



Search for $ZH \rightarrow \ell^+ \ell^- b\bar{b}$ production in 6.2 fb^{-1} of $p\bar{p}$ collisions

The DØ Collaboration
URL <http://www-d0.fnal.gov>

We present a search for a standard model (SM) Higgs boson produced in association with a Z boson in 6.2 fb^{-1} of $p\bar{p}$ collisions, collected with the DØ detector at the Fermilab Tevatron at $\sqrt{s} = 1.96 \text{ TeV}$. Selected events contain one reconstructed $Z \rightarrow e^+e^-$ or $Z \rightarrow \mu^+\mu^-$ candidate and at least two jets, including at least one b -tagged jet. The data are consistent with the background expected from other SM processes. Upper limits at 95% C.L. on the ZH production cross section times branching ratio are set for Higgs boson masses $100 < M_H < 150 \text{ GeV}$. The observed (expected) limit for $M_H = 115 \text{ GeV}$ is a factor of 8.0 (5.7) larger than the SM prediction.

I. INTRODUCTION

In the standard model (SM), the spontaneous breakdown of the electroweak gauge symmetry generates masses for the W and Z bosons and produces a residual massive particle, the Higgs boson, which has so far eluded detection. The discovery of the Higgs boson would be a remarkable addition to the list of experimentally confirmed SM predictions. For Higgs boson masses $M_H \lesssim 135 \text{ GeV}$, the primary Higgs boson decay in the SM is to $b\bar{b}$, which has copious background at the Tevatron collider. Consequently, sensitivity to a low-mass Higgs boson is predominantly from its production in association with a W or Z boson.

We present a search for $ZH \rightarrow \ell^+ \ell^- b\bar{b}$, where ℓ is either a muon or an electron. The search for $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ is treated elsewhere [1]. The data for this analysis were collected at the Fermilab Tevatron Collider by the DØ detector [2] from April 2001 to February 2006 (Run IIa), and from June 2006 to March 2010 (Run IIb). Between Run IIa and Run IIb a new layer of the silicon microstrip tracker was installed, and the trigger system was upgraded. The analyzed events were acquired predominantly with triggers that provide real-time identification of electron and muon candidates. However, events satisfying any trigger requirement are considered.

This note documents a preliminary update to a previous search in the same final states [3] that used 4.2 fb^{-1} , of which 1 fb^{-1} was collected during Run IIa, with the remaining data collected during Run IIb. In addition to the $\mu\mu$ and ee selections (to be discussed in Section II), the previous analysis also used additional selections: $\mu\mu_{\text{trk}}$, in which the second muon was identified as an isolated track not observed in the muon detectors; and ee_{ICR} , in which the second electron was identified in the inter-cryostat region ($1.1 < |\eta_{\text{det}}| < 1.5$, where η_{det} is the pseudorapidity measured with respect to the center of the detector).

In this update we discuss the reanalysis of the $\mu\mu$ and ee selections of the Run IIb data considered in [3], including an additional 2 fb^{-1} of data. Thus, the integrated luminosity of the Run IIb data set to 5.2 fb^{-1} . In the reanalyzed samples, we make use of a new multivariate b -tagging algorithm that offers improved discrimination against light-flavor jets.

The 1 fb^{-1} Run IIa data, as well as the $\mu\mu_{\text{trk}}$ and ee_{ICR} selections, are not reanalyzed, but are combined with this analysis to obtain the final results. This brings the total integrated luminosity for this result to 6.2 fb^{-1} for the ee and $\mu\mu$ channels, and 4.2 fb^{-1} for the ee_{ICR} and $\mu\mu_{\text{trk}}$ channels.

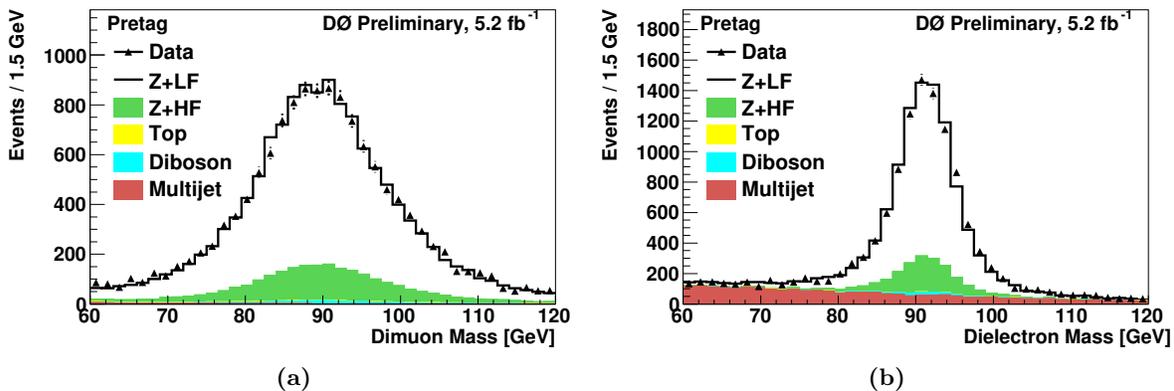


FIG. 1: The dilepton mass spectra in the (a) $\mu\mu$, and (b) ee channels. Distributions are shown in the “pretag” control sample, in which all selection requirements except b -tagging are required.

II. EVENT SELECTION

The selection of signal-like events requires a primary $p\bar{p}$ interaction vertex (PV) that has at least three associated tracks, and is located within ± 60 cm of the center of the detector along the direction of the beam. Selected events must also contain a $Z \rightarrow \ell^+\ell^-$ candidate with a dilepton invariant mass $60 < m_{\ell\ell} < 150$ GeV.

The dimuon ($\mu\mu$) selection requires at least two muons identified in the outer muon system, matched to central tracks with transverse momenta $p_T > 10$ GeV. Combined tracking and calorimeter isolation requirements are applied to the muon pair, such that one muon does not have to be isolated if the other muon is sufficiently well isolated. Each muon track must satisfy $|\eta_{\text{det}}| < 2$. At least one muon must have $|\eta_{\text{det}}| < 1.5$ and $p_T > 15$ GeV. The distance of closest approach of each track to the PV in the plane transverse to the beam direction, d_{PV} , must be less than 0.04 cm for tracks with at least one hit in the silicon microstrip tracker (SMT). A track without any SMT hits must have $d_{PV} < 0.2$ cm, and its p_T is corrected through a constraint to the position of the PV. To reduce contamination from cosmic rays, the muon tracks must not be back-to-back in η and ϕ . The two muon tracks must also have opposite charge.

The dielectron (ee) selection requires at least two electrons with $p_T > 15$ GeV that pass selection requirements based on the shapes of the electromagnetic showers in the calorimeter and separation of the showers from other depositions of energy. At least one electron must be identified in the central calorimeter (CC, $|\eta_{\text{det}}| < 1.1$), and a second electron either in the CC or end calorimeter (EC, $1.5 < |\eta_{\text{det}}| < 2.5$). The CC electrons must match central tracks or a pattern of hits consistent with a charged particle.

Jets are reconstructed in the calorimeter using the iterative midpoint cone algorithm [4] with a cone of radius 0.5 in the plane of rapidity and azimuth. The energy scale of jets is corrected for detector response, the presence of noise and multiple interactions, and energy deposited outside of the reconstructed jet cone. At least two jets with $|\eta_{\text{det}}| < 2.5$ are required, with the leading jet of $p_T > 20$ GeV and additional jets with $p_T > 15$ GeV. For dielectron events, all jets are required to be isolated from each electron by $\Delta\mathcal{R} = \sqrt{\Delta\eta^2 + \Delta\phi^2} > 0.5$. To reduce the impact from multiple interactions at high instantaneous luminosities, jets must contain at least two tracks matched to the PV. The dimuon and dielectron mass spectra, after requiring two leptons and two jets are shown in Fig. 1. The invariant mass of the dijet system (constructed from the two jets with the highest p_T) in the combination of the ee and $\mu\mu$ samples is shown in Fig 2.

To distinguish the decay $H \rightarrow b\bar{b}$ from background processes involving light quarks and gluons, jets are identified as likely to contain b -quarks (b -tagged) if they pass “loose” or “tight” requirements on the output of a boosted decision tree trained to separate b -jets from light jets. For $|\eta| < 1.1$ and $p_T \approx 50$ GeV, the b -tagging efficiency for b -jets and the misidentification rate of light jets are, respectively, 72% and 6.7% for loose b -tags, and 47% and 0.4% for tight b -tags. Events with at least one tight and one loose b -tag are classified as double-tagged (DT). Events not in the DT sample that contain a single tight b -tag are classified as single-tagged (ST). The dijet $H \rightarrow b\bar{b}$ candidate is composed of the two highest p_T tagged jets in DT events, and the tagged jet plus the highest p_T non-tagged jet in ST events.

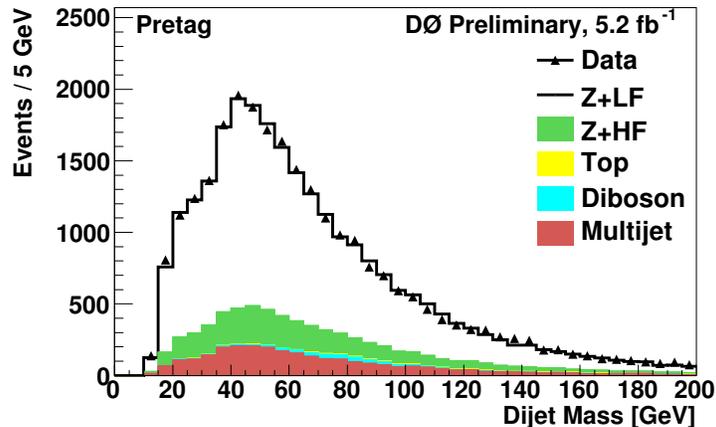


FIG. 2: Distribution of the dijet invariant mass in the combined ee and $\mu\mu$ sample.

III. BACKGROUND ESTIMATION

The dominant background process is the production of a Z boson in association with jets, with the Z decaying to dileptons (Z +jets). The light-flavor component (Z +LF) includes jets from only light quarks (uds) or gluons. The heavy-flavor component (Z +HF) includes non-resonant $Z + b\bar{b}$ which has the same final state as the signal, and non-resonant $Z + c\bar{c}$ production. The remaining backgrounds are from $t\bar{t}$ and diboson production, and from multijet events in which jets are misidentified as leptons. We simulate $ZH \rightarrow \ell^+ \ell^- b\bar{b}$ and inclusive diboson production with PYTHIA [5] and Z +jets and $t\bar{t} \rightarrow \ell^+ \nu b \ell^- \bar{\nu} \bar{b}$ processes with ALPGEN [6]. All simulated samples are generated using the CTEQ6L1 [7] leading-order parton distribution functions (PDFs).

The events generated with ALPGEN use PYTHIA for parton showering and hadronization. Because this procedure can generate additional jets, we use a matching procedure to avoid double counting partons produced by ALPGEN and those subsequently added by the showering in PYTHIA [6]. All samples are processed using a detector simulation program based on GEANT3 [8], and the same offline reconstruction algorithms used to process the data. Events from randomly chosen beam crossings are overlaid on the generated events to model the effect of multiple $p\bar{p}$ interactions and detector noise.

The cross section and branching ratio for signal are taken from Refs. [9, 10]. For the $t\bar{t}$ and diboson processes, the cross sections are taken from MCFM [11], which are calculated at next-to-leading order (NLO). The inclusive Z -boson cross section is scaled to next-to-NLO [12]. Additional NLO heavy-flavor corrections, calculated from MCFM, are applied to $Z + b\bar{b}$ and $Z + c\bar{c}$. To improve the modeling of the p_T distribution of the Z boson, the simulated Z boson events are reweighted to be consistent with the observed dilepton mass spectrum in data before requiring any b -tags [13].

The energies of simulated jets are modified to reproduce the resolution observed in data. Scale factors are applied to account for differences in reconstruction efficiency between the data and simulation. Additional corrections are applied to improve agreement between data and background simulation, using a “pretag” control sample with negligible signal contribution that is obtained by applying all selection requirements except b -tagging. The $\mu\mu$ distributions are corrected for trigger efficiencies. For the ee channel, no correction is applied as the combination of all triggers is nearly 100% efficient. To improve upon the ALPGEN modeling of Z +jets, motivated by a comparison with the SHERPA generator [14], events are reweighted so that the pseudorapidities of the two jets with the highest p_T , and the $\Delta\mathcal{R}$ between them, reproduce those distributions measured in the pretag data. To model the tagged samples, simulated events are weighted by their probability to satisfy the ST or DT criteria as measured in data.

The multijet backgrounds are estimated from control samples in the data. For the $\mu\mu$ channel, the multijet control sample consists of events that fail the muon isolation requirements but otherwise pass the event selection. For the ee channel, the electrons must fail isolation and shower shape requirements.

The normalizations of the simulated and the multijet backgrounds are adjusted by scale factors determined from a fit to the $m_{\ell\ell}$ distributions in the inclusive and pretag data. This improves the accuracy of the background model and reduces the impact of systematic uncertainties that affect pretag event yields (e.g., uncertainties on luminosity and lepton identification). The region $40 < m_{\ell\ell} < 60$ GeV, where the multijet contribution is most prominent, is included in the fit to normalize the multijet control sample to the actual multijet contribution. The inclusive control sample constrains the lepton trigger and identification efficiencies, while the pretag control sample, which includes

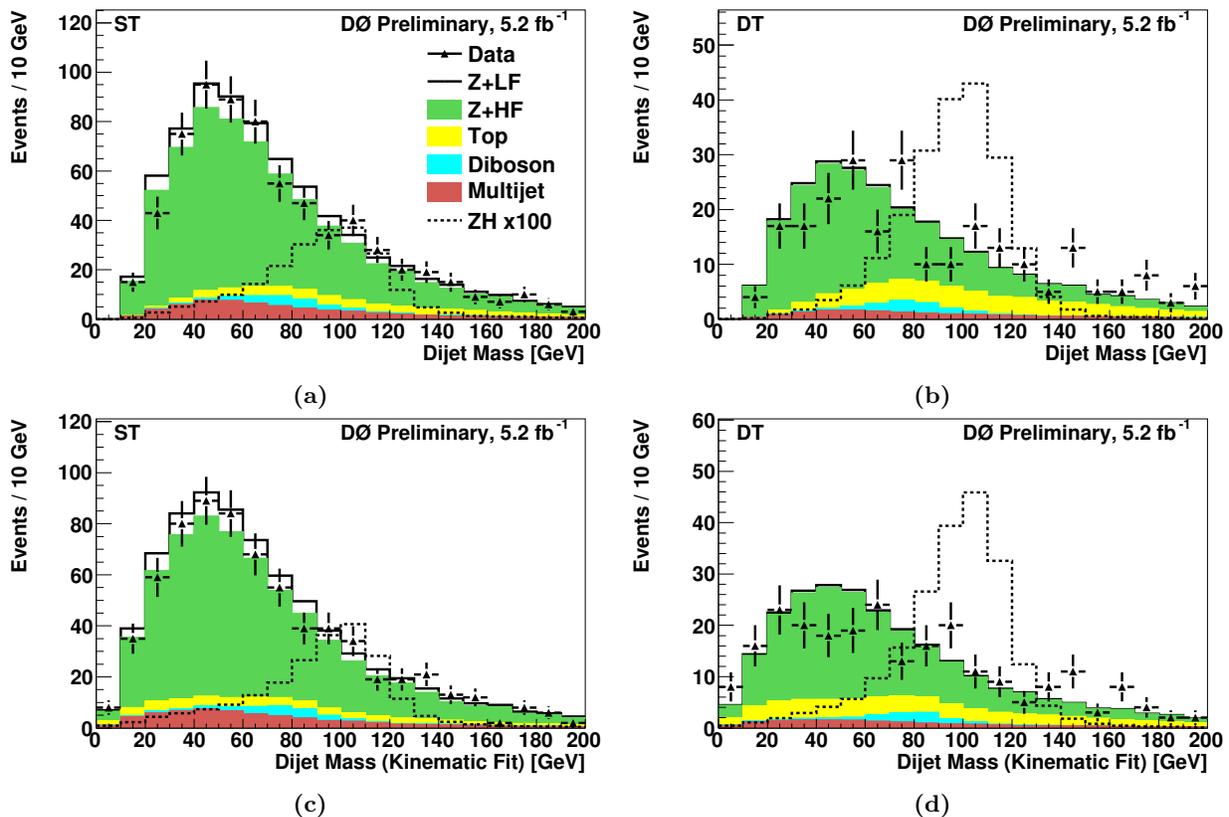


FIG. 3: Dijet invariant mass distributions before the kinematic fit in (a) ST events, and (b) DT events; and after the kinematic fit in (c) ST events and (d) DT events, combined for both lepton channels.

jet requirements, is used to correct the Z +jets cross section by a common scale factor k_{Z+jets} . The total event yields after applying all corrections and normalization factors are shown in Table I.

	Data	Total Background	Multijet	Z+LF	Z+HF	Other	ZH
inclusive	846550	841099	129559	690838	19244	1458	9.62
pretag	26749	26468	2815	19355	3797	500	7.34
ST	638	694 ± 151	32.3	73.8	529	59.3	2.13 ± 0.35
DT	220	227 ± 52	11.2	4.47	165	46.3	2.01 ± 0.41

TABLE I: Expected and observed event yields for all lepton channels combined after requiring two leptons (inclusive), after also requiring two jets (pretag), and after requiring at least one (ST) or two (DT) b -tags. The total statistical and systematic uncertainties are indicated for the Total Background and ZH columns of the ST and DT samples. No systematic uncertainties are assessed for the inclusive and pretag control samples. The “Other” column includes diboson and $t\bar{t}$ event yields. The ZH sample yields are for $M_H = 115$ GeV.

IV. MULTIVARIATE ANALYSIS

To exploit the kinematics of the $ZH \rightarrow \ell^+ \ell^- b\bar{b}$ process, the energies of the candidate leptons and jets are adjusted within their experimental resolutions with a χ^2 fit that constrains $m_{\ell\ell}$ to the mass and width of the Z , and the p_T of the $\ell^+ \ell^- b\bar{b}$ system to the distribution expected for ZH events. Distributions of the dijet invariant mass spectra, before and after adjustment by the kinematic fit, may be seen in Fig. 3. In the dielectron channel, a cut is applied on the maximum χ^2 from the kinematic fit, optimized to maximize the sensitivity in each ST and DT sample.

A multivariate analysis combines the most significant kinematic information into a single discriminant. Well-modeled kinematic variables are chosen as inputs for the analysis: the transverse momenta of the two b -jet candidates

and the dijet mass, before and after the jet energies are adjusted by the kinematic fit; angular differences within and between the dijet and dilepton systems; the opening angle between the proton beam and the Z candidate in the rest frame of the Z boson [16]; and composite kinematic variables, such as the p_T of the dijet system and the scalar sum of the p_T values of the leptons and jets. We train a random forest (RF), consisting of 200 decision trees [15]. Each tree uses a randomly chosen subset of the simulated signal and background events. At each decision of each tree, a subset of ten of the twenty inputs is randomly chosen for consideration. The RF output is a performance-weighted average of the output from each decision tree. The RF outputs for the ee and $\mu\mu$ channels combined are shown separately for ST and DT events in Fig. 4.

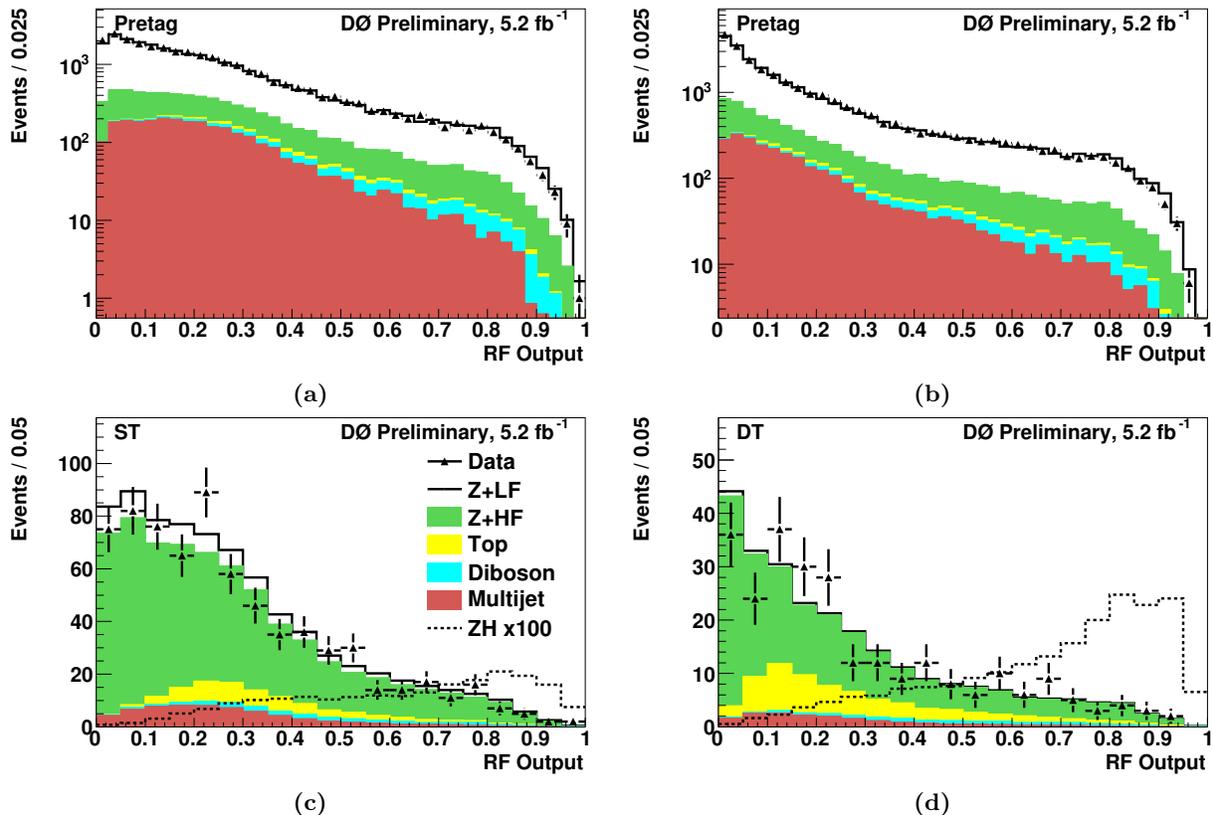


FIG. 4: RF output distributions for the ee and $\mu\mu$ channels combined assuming $M_H = 115$ GeV for (a) pretag events evaluated with the ST-trained RF, (b) pretag events evaluated with the DT-trained RF, (c) ST events evaluated with the ST-trained RF, (d) DT events evaluated with the DT-trained RF.

V. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties resulting from the background normalization are assessed for the multijet contribution (20–60%) and from lepton efficiency effects (2–10%), some of which are correlated between lepton-channels (6%). The normalization of the Z +jets sample to match the pretag data constrains the Z +jets cross section times any jet-dependent efficiency to within the statistical uncertainty of the pretag data (1–2%). Additional systematic uncertainties (10–20%) for possible jet-dependent efficiency effects absorbed into k_{Z+jets} are applied to the $t\bar{t}$, diboson and ZH samples. The normalization to the pretag data, which is dominated by $Z+LF$, does not strongly constrain the cross sections of other processes. For $Z+HF$, a cross section uncertainty of 20% is determined from Ref. [11]. For other backgrounds, the uncertainties are 6%–10%. For the signal, the uncertainty is 6% [9]. The normalization reduces the impact of many of the remaining systematic uncertainties on the background size (except those related to b -tagging), but changes to the shape of the RF output distribution persist and are accounted for. Additional sources of systematic uncertainty include: jet energy scale, jet energy resolution, jet identification efficiency, b -tagging and trigger efficiencies, PDFs, data-determined corrections to the model for Z +jets, and modeling of the underlying event. The uncertainties from the factorization and renormalization scales in the simulation of Z +jets are estimated

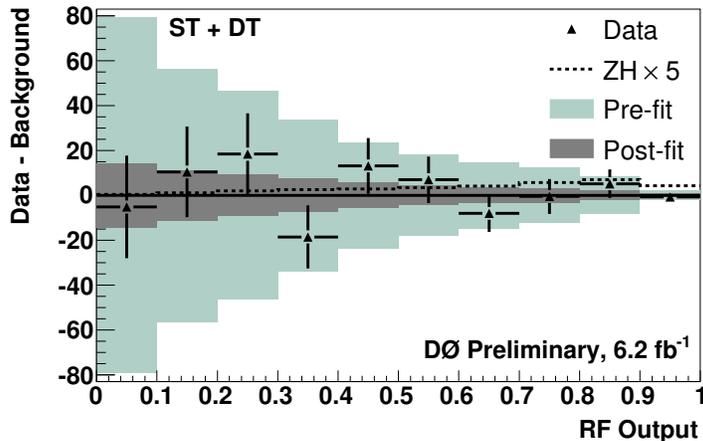


FIG. 5: Background subtracted RF distribution for $M_H=115$ GeV. Both ST and DT events are included. The green (light-shaded) region indicates the systematic uncertainty before the fit from the limit-setting program, while the gray (dark-shaded) region indicates the uncertainty after the fit. Results were obtained using the reanalyzed 5.2 fb^{-1} $\mu\mu$ and ee Run IIb data discussed in this note, as well as the orthogonal samples from Ref. [3] that were not reanalyzed.

by scaling these parameters by factors of 0.5 and two.

VI. RESULTS

No significant excess is observed above the background expectation. Therefore we use the distributions in RF output for the ST and the DT samples in each channel and the corresponding systematic uncertainties to set limits. In addition to the reanalyzed 5.2 fb^{-1} Run IIb dataset, we also use data from Ref. [3] that have not been re-analyzed: the Run IIa dataset; the ee_{ICR} selection, and the $\mu\mu_{\text{trk}}$ selection. The total analyzed luminosity is thus 6.2 fb^{-1} . To take advantage of the sensitivity in the more discriminating channels, we provide separate distributions for each channel to the limit-setting program.

We obtain 95% C.L. limits on the ZH cross section with a modified frequentist (CLs) method that uses a log likelihood ratio (LLR) of the signal+background hypothesis to the background-only hypothesis [17]. To minimize the effect of systematic uncertainties, the likelihoods of the B and S+B hypotheses are maximized by independent fits which vary nuisance parameters that model the systematic effects [18]. The correlations among systematic uncertainties are maintained across channels, backgrounds and signal, as appropriate. The post-fit background-subtracted RF distribution, combined for all channels, with systematic uncertainty bands both before and after the fitting procedure, is shown in Fig. 5.

Figure 6 shows the observed LLR for all data combined, as a function of Higgs boson mass. Also shown are the expected (median) LLRs for the background-only and signal+background hypotheses, together with the one and two standard deviation bands about the background-only expectation. A signal-like excess would result in a negative excursion in observed LLR. For values of $M_H \lesssim 115$ GeV, the data are within one standard deviation of the background-only expectation and within two standard deviations for the entire range of M_H .

The upper limit on the cross section, expressed as a ratio to the SM cross section, as a function of M_H is presented in Table II and Fig. 7. At $M_H = 115$ GeV, the observed (expected) limit on this ratio is 8.0 (5.7). This represents a 20% improvement in the expected limit over that in Ref. [3].

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); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC and the Royal Society (United Kingdom); MSMT and GACR (Czech Republic); CRC Program and NSERC (Canada); BMBF and DFG (Germany);

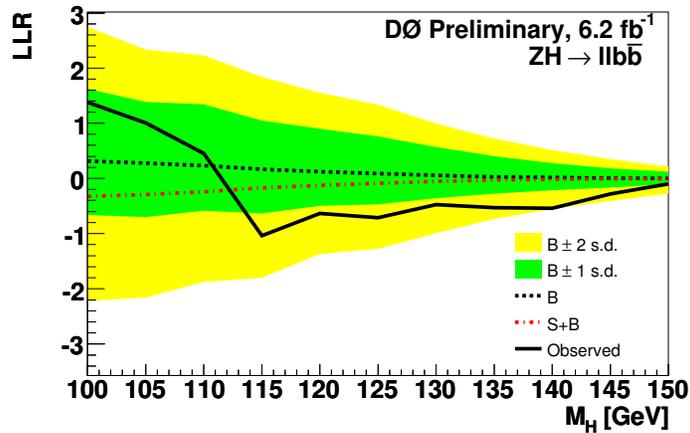


FIG. 6: Observed LLR as a function of Higgs boson mass. Also shown are the expected LLRs for the background-only (B) and signal+background (S+B) hypotheses, together with the one and two standard deviation (s.d.) bands about the background-only expectation. Results were obtained using the reanalyzed 5.2 fb^{-1} $\mu\mu$ and ee Run IIb data discussed in this note, as well as the orthogonal samples from Ref. [3] that were not reanalyzed.

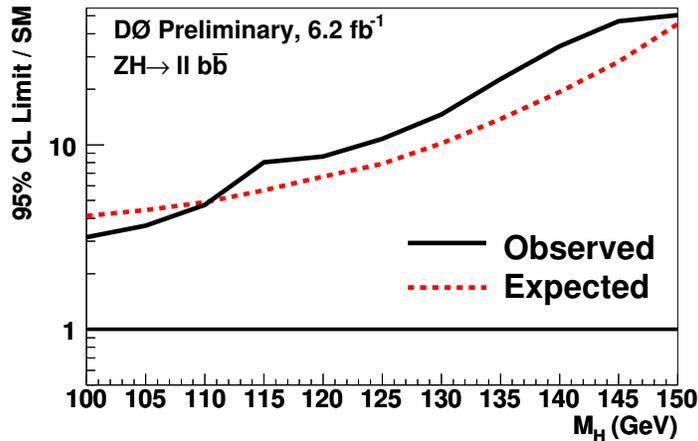


FIG. 7: Expected and observed 95% C.L. cross section upper limits compared to expected limit for the background-only hypothesis as a function of M_H . Limits were obtained using the reanalyzed 5.2 fb^{-1} $\mu\mu$ and ee Run IIb data discussed in this note, as well as the orthogonal samples from Ref. [3] that were not reanalyzed.

SFI (Ireland); The Swedish Research Council (Sweden); and CAS and CNSF (China).

-
- [1] V.M. Abazov *et al.* (D0 Collaboration), DØ Note 6087-CONF (2010)
V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **104**, 071801 (2010).
[2] V. M. Abazov *et al.* (D0 Collaboration), Nucl. Instrum. Methods Phys. Res. Sect. A **565**, 463 (2006);
[3] V.M. Abazov *et al.* (D0 Collaboration), “Search for $ZH \rightarrow \ell^+ \ell^- b \bar{b}$ production in 4.2 fb^{-1} of $p\bar{p}$ collisions” (draft in preparation).
[4] G. C. Blazey *et al.*, arXiv:hep-ex/0005012.
[5] T. Sjöstrand *et al.*, Comput. Phys. Commun. **135**, 238 (2001). Version 6.409 was used.
[6] M. L. Mangano *et al.*, J. High Energy Phys. **0307**, 001 (2003). Version 2.11 was used
[7] J. Pumplin *et al.*, J. High Energy Phys. **07**, 012 (2002).
[8] R. Brun, F. Carminati, CERN Program Library Long Writeup W5013 (1993).
[9] J. Baglio and A. Djouadi, arXiv:1003.4266 [hep-ph].
[10] A. Djouadi, J. Kalinowski and M. Spira, Comput. Phys. Commun. **108**, 56 (1998).
[11] J. Campbell and R.K. Ellis, <http://mcfm.fnal.gov/>.

M_H (GeV)	100	105	110	115	120	125	130	135	140	145	150
Expected	4.1	4.4	4.9	5.7	6.7	7.9	10	14	19	28	45
Observed	3.2	3.6	4.7	8.0	8.7	11	15	23	34	47	50

TABLE II: The expected and observed 95% C.L. upper limits on the SM Higgs boson production cross section for $ZH \rightarrow \ell^+\ell^-b\bar{b}$, expressed as a ratio to the SM cross section. Limits were obtained using the reanalyzed 5.2 fb⁻¹ $\mu\mu$ and ee Run IIb data discussed in this note, as well as the orthogonal samples from Ref. [3] that were not reanalyzed.

- [12] R. Hamberg, W.L. van Neerven and W.B. Kilgore, Nucl. Phys. B**359**, 343 (1991), [*Erratum ibid.* B **644**, 403 (2002)].
- [13] V. M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **100**, 102002 (2008).
- [14] T. Gleisberg *et al.*, J. High Energy Phys. 02 (2004) 056; J. Alwall *et al.*, Eur. Phys. J. C **53**, 473 (2008).
- [15] A. Hoecker *et al.*, arXiv:physics/0703039v5 [physics.data-an].
- [16] S. Parke and S. Veseli, Phys Rev D **60**, 093003 (1999).
- [17] T. Junk, Nucl. Instrum. Methods Phys. Res., Sect. A **434**, 435 (1999); A. Read. J. Phys. G **28** 2693 (2002).
- [18] W. Fisher, FERMILAB-TM-2386-E (2007).