PROCEEDINGS OF SCIENCE

The search for standard model Higgs boson in $ZH \to \ell^+\ell^- b\bar{b}$ channel at DØ

Peng Jiang*
University of Science and Technology of China
E-mail: pengj@fnal.gov
on behalf of the DØ collaboration

We present a search for the Standard Model Higgs boson in the $ZH \to \ell^+\ell^- b\bar{b}$ channel, using data collected with the DØ detector at the Fermilab Tevatron Collider. This analysis is based on a sample of reprocessed data incorporating several improvements relative to a previous published result, and a modified multivariate analysis strategy. For a Standard Model Higgs boson of mass 125 GeV, the expected cross section limit over the Standard Model prediction is improved by about 5% compared to the previous published results in this channel from the DØ Collaboration.

XXII. International Workshop on Deep-Inelastic Scattering and Related Subjects,
28 April - 2 May 2014
Warsaw, Poland

*Speaker.

PoS(DIS2014)101

1. Introduction

In the standard model (SM), spontaneous breaking of the electroweak gauge symmetry generates masses for the W and Z bosons and produces a new scalar elementary particle, the Higgs boson [1]. A narrow boson resonance with a mass of about 126 GeV has been discovered in the $\gamma\gamma$, ZZ and WW decay channels by the ATLAS and CMS experiments at the CERN Large Hadron Collider (LHC) [2] [3]. The CDF and D0 experiments at the Fermilab Tevatron Collider supported the interpretation of the LHC boson as the Standard Model Higgs boson by reporting evidence for $H \to b\bar{b}$ in VH production [4] and overall consistency when combining all production and decay channels [5]. Subsequent experimental results have continued to support the discovery of the Standard Model Higgs boson with a mass of 126 GeV [6].

We present an analysis of Higgs boson associated production with a Z boson, with subsequent decay of the Z boson into a lepton pair ($Z \to ee$ and $Z \to \mu\mu$), and the decay of the Higgs boson into a b quark-antiquark pair ($H \to b\bar{b}$). This analysis is based on a fully reprocessed D0 dataset that uses improved reconstruction algorithms and incorporating a number of improvements relative to the previously published D0 result [7] including optimization of the lepton selection, optimized corrections to our simulated Monte Carlo (MC) samples, improved identification algorithms for b-quark jet, and a new two-step multivariate analysis strategy.

2. Data Sample

The data sample was collected by the D0 detector [8], between April 2002 and September 2011, and is split into two periods, one prior to March 2006 (Run IIa), corresponding to 1.1 fb^{-1} of data, while the period after is referred to as Run IIb (8.6 fb^{-1}), and benefits from the installation of an additional layer of silicon vertex detector, trigger upgrades, and a significant increase in the rate of delivered luminosity. The D0 experiment reprocessed the Run IIb data sample with improved reconstruction algorithms. The reprocessing uses an improved vertexing algorithm (previously different vertexing algorithms were used in different run periods) and better tracker alignment, which improves the track impact parameter resolution by up to 15%. The biggest improvement in the reprocessing is the newly added "Track-in-Road" algorithm, which uses electrons or muons reconstructed outside of the central tracker as a start point for building tracks in the central tracker, improving the lepton efficiency in reconstruction. New sets of dedicated Monte Carlo (MC) samples have been created to model the Run IIb reprocessed data. We retain the identical analysis strategy and selection requirements as in the previous analysis [7] for the Run IIa data, but then combine with new results presented here for the Run IIb reprocessed data. The total integrated luminosity is 9.7 fb^{-1} after imposing the data quality requirements, and 8.6 fb^{-1} is from reprocessed Run IIb data.

3. Event Selection

The signature for ZH events is two leptons compatible with a Z decay and two b-jets compatible with a Higgs boson decay. We therefore first select the events with two oppositely charged leptons (two electrons or two muons), plus two jets. We separate the lepton channel selection into

di-electron channels and a di-muon channel. For electron channels, we first select events containing one electron in the central calorimeter(CC) region with $|\eta_{\rm det}| < 1.1$, where $\eta_{\rm det}$ is the pseudorapidity measured with respect to the center of the detector; and another electron either in the CC or the end calorimeters(EC) region with $1.5 < |\eta_{\rm det}| < 2.5$, denoted as the *ee* channel. We accept events that satisfy any trigger requirement for the *ee* channel, with a measured efficiency consistent with 100% to within 1%. Both of the electrons are required to have the transverse momenta $p_{\rm T} > 15$ GeV, unless the sub-leading electron is in CC region, for which we lowered the pT requirement to 10 GeV to enlarge our acceptance comparing to the previous analysis [7]. All electrons are selected from electromagnetic(EM) clusters reconstructed within a cone of radius $\mathscr{R}$ 0.4, and have to pass the selection requirements [9] based on the energy deposition and shower shape in the calorimeter and the central preshower(CPS). We further select events containing one electron in CC and the other one in the inter-cryostat regions(ICR) region with $1.1 < |\eta_{\rm det}| < 1.5$ ($ee_{\rm ICR}$ channel), that require triggers that require electrons and jets with an efficiency of 90-100% depending on the region of the detector toward which the electron points. An $ee_{\rm ICR}$ event has to contain one electron in the CC with $p_{\rm T} > 15$ GeV, and a track pointing towards a ICR region. The electron in the CC has the same requirements as in the *ee* channel. The ICR track must have a measured transverse energy in the plastic scintillators satisfying $E_{\rm T} > 15$ GeV. For the di-muon channel, we select events containing two muons from events satisfying any trigger requirement. The two muons must have transverse momenta $p_{\rm T} > 10$ GeV and $|\eta_{\rm det}| < 2$. At least one muon must have $p_{\rm T} > 15$ GeV and $|\eta_{\rm det}| < 1.5$. For all lepton channels, we require a $p\overline{p}$ interaction vertex (PV) that has at least three associated tracks, and is located within $\pm$60 cm of the center of the detector along the beam direction. The distance between the PV and each of the muon or electron tracks along the *z* axis, $d^z_{\rm PV}$, must be less than 1 cm.

We reconstruct two jets in the calorimeter using an iterative midpoint cone algorithm [10] with a cone size of $\mathscr{R} = 0.5$. The energies of jets are corrected for detector response, the presence of noise and multiple $p\overline{p}$ interactions, and energy measured outside (inside) the jet cone from particles produced inside (outside) the cone [11]. All jets must have $p_{\rm T} > 20$ GeV and $|\eta_{\rm det}| < 2.5$ and contain at least two associated tracks originating from the PV and have at least one hit from the central silicon detector. Jets meeting these criteria are considered "taggable" by the *b*-tagging algorithm described below. We select events that have at least two taggable jets with $p_{\rm T} > 20$ GeV and $|\eta_{\rm det}| < 2.5$, and $70 < M_{\ell\ell} < 110$ GeV. Figure 1 shows the selected events dilepton and dijet invariant mass distributions.

An improved multivariate *b*-tagging discriminant [12] was developed to distinguish the taggable jets decay from *b* quarks and from light flavor (LF) quarks (*uds*), using several boosted decision trees as inputs, combining information related to the associated tracks, primary vertex and secondary vertices. The *b*-tagging tools provides several operating points for which the corresponding efficiencies are measured in our data and MC samples. We define two orthogonal samples based on the selected sample described above:

- a Loose-Tight (LT) sample, which has one jet satisfying a tight *b*-tagging operating point A, and another jet that satisfying a loose *b*-tagging operating point B;
- a Single-Tight (ST) sample, which does not satisfy the LT requirements, but contains one jet satisfying a tight *b*-tagging operating point C.

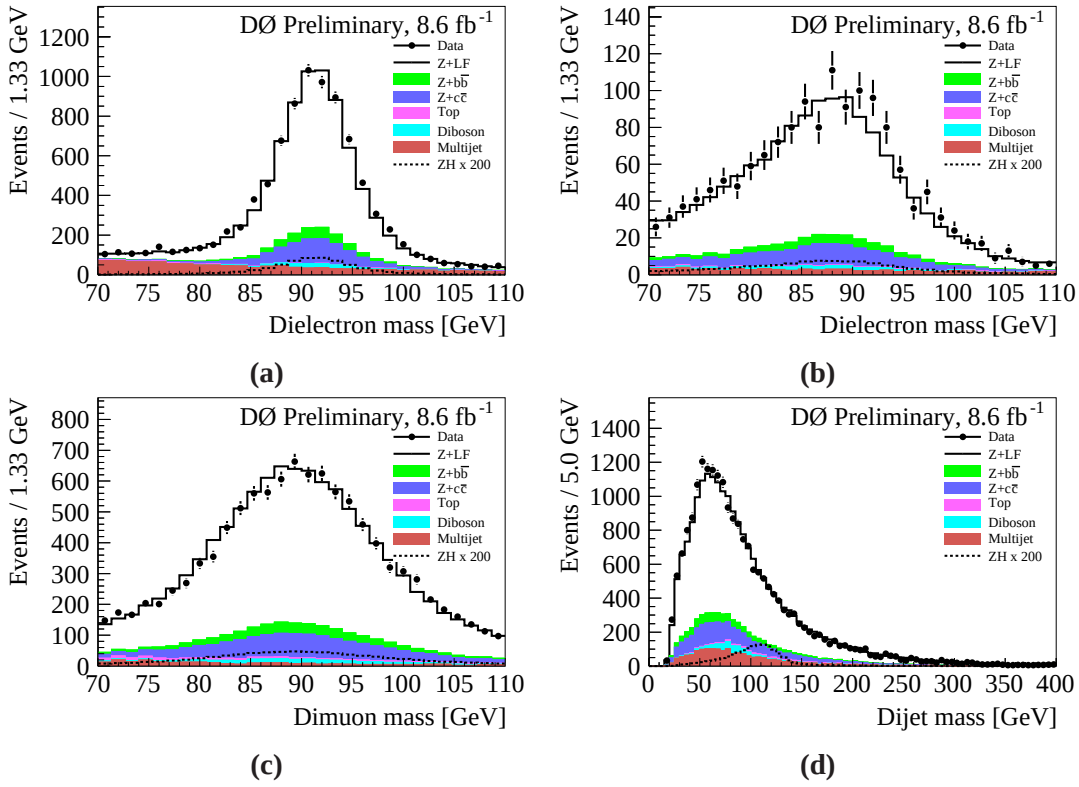

Figure 1: The dilepton mass spectra in the (a) *ee*, (b) ee_{ICR} and (c) $\mu\mu$ channels. (d) The dijet mass spectra for all lepton channels combined. Signal distributions, for $M_H = 125$ GeV, are scaled by a factor of 200.

We have done an optimization over all combinations of operation points (OPs) A, B, and C to find the best expected limits on *ZH* production. The *b*-tagging efficiency and fake rate for our chosen OP combinations are: (A) $\sim 60\%/1\%$, (B) $\sim 80\%/8\%$, and (C) $\sim 70\%/2\%$ depending on the jet p_{T} and η.

To improve the resolution of the dijet invariant mass, we perform a kinematic fit based on three constraints: the reconstructed dilepton mass should be consistent to a Gaussian distribution with a width of $\Gamma_Z = 2.50$ (GeV). and the *x* and *y* components of the vector sum of the transverse momenta of the leptons and jets should be consistent with zero. The energies and angles of the two leptons from the *Z* boson candidate, and of the two jets that arise from the Higgs boson candidate (and of a third jet,if present) are allowed to vary within their experimental resolutions in the fit. The fit yields about 12% improvement in the dijet mass resolution from a resolution of 13.5% to 12% for $m_H = 125$ GeV after the kinematic fit. Figure 2 shows the dijet invariant mass distributions before and after kinematic fit.

4. Multivariate Analysis

The main backgrounds for this analysis are $t\bar{t}$, $Z + b\bar{b}$, $Z + c\bar{c}$, *Z*+LF, diboson. To further distinguish our signal from backgrounds, we use a two-step multivariate analysis (MVA) strategy based on random forest discriminants (RF), an ensemble classifier that consists of many decision trees [13], as implemented in the TMVA software package [14]. In the first step, we train four

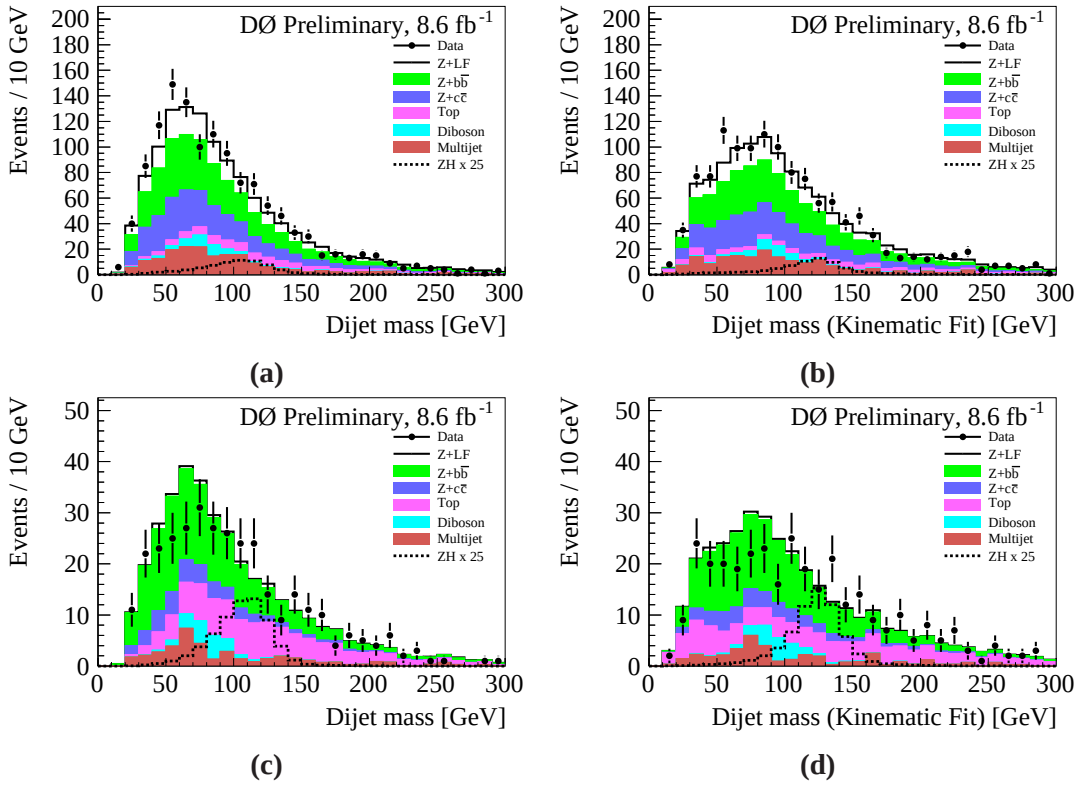

Figure 2: The dijet mass distribution before kinematic fit for (a) ST , (c) LT and after kinematic fit for (b) ST , (d) LT.

intermediate RFs for our signal against each group of backgrounds, (i) $t\bar{t}$, (ii) $Z+b\bar{b}$, (iii) $Z+c\bar{c}$, Z+LF and (iv) dibosons. For each RF output distribution we identify a cut value that best separates the sample into subsamples enriched and depleted in the background under consideration. We then define five orthogonal regions, each of which is rich in one of the backgrounds (or signal) but depleted in the others:

- Region 1: $t\bar{t}$ enriched region;
- Region 2: $t\bar{t}$ depleted, $Z+b\bar{b}$ enriched region;
- Region 3: $t\bar{t}$ and $Z+b\bar{b}$ depleted, $Z+c\bar{c}$ and Z+LF enriched region;
- Region 4: $t\bar{t}$, $Z+b\bar{b}$, $Z+c\bar{c}$, Z+LF depleted, diboson enriched region;
- Region 5: $t\bar{t}$, $Z+b\bar{b}$, $Z+c\bar{c}$, Z+LF, diboson depleted region.

In the second step, we train global RFs which separate our signal from all backgrounds in these five separated regions for the LT and ST events separately. Each lepton channel samples corresponding to each region and tag category are passed through the appropriate RF. The resulting global RFs for five separated region, two tagging categories and each lepton channel are used for calculation of limits on cross section times branching ratio in the $ZH \to \ell^+\ell^- b\bar{b}$ channel.

5. Conclusion

Since we do not have significant sensitivity to the SM Higgs boson with this channel alone, we set upper limits on the *ZH* production cross section times branching ratio for a Higgs decaying into $b\bar{b}$ with *Z* decaying into two leptons, using the final discriminants in the five orthogonal regions for LT and ST events in four lepton channels described in Section 4. The limits are calculated at 95% C.L. with a modified frequentist (CLs) method that uses the log likelihood ratio (LLR) of the signal+background hypothesis to the background-only hypothesis [15]. We decrease the effect of systematic uncertainties on the limit sensitivity by fitting the parameters controlling them, constrained by their priors. We estimate that the expected limit from this analysis is 5.4 times the Standard Model cross section times branching ratio using an approximation method AWW described in Ref [16] [17]. The expected limit is improved by about 5% compared with those for the same channels in the Run IIb data set obtained in Ref. [7]. This analysis thus confirms that the results of Ref. [7] are close to optimum.

References

[1] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964); P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964); G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. **13**, 585 (1964).

[2] G. Aad et al. (ATLAS Collaboration), Phys. Lett. B **716**, 1 (2012).

[3] S. Chatrchyan et al. (CMS Collaboration), Phys. Lett. B **716**, 30 (2012)

[4] T. Aaltonen et al. (CDF Collaboration, D0 Collaboration), Phys. Rev. Lett. **109**, 071804 (2012) V.M. Abazov et al. (D0 Collaboration), Phys. Rev. Lett. **109**, 121802 (2012).

[5] T. Aaltonen et al. (CDF Collaboration, D0 Collaboration), Phys. Rev. D **88**, 052014 (2013)

[6] CMS Collaboration, J. High Energy Phys. 01 (2014) 096 CMS Collaboration, Phys. Rev. D **89** (2014) 012003 ATLAS Collaboration, Phys. Lett. B **726** (2013) 88-119

[7] V.M. Abazov et al. (D0 Collaboration), Phys. Rev. D **88**, 052010 (2013).

[8] V.M. Abazov *et al.*, (DØ Collaboration), Nucl. Instrum. Methods Phys. Res. A **565**, 463 (2006).

[9] V.M. Abazov *et al.*, (DØ Collaboration), Nucl. Instrum. Methods Phys. Res. A **750**, 78 (2014).

[10] G. C. Blazey *et al.*, arXiv:hep-ex/0005012.

[11] V.M. Abazov *et al.*, (DØ Collaboration), arXiv: 1312.6873.

[12] V. M. Abazov *et al.* (D0 Collaboration), arXiv: 1312.7623.

[13] L. Breiman, Machine Learning **45**, 5 (2001).

[14] H. Voss *et al.*, PoS (**ACAT**), 040 (2007), arXiv:physics/0703039.

[15] T. Junk, Nucl. Instrum. Methods Phys. Res., Sect. A **434**, 435 (1999); A. L. Read, Workshop on Confidence Limits, CERN-OPEN-2000-205 (2000).

[16] Eric Burns, Wade Fisher, arXiv:1110.5002.

[17] W. Fisher, FERMILAB-TM-2386-E (2006).