DØNote 5205 - v2.3 - CONF

Evidence for $Z \to b\bar{b}$ Decays at DØ

The DØ Collaboration
URL: http://www-d0.fnal.gov
Contact Person for This Analysis: Andrew Haas (haas@fnal.gov)

(Dated: July 25, 2006)

A search has been performed at DØ for $Z \to b\bar{b}$ decays using the initial 300 pb^{-1} of data from Run II. Candidate events are selected by triggering on muons from the semileptonic decay of b quarks. Trigger and offline b-tagging requirements remove much of the background from light quarks. The remaining $b\bar{b}$ background is estimated using b-tag rate functions (TRFs) derived from data. An excess of 1168 ± 217 (stat.) ± 150 (sys.) events are observed over the QCD background, an excess of about 4.4σ. The mean and width of the observed peak are in agreement with the Standard Model prediction from $Z \to b\bar{b}$, as simulated in MC. The observed event excess in a trigger-selected data subset is also in agreement with the number of expected events from $Z \to b\bar{b}$, after accounting for trigger and selection efficiencies.

Preliminary Results for Summer 2006 Conferences

I. MOTIVATION

The measurement of $Z \rightarrow b\bar{b}$ is an important ingredient of the Run II physics program at DØ. $Z \rightarrow b\bar{b}$ is an essential tool in the calibration of b-jets, which affects much of the high p_T physics studied at the Tevatron. The current uncertainty on the b-jet energy scale - of the order of 3% - dominates the uncertainty on the mass of the top quark. In addition, $Z \rightarrow b\bar{b}$ can be used to better understand the b-jet energy resolution. $Z \rightarrow b\bar{b}$ also serves as an important benchmark signal for Higgs physics, since it is the closest observable process to the dominant decay of a light Higgs to $b\bar{b}$.

The upgraded tracking system, improved muon triggering, and the possibility of online b-jet tagging make the accumulation of a significant $Z \rightarrow b\bar{b}$ signal possible during Run II at DØ. The main challenge is to sufficiently reduce the QCD backgrounds such that $b\bar{b}$ events from the Z can be observed, without biasing the shape of the signal. To this end, careful analysis techniques are critical and effective triggers must be employed. This paper describes the first search for $Z \rightarrow b\bar{b}$ at DØ in Run II of the Tevatron collider.

II. EVENT SAMPLES

The data used in the analysis come from the "BID" data subset, processed and fixed with the p14 versions of the reconstruction software with PASS2 corrections and has the following properties:

- A "loose" offline reconstructed muon in each event, with $p_T > 4$ GeV/c, matched within ΔR <0.7 to a Run II cone jet [4] of radius 0.5. This requirement enhances the fraction of heavy-flavor events due to the decays of $b \rightarrow \mu$ and $b \rightarrow c \rightarrow \mu$.
- Contains about 90 million events and corresponds to an integrated luminosity of about 300 pb^{-1}.

The following run selection is imposed to ensure data quality:

- Events are required to be in good luminosity blocks;
- Good calorimeter and jet/Missing E_T runs are selected;
- Bad muon runs are removed;
- Runs with bad central tracking system information (SMT and CFT) are excluded.

In addition, MC was generated for $Z \rightarrow b\bar{b}$ signal with Pythia, overlaid with an average of 0.8 minimum bias events, passed through a full GEANT simulation of the detector, and processed with version p14 of the reconstruction software.

The data and MC samples are processed using a custom ROOT ntuple-generating package (higgs_multijet). Jet energy scale (JES) corrections are applied to all jets, including corrections for muons in jets, using version 5.3 of the JES correction software. No specific b-jet energy scale corrections are applied, as no officially approved such correction yet exists.

Monte Carlo events are corrected to account for the b-tagging efficiency and jet measurement effects observed in data. Jet energies are smeared to accurately reflect the jet energy resolution measured in data, and event weighting is applied in order to account for the fact that jet reconstruction/identification and b-tagging are less efficient in data than MC.

A. Triggers

$Z \rightarrow b\bar{b}$ events are characterized by the presence of two b-tagged jets. Light quark rejection is needed at the trigger level, prior to any offline event selection, in order to achieve an acceptable trigger rate at instantaneous luminosities in excess of 50 x 10^{30}cm^{-2}s^{-1}. Without this rejection the data would have to be undesirably prescaled, and candidate events would be thrown away.

Ideally one would trigger on di-jet events with displaced vertices at Level 2; such a capability is however only since recently provided by the Level 2 Silicon Track Trigger (STT). A large reduction in rate must already be achieved at Level 1. Candidate events have been selected at the first trigger level using the semileptonic decay of the b-jets. Triggering on muons from one or both of the b-jets from the Z limits the signal efficiency however, because of the small $b \rightarrow \mu$ branching fraction[5].

III. EVENT SELECTION

There are few kinematic handles with which to discriminate between the $Z \rightarrow b\bar{b}$ signal and the QCD $b\bar{b}$ background. The two most powerful variables, besides the invariant di-jet mass, are the number of jets in the event, n_{jet}, and the angular separation of the two leading b-jets in the plane perpendicular to the beam, $\Delta\varphi$. Both of these variables are sensitive to the prediction that QCD background, due to the color connection between initial and final state in the QCD processes, should have more gluon radiation. With this in mind, candidate events are selected offline using the following prescriptions:

- The event must contain two, and only two, good quality jets.
- The two jets must both have $\mid \eta \mid < 2.5$ and *JES corrected* $p_T > 20$ GeV.
- For b-tagging, the two jets must both be taggable.
- The primary vertex of the event must have at least 4 tracks attached and be located within ± 35 cm of the detector center in the z-direction.
- The two jets must both be "loose" SVT-tagged (using the standard p14 definitions).
- The two b-jets must be roughly back-to-back in the plane perpendicular to the beam-line, i.e. $\Delta\varphi > 2.5$ radians.

The b- / light-jet fraction is about 10% after a single b-tag requirement. After requiring a double offline b-tag in the data, the light-quark QCD background component is reduced to about 10% of the sample. However, a large heavy-flavor component remains which still swamps the signal; S:(S+B) is of the order of 1:30.

IV. BACKGROUND SUBTRACTION

Understanding the shape of the background to the $Z \rightarrow b\bar{b}$ signal in the double b-tagged data sample is central to this analysis. The background is composed of heavy-flavor di-jet production and mis-tagged gluon/light-quark jet production, none of which can be accurately simulated using current techniques, particularly in the quantities needed for the statistical accuracy required. Thus the background is derived from data, using either single-tagged and/or untagged events.

We use a *tag rate function (TRF)* to estimate the shape of the double b-tagged background in the data. We measure the *tag ratio* of double to single b-tags as a function of jet p_T and jet η to form the TRF. The TRFs are applied, as an event weight, to the single tagged events. A similar approach was adopted for the DØ Run II $hb\bar{b}$ analysis [2]. Consider those events in which the first-leading-p_T jet is b-tagged. For these events, the second-leading-p_T jet is categorized according to its location in one of three different eta regions of the detector:
$\mid \eta \mid < 1.1$, $1.1 < \mid \eta \mid < 1.5$ and $1.5 < \mid \eta \mid < 2.5$. For each η region, the TRF is then parameterized as a function of the p_T of the second-leading-p_T jet. This generates a TRF *per jet*.

To estimate the background, each event passing cuts with the leading-p_T jet b-tagged is weighted by the TRF corresponding to the second-leading-p_T jet. Figure 1 illustrates the jet-based TRF as a function of the second leading jet p_T, for jets in the three different η-bins.

Because the TRF is applied to the single-tagged data sample, which is less rich in heavy-flavor jets than the data sample to which it is compared (the double b-tagged data), any differences in either the jet-energy-scale of heavy-flavor jets as opposed to gluon and light-quark jets or their natural invariant mass spectrum will result in a shift of the invariant mass distribution expected for the background, as compared to the data. This shift is first observed and measured using an *untagged* data sample. A TRF is derived for the untagged data and applied back to the untagged data, and then compared to the actual single-tagged data. The shift which is derived is called the "0→1" correction, which is then subtracted from the expected background in the double-tagged data sample. This correction relies on the fact that each successive b-tag that is required increases the fraction of heavy-flavor by the same amount. This is verified by comparing the data with MC $b\bar{b}$ events.

This method also makes an additional correction for the effects of $Z \rightarrow b\bar{b}$ events which are present in both the untagged and single-tagged data samples from which the 0→1 correction and the single-tag TRF are derived, respectively. A signal peak is first measured in double-tagged data, including the 0→1 correction. The number of events which would exist from this signal peak in the untagged and single-tagged data samples is then extrapolated using a $Z \rightarrow b\bar{b}$ MC sample with no, one, and two b-tags required. The signal peak, *measured in data*, is then scaled up

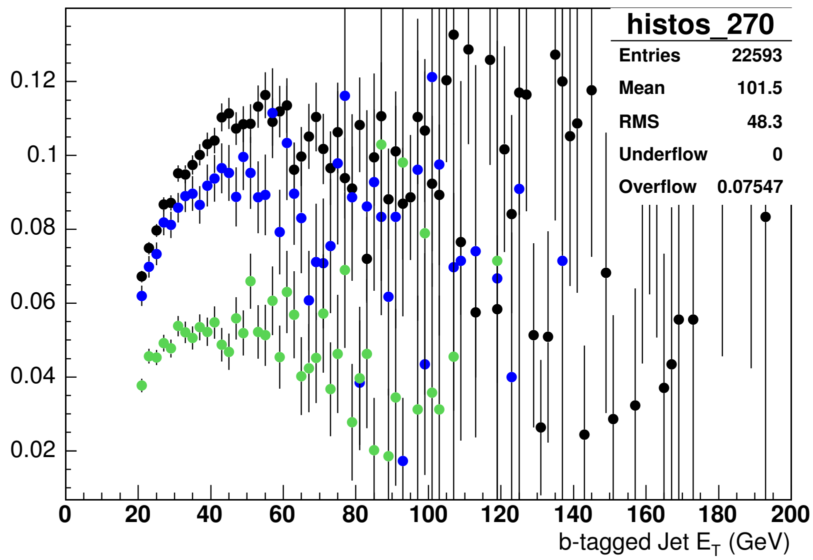

FIG. 1: The TRF's derived on the single-tagged data sample, which are used to estimate the double-tagged background. Each TRF is a function of corrected jet p_T in one of three η-bins: $\eta < 1.1$ (black), $1.1 < \eta < 1.5$ (blue), and $1.5 < \eta < 2.5$ (green).

by these factors and subtracted from the untagged and single-tagged samples. Then the TRF and the 0→1 correction are re-derived, the expected background in the double-tagged data is re-calculated, and a new signal peak is observed. This correction process is then repeated, using the new signal peak to estimate the $Z \to b\bar{b}$ events in the untagged and single-tagged data samples, a total of three times, after which the correction and signal peak are stable. Effects of contributions from W→cs production and decay have been investigated and determined to be small.

A. The $0 \to 1$ tag Shift

The shift in the invariant mass spectrum of the di-jet system caused by applying a single b-tag is first measured. Later this will be subtracted from the estimated background of the double b-tagged data. A TRF is derived on the untagged sample and re-applied to the same untagged data sample, to predict the background to the single-tagged data sample, as shown in Figure 2. A comparison is also shown in this figure to the $b\bar{b}$ MC di-jet invariant mass spectrum (which was normalized using double-tagged data). The $b\bar{b}$ events make up about 10% of the single-tagged data sample, as opposed to about 2% in the untagged data sample. The small contribution of $Z \to b\bar{b}$ expected in the single-tagged data is also shown. Most of the events in the single-tagged data sample contain only gluon/light-quark jets.

The difference between the single-tagged data and the expected background from the TRF is shown in Figure 3. A comparison of the background-subtracted single-tagged data to the MC $Z \to b\bar{b}$ invariant mass distribution shows that it is not the result of a signal peak in the single-tagged data, but rather an overall shift in the invariant mass distribution. As mentioned above, this shift is understood to come from either a difference between the gluon/light-quark and the b-quark jet energy scale factors or a difference in the ratio of true cross-section of heavy-flavor jets to light-jets as a function of di-jet invariant mass.

B. The 0- and 1-tag Z Peak Correction

The untagged and single-tagged data samples contain signal events, which will now be corrected for. The signal peak observed in double-tagged data, after subtracting the estimated background using the TRF method, is scaled up by a factor of 6.5, which is the ratio of single-tagged to double-tagged events in the $Z \to b\bar{b}$ MC sample. It is important to note that the MC is only used for an overall normalization (the b-tagging efficiency), and does not affect

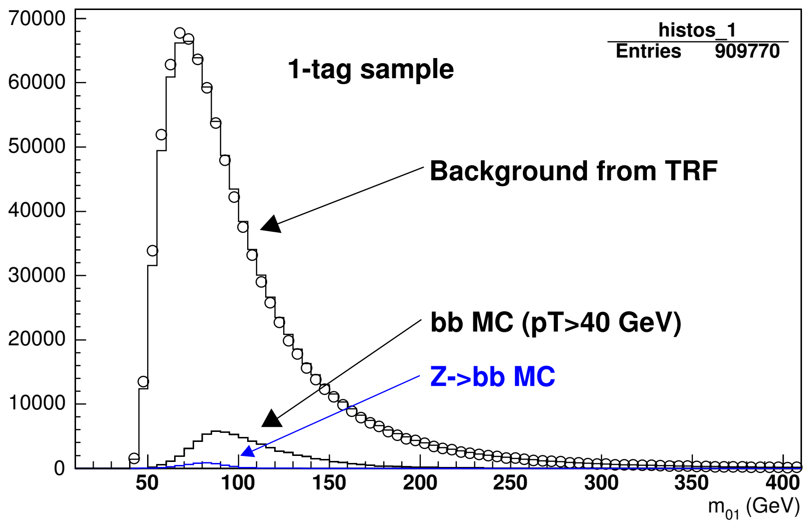

FIG. 2: Comparison between the single-tagged data and the background expected using the TRF method. Comparisons are also shown to the $b\bar{b}$ MC di-jet invariant mass spectrum (which was normalized using double-tagged data) and to the $Z \to b\bar{b}$ MC di-jet invariant mass spectrum. The rest of the events are thought to be gluon/light-quark jet events.

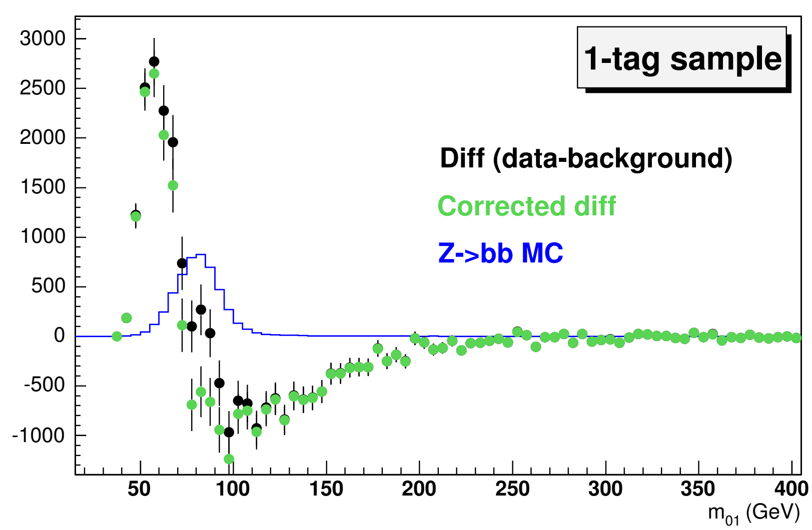

FIG. 3: Difference between the single-tagged data and the background expected using the TRF method, a measure of the 0→1 tag shift, which will be subtracted from the expected double b-tagged data (after proper normalization). The difference is also shown after correcting for the expected $Z \to b\bar{b}$ events in the untagged and single-tagged data samples, using the methods described below in Section IV B.

the shape of the signal peak. This scaled signal peak is then divided (bin by bin) by the single-tag data, to determine the estimated fraction of $Z \to b\bar{b}$ events in each bin of the single-tag data.

The expected fraction of $Z \to b\bar{b}$ events in each invariant mass bin is then subtracted from the signal-tagged data (the events are weighted by 1-f, where f is the fraction of $Z \to b\bar{b}$ expected). Then the TRFs are re-derived and re-applied to this corrected, weighted, single-tagged data. The 0→1 tag correction is also re-derived, using the same principles to estimate the fraction of $Z \to b\bar{b}$ in the untagged data sample. The effect from $Z \to b\bar{b}$ events in the untagged sample is relatively small, as expected, so the modification to the 0→1 tag correction is slight, as seen in Figure 3.

Since the size and shape of the original Z peak in data changes slightly after the corrections to the untagged and single-tagged data, the corrections themselves can be re-derived using the new Z peak observed. A solution is stable after only a few iterations, which results in the final Z peak and set of corrections.

C. Applying to Data

The final $Z \to b\bar{b}$ peak derived from data, after all corrections, is considered to be the difference between the background-subtracted double-tagged data and the corrected 0→1 tag shift. This invariant mass distribution difference is shown in Figure 4, and compared to the shape of the $Z \to b\bar{b}$ distribution in MC.

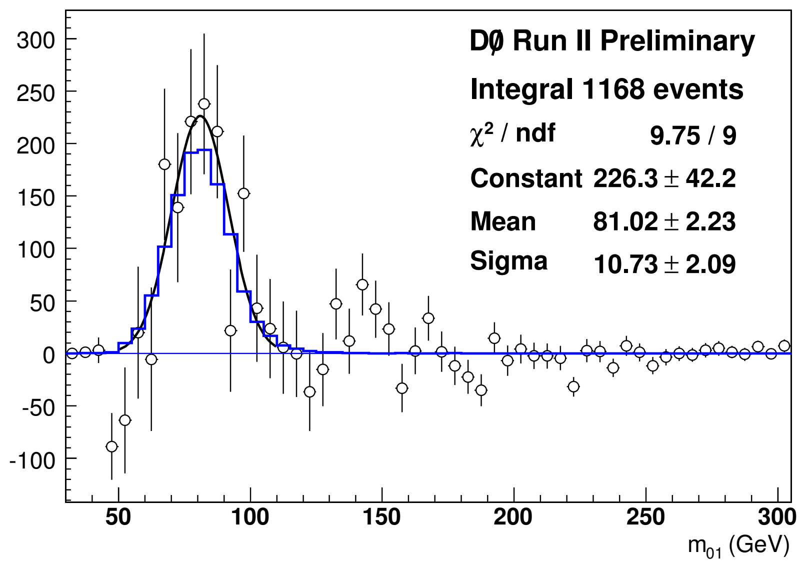

FIG. 4: The final $Z \to b\bar{b}$ peak derived from data, after all corrections (points with error bars, black), compared to the shape of the $Z \to b\bar{b}$ distribution in MC (histogram, blue).

The peak in data is fit to a Gaussian distribution, which describes the shape of the $Z \to b\bar{b}$ peak well in MC. Both the mean (81.0±2.2 GeV) and width (10.7±2.1 GeV) of the distribution observed for data are comparable to those derived from $Z \to b\bar{b}$ MC: 83.3 GeV and 13.0 GeV, respectively.

In the peak, 1168±217 $Z \to b\bar{b}$ events are observed, where the error is taken from the uncertainty on the height of the Gaussian fit. The total number of events expected from $Z \to b\bar{b}$ MC can not be estimated very accurately, because the data were not selected by a given trigger. However, in a trigger-selected subset of the data, a peak containing 651±174 $Z \to b\bar{b}$ events is observed, compared to an expectation of 754±151 events, indicating that the rate observed is compatible with expectations from SM production.

V. SYSTEMATICS

Two types of systematic uncertainties are studied: those influencing the selection efficiency for signal, and those which modify the shape, size, or location of the signal peak which is derived in data.

For the signal efficiency, the dominant source of uncertainty is the trigger efficiency, estimated to be 80%, with an uncertainty of ±20%. Changing the b-tagging efficiency by $\pm 1\sigma$ modified the signal efficiency, measured in MC, by ±12%. A 7% uncertainty is assigned due to the requirement that there not be a third jet above threshold (8 GeV) in the event, measured by moving the third jet threshold to 10 GeV and re-measuring the signal efficiency. As the relative data/MC jet energy scale is varied by $\pm 1\sigma$, the signal efficiency changes by +8% and -6%, respectively, due to the requirement of having 2 jets with p_T>20 GeV. Of course, the signal peak position is linearly sensitive to the data JES. The uncertainty on the peak position due to standard JES uncertainty is about ±4%. Other uncertainties due to primary vertex requirements and $\Delta\phi$ selection are negligible. The total signal efficiency uncertainty is thus about 25%.

The change in the signal peak was observed when varying the $Z \rightarrow b\bar{b}$ TRF correction factor, and the 0→1 correction factor, by ±50% each. The signal peak mean and width shift by a negligible amount when adjusting the amount of $Z \rightarrow b\bar{b}$ in the TRFs. The signal size varies by ± 113 events, or 10%. The signal peak mean shifts by ±4 GeV when modifying the amount of 0→1 correction applied, or about 5%. The signal peak width is not changed much. And the signal size varies by +10,-0%, or about 100 events in the positive direction only. The total signal peak size systematic uncertainty is thus about ±13%, the peak mean about ±4%, and the width is dominated by statistical uncertainty.

VI. CONCLUSIONS

We have extracted a $Z \rightarrow b\bar{b}$ signal in approximately 300 pb^{-1} of data, with an excess of 1168 ± 217 (stat.) ±150 (sys.) events observed over the QCD background. Adding statistical and systematic uncertainties in quadrature gives 1168 ± 264 events, an excess of about 4.4σ. The position and width of the observed mass peak are in agreement with MC expectations. The expected number of events, after trigger selections, is also in good agreement with the excess observed in a trigger-selected data subset.

Thanks to new, dedicated, $Z \rightarrow b\bar{b}$ triggers and additional integrated luminosity, we have already collected a several times larger data sample which is now being analyzed. The larger statistics from this sample is expected to not only provide a larger $Z \rightarrow b\bar{b}$ peak but also enable a better background determination, thus enabling a precision measurement.

Acknowledgments

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA), CEA and CNRS/IN2P3 (France), FASI, Rosatom and RFBR (Russia), CAPES, CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil), DAE and DST (India), Colciencias (Colombia), CONACyT (Mexico), KRF (Korea), CONICET and UBACyT (Argentina), FOM (The Netherlands), PPARC (United Kingdom), MSMT (Czech Republic), CRC Program, CFI, NSERC and WestGrid Project (Canada), BMBF and DFG (Germany), SFI (Ireland), A.P. Sloan Foundation, Research Corporation, Texas Advanced Research Program, Alexander von Humboldt Foundation, and the Marie Curie Fellowships.

[1] T. Dorigo,*Observation of Z Decays to b Quark Pairs at the Tevatron Collider*, hep-ex/9806022.

[2] A. Haas *et.al.*,*DØ Search for Neutral Higgs Bosons at High tanβ in Multi-jet Events Using p14 Data*, DØNote 4671, hep-ex/0504018.

[3] A. Jenkins and A. Goussiou, *An Investigation of b-jet Energy Resolution in* $Z^0H^0 \rightarrow eeb\bar{b}$ *and* $Z^0 \rightarrow b\bar{b}$, DØNote 4136.

[4] G. C. Blazey *et al.*, in *Proceedings of the Workshop: "QCD and Weak Boson Physics in Run II,"* edited by U. Baur, R. K. Ellis, and D. Zeppenfeld, (Fermilab, Batavia, IL, 2000) p. 47; see Sec. 3.5 for details.

[5] B.R.(b $\rightarrow \mu$) = 0.11. Muons from the cascade decays $b \rightarrow c \rightarrow \mu$ and $b \rightarrow \tau \rightarrow \mu$ are significant, and the total probability for a b-jet to have a muon is nearly 20%. For $Z \rightarrow b\bar{b}$ events in which *at least one* b-jet has a muon which passes muon cuts, the maximum signal efficiency hence is $1-(1-0.2)$x$(1-0.2) = 36\%$. For events with two muons, this number reduces to 4%. However, not all of these muons will have sufficient p_T to pass the trigger requirements.