beam is the positive $+z$ axis. The $+x$ axis is horizontal, pointing away from the center of the Tevatron ring. The $+y$ axis points vertically upwards. The polar angle, $\theta$, is defined such that $\theta = 0$ is the $+z$ direction. The rapidity is defined as $y = -\ln [(1 + p_z)/(1 - p_z)]$, where $E$ is the energy and $p_z$ is the momentum component along the proton beam direction. Pseudorapidity is defined as $\eta = -\ln (\tan \theta/2)$. $\phi$ is defined as the azimuthal angle in the plane transverse to the proton beam direction. Also, $\eta_{det}$ and $\phi_{det}$ are the pseudorapidity and the azimuthal angle measured with respect to the center of the detector.
[19] The primary $p\bar{p}$ interaction vertex is that found to be the most likely collision point, among possibly several collisions within a specific beam crossing, from which selected objects emanate. The algorithm for defining it can be found in [20].