CDF/PHYS/EXOTIC/PUBLIC/9718

# A Search for the Production and Decay Process $ZH \to l^- l^+ b\bar{b}$ in 2.7 fb$^{-1}$ using the Matrix Element Method

The CDF Collaboration
*URL http://www-cdf.fnal.gov*
(Dated: March 10, 2009)

We present a method for searching for associated production of the standard model (SM) Higgs boson and a $Z$ boson where the $Z$ boson decays to two leptons and the Higgs decays to a pair of $b$ quarks, $p\bar{p} \to ZH \to l^- l^+ b\bar{b}$. We use a matrix element weighting technique to construct the likelihood functions of the Higgs content of the data sample. In 2.7 fb$^{-1}$ of CDF data we observe a 95% C.L. of 7.8 times the SM prediction while the median expected 95% C.L. is 12.3 times the SM prediction for $m_H$ =115 GeV/$c^2$.

## I. INTRODUCTION

Discovery of the Higgs boson remains one of the highest priorities of the Tevatron program. At high masses, sensitivity to a standard model (SM) Higgs boson can be achieved by searching for direct production $gg \to H$ of the Higgs as its decay is dominated by relatively clean decays to diboson final states, $WW$ and $ZZ$. For low masses, however, background swamps the dominant decay channel of $b\bar{b}$ making it necessary to rely on associated production of the Higgs boson with a $W$ or $Z$ boson. Of these, $ZH$ production where the $Z$ decays to two charged leptons offers a clean final state with no neutrinos but has a very small branching ratio. The size of the potential signal in this channel makes a simple counting experiment prohibitive. We utilize a method that uses leading-order matrix element calculations to form per-event likelihoods. We further increase our sensitivity by modeling the dominant background processes in a similar fashion. By combining these per-event likelihoods we obtain a sample likelihood of its signal content. This is one of two complementary techniques used at CDF for a search in this channel, the other being a 2 dimensional neural network [1].

| Source | $ee$ | $\mu\mu$ | $ll$ |
|---|---|---|---|
| $Z \to ee$ | $70.7 \pm 18.8$ | $0.0 \pm 0.0$ | $70.7 \pm 18.8$ |
| $Z \to ee + hf$ | $63.0 \pm 11.4$ | $0.0 \pm 0.0$ | $63.0 \pm 11.4$ |
| $Z \to \mu\mu$ | $0.0 \pm 0.0$ | $58.9 \pm 14.9$ | $58.9 \pm 14.9$ |
| $Z \to \mu\mu + hf$ | $0.0 \pm 0.0$ | $44.2 \pm 8.1$ | $44.2 \pm 8.1$ |
| $Z \to \tau\tau$ | $0.1 \pm 0.1$ | $0.0 \pm 0.1$ | $0.1 \pm 0.1$ |
| $Z \to \tau\tau + hf$ | $0.0 \pm 0.1$ | $0.0 \pm 0.1$ | $0.0 \pm 0.1$ |
| $WW, WZ, ZZ$ | $6.6 \pm 1.0$ | $5.0 \pm 0.8$ | $11.6 \pm 1.3$ |
| Fakes | $12.8 \pm 6.4$ | $3.1 \pm 1.2$ | $15.9 \pm 6.5$ |
| $t\bar{t}$ | $7.7 \pm 1.5$ | $6.2 \pm 1.2$ | $13.9 \pm 2.0$ |
| $ZH(m_H = 115$ GeV$/c^2)$ | $0.7 \pm 0.1$ | $0.5 \pm 0.1$ | $1.3 \pm 0.2$ |
| Total | $161.6 \pm 23.0$ | $117.9 \pm 17.1$ | $279.5 \pm 28.6$ |
| Data | 152 | 106 | 258 |

**TABLE I:** Expected yields with the $\geq 1$ tight $b-$tag requirement, and the corresponding number of observed data events.

| Source | $ee$ | $\mu\mu$ | $ll$ |
|---|---|---|---|
| $Z \to ee$ | $2.9 \pm 0.7$ | $0.0 \pm 0.0$ | $2.9 \pm 0.7$ |
| $Z \to ee + hf$ | $11.4 \pm 2.8$ | $0.0 \pm 0.0$ | $11.4 \pm 2.8$ |
| $Z \to \mu\mu$ | $0.0 \pm 0.0$ | $2.6 \pm 0.6$ | $2.6 \pm 0.6$ |
| $Z \to \mu\mu + hf$ | $0.0 \pm 0.0$ | $8.1 \pm 1.9$ | $8.1 \pm 1.9$ |
| $WW, WZ, ZZ$ | $1.7 \pm 0.3$ | $1.2 \pm 0.2$ | $2.9 \pm 0.4$ |
| Fakes | $0.2 \pm 0.1$ | $0.2 \pm 0.1$ | $0.4 \pm 0.2$ |
| $t\bar{t}$ | $4.4 \pm 0.9$ | $3.3 \pm 0.7$ | $7.7 \pm 1.1$ |
| $ZH(m_H = 115$ GeV$/c^2)$ | $0.4 \pm 0.1$ | $0.4 \pm 0.1$ | $0.7 \pm 0.1$ |
| Total | $20.8 \pm 3.0$ | $15.5 \pm 2.1$ | $36.3 \pm 3.7$ |
| Data | 16 | 16 | 32 |

**TABLE II:** Expected yields with the $\geq 2$ loose $b-$tag requirement, and the corresponding number of observed data events.

## II. EVENT SELECTION

### A. Dataset and Trigger

This analysis is based on an integrated luminosity of 2.7 fb$^{-1}$ collected with the CDF II detector between March 2002 and April 2008. The CDF II detector is a general purpose detector described elsewhere [2]. The data are collected with an inclusive lepton trigger that requires events to have a lepton with $E_T > 18$ GeV (for an electron) or $p_T > 18$ GeV$/c$ (for a muon).

### B. Final Sample

The event selection used in this analysis closely follows that of the first $ZH \to l^- l^+ b\bar{b}$ analysis performed at CDF II [3]. Following the trigger requirement outlined above, events with two leptons are selected where the invariant mass of the leptons falls within $76 < m_{ll} < 106$ GeV$/c^2$. Events are then required to have at least two jets with $E_T > 15$ GeV and $\eta_{\text{jet}} < 2.0$ where at least one jet has $E_T > 25$ GeV. Validation of the data with these cuts is shown for a variety of kinematic variables in Fig. 1. Finally, events are required to have either 2 loose SecVtx [4] $b$-tags or $\geq 1$ tight SecVtx $b$-tags. Validation of the data with all cuts is shown for a variety of kinematic variables in Figs. 2 and 3. After event selection, the remaining sample is dominated by several sources of background: Drell-Yan with associated jets, which are modeled separately for sources with real heavy flavor and sources where light flavor mimics heavy flavor; $WW$, $WZ$ and $ZZ$ events with heavy flavor; $t\bar{t}$; and events with one real lepton, and an object such as a jet mimicking a lepton (Fakes). The predicted number of events for each source of background and for signal for our data sample as well as the total number of events observed in data are shown in Table I for events with $\geq 1$ tight SecVtx $b$-tags and in Table II for events with 2 loose SecVtx [4] $b$-tags.

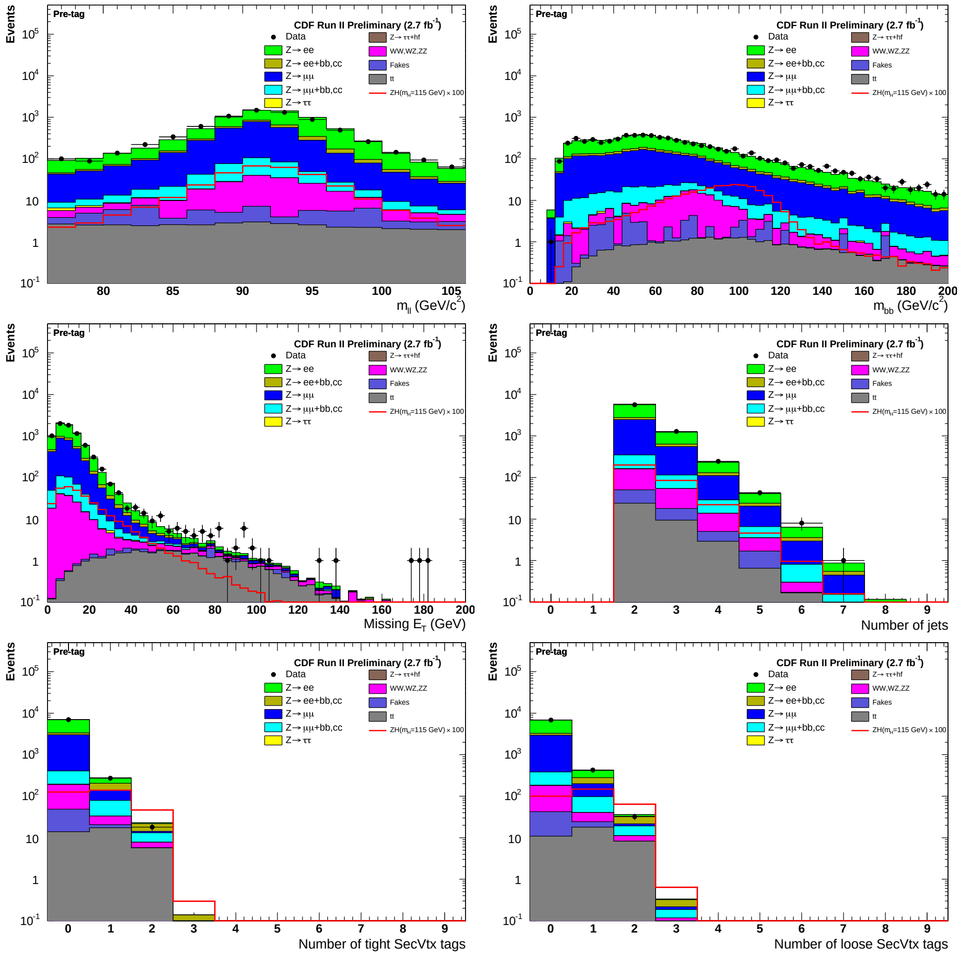


**FIG. 1:** Validation of dilepton invariant mass $m_{ll}$, dijet invariant mass $m_{jj}$, missing transverse energy $\not{E}_T$, jet multiplicity, number of tight $b$-tags and number of loose $b$-tags for events passing all selection cuts prior to tag requirement.

## III. METHOD

### A. Likelihood Determination

We denote the $ZH$ signal probability by $P_{ZH}(\mathbf{x}|m_H)$ where the Higgs boson mass $m_H$ is an input parameter and $\mathbf{x}$ is a vector of event kinematics. Similarly we denote the background probability as $P_b(\mathbf{x})$. We then construct the likelihood for event $i$ as

$$L(s, \mathbf{x_i}|m_H) = sP_{ZH}(\mathbf{x_i}|m_H) + (1-s)P_b(\mathbf{x_i}) \tag{1}$$

where $s$ denotes a possible value of the signal fraction. To formulate $P_{ZH}$ and the $P_b$, we utilize leading-order matrix-element calculations. Thus, $P_{ZH}$ is given by

$$P_{ZH}(\mathbf{x}|m_H) = \frac{1}{\sigma(m_H)}\frac{d\sigma(m_H)}{d\mathbf{x}} \tag{2}$$

where $d\sigma/d\mathbf{x}$ is the differential cross section evaluated with respect to the event measurements contained in $\mathbf{x}$. We evaluate the differential cross section by convoluting a leading order matrix element for the process with detector resolution functions and integrate over unmeasured quantites. Thus, the probability density can be expressed as

$$P_{ZH}(\mathbf{x}|m_H) = \frac{1}{\sigma(m_H)}\int d\Phi|\mathcal{M}_{ZH}(q,p;m_H)|^2\prod_i[W(p_i,\mathbf{x})]f_{PDF}(q_1)f_{PDF}(q_2) \tag{3}$$

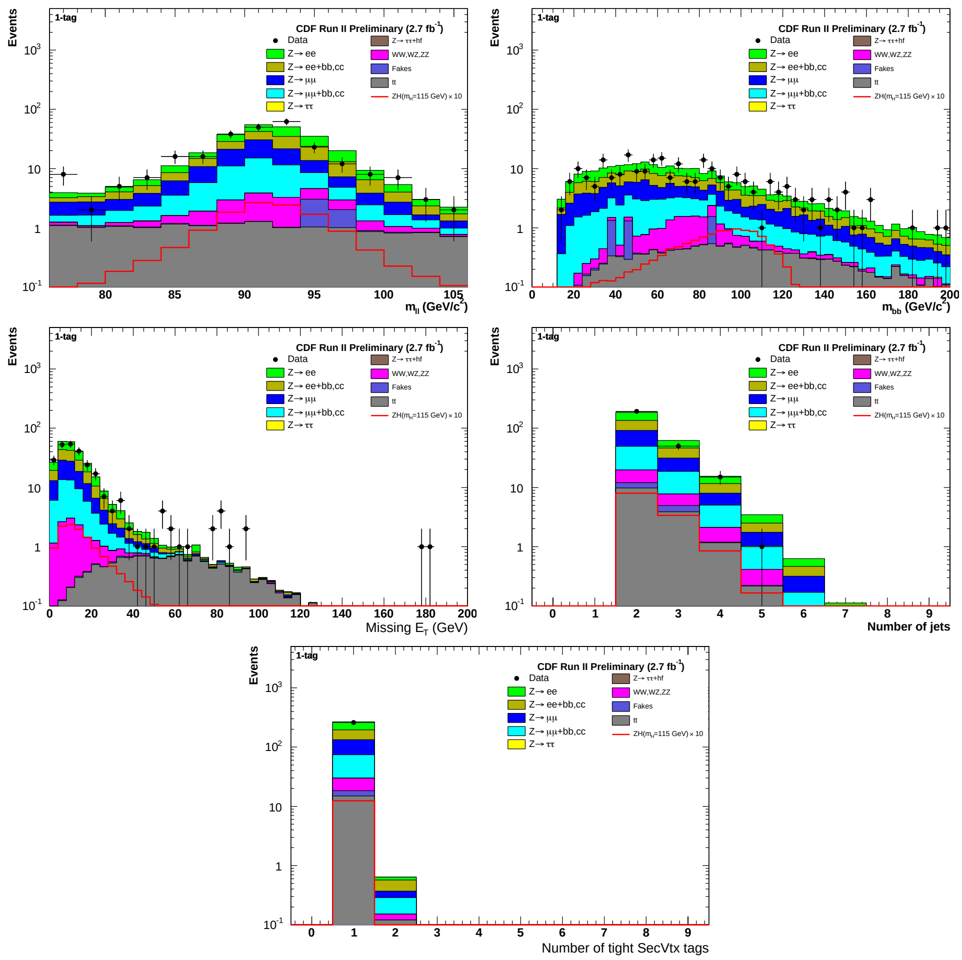


**FIG. 2:** Validation of dilepton invariant mass $m_{ll}$, dijet invariant mass $m_{jj}$, missing transverse energy $\not\!\!E_T$, jet multiplicity, and number of tight $b$-tags for events with $\geq 1$ $b$-tag.

where $\mathcal{M}_{ZH}$ is the leading order matrix element for the process $q\bar{q} \to ZH \to l^-l^+b\bar{b}$ evaluated for a set of incoming partons $q$ and outgoing partons $p$, $W(p_i, \mathbf{x})$ are transfer functions linking the outgoing parton momenta $p_i$ to measured quantities $\mathbf{x}$ and the $f_{PDF}$ are parton density functions for the initial state. We calculate probabilities for the signal process and for three dominant backgrounds: Drell-Yan with associated jets, $ZZ$ with associated jets and $t\bar{t}$. Validation of these four probabilities is shown in Fig. 4.

The sample likelihood is obtained by taking the product over all events $i$ in the sample

$$\mathcal{L}(s|m_H) = \prod_i L(s, \mathbf{x_i}|m_H). \tag{4}$$

$\mathcal{L}(s|m_H)$ is evaluated for all values of $s$ and the value of $s = s_m$ is evaluated for which $\mathcal{L}$ attains its maximum value. Thus $s_m$ denotes the maximum-likelihood estimate of the Higgs signal fraction in the event sample.

We further enhance our statistical sensitivity by splitting our sample into events with 1 $b$-tag and events with 2 $b$-tags. Thus, in a sample of $N$ events, where $N_1$ events have 1 $b$-tag and $N_2$ events have 2 $b$-tags, we define the 1-tag and 2-tag signal fractions $s_1$ and $s_2$ such that the likelihood for the sample of $n$-tagged events is

$$\mathcal{L}_n(s_n|m_H) = \prod_i [s_n P_{ZH}(\mathbf{x}|m_H) + (1 - s_n) P_{b,n}(\mathbf{x})] \tag{5}$$

where $s_n = N_{ZH,n}/N_n$ is the signal fraction in the $n$-tagged sample and the final sample likelihood is given by

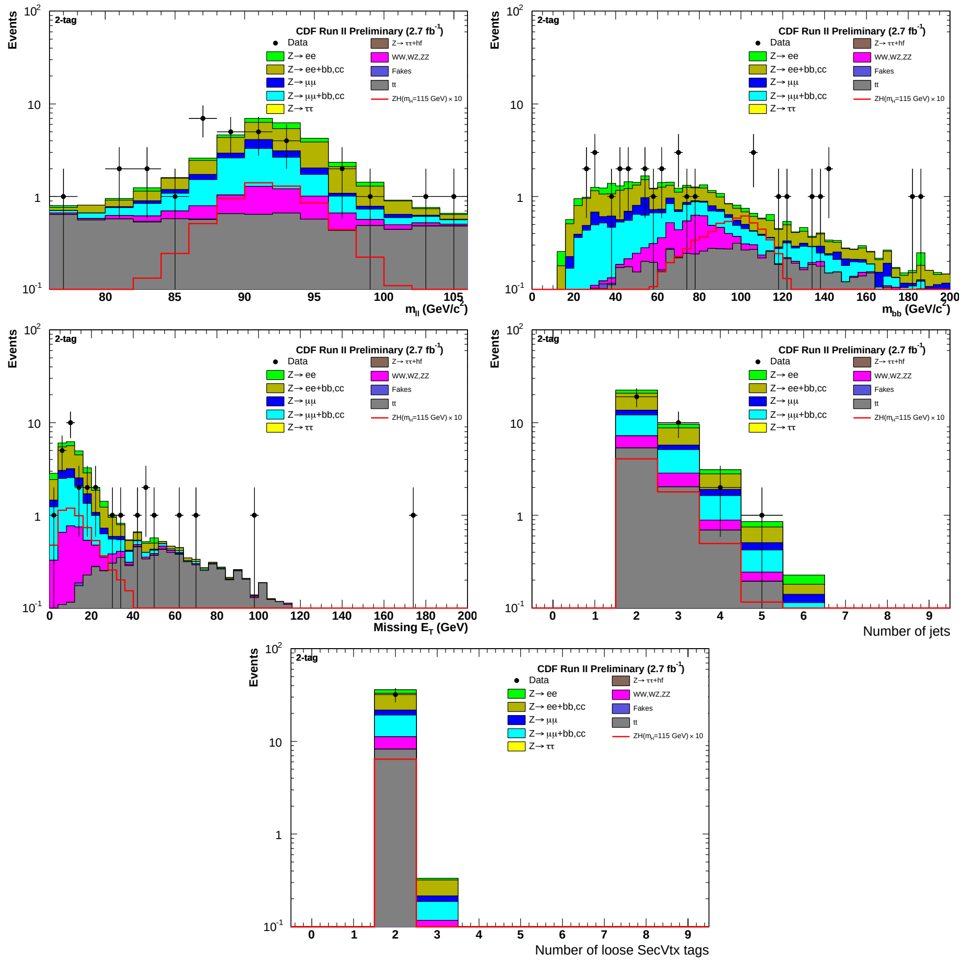


**FIG. 3:** Validation of dilepton invariant mass $m_{ll}$, dijet invariant mass $m_{jj}$, missing transverse energy $\not\!\!E_T$, jet multiplicity, and number of loose $b$-tags for events with 2 loose $b$-tags.

$$\mathcal{L}(s|m_H) = \mathcal{L}_1 \times \mathcal{L}_2 \tag{6}$$

In our data sample, we expect backgrounds from different sources. The dominant backgrounds are due to the $Zjj$, $t\bar{t}$ and $ZZ$ processes, in the expected proportions denoted by $\lambda_{Zjj}^{n-\mathrm{tag}}$, $\lambda_{t\bar{t}}^{n-\mathrm{tag}}$ and $\lambda_{ZZ}^{n-\mathrm{tag}}$ respectively for events with $n = 1$ or 2 tags. Thus, the background probability is given by

$$P_{b,n}(\mathbf{x}) = \lambda_{Zjj}^{n-\mathrm{tag}} P_{Zjj}(\mathbf{x}) + \lambda_{t\bar{t}}^{n-\mathrm{tag}} P_{t\bar{t}}(\mathbf{x}) + \lambda_{ZZ}^{n-\mathrm{tag}} P_{ZZ}(\mathbf{x}) \tag{7}$$

where $P_{Zjj}(\mathbf{x})$, $P_{t\bar{t}}(\mathbf{x})$ and $P_{ZZ}(\mathbf{x})$ are the respective probabilities for the $Z + jj$, $t\bar{t}$, and $ZZ$ background processes. The sample likelihood of Eq. 4 can thus be expressed as

$$\begin{aligned}\mathcal{L}(s|m_H) = &\prod_{i_1} f_1 s P_{ZH}(\mathbf{x}|m_H) + \Big(1 - f_1 s\Big)[\lambda_{Zjj}^{1-\mathrm{tag}} P_{Zjj}(\mathbf{x}) + \lambda_{t\bar{t}}^{1-\mathrm{tag}} P_{t\bar{t}}(\mathbf{x}) + \lambda_{ZZ}^{1-\mathrm{tag}} P_{ZZ}(\mathbf{x})]\} \\ &\times \prod_{i_2} f_2 s P_{ZH}(\mathbf{x}|m_H) + \Big(1 - f_2 s\Big)[\lambda_{Zjj}^{2-\mathrm{tag}} P_{Zjj}(\mathbf{x}) + \lambda_{t\bar{t}}^{2-\mathrm{tag}} P_{t\bar{t}}(\mathbf{x}) + \lambda_{ZZ}^{2-\mathrm{tag}} P_{ZZ}(\mathbf{x})]\}\end{aligned} \tag{8}$$

where the products are over all 1-tag and 2-tag events respectively and the constant $f_n$ are defined as

$$f_n = \frac{N_{ZH,n}}{N_n}\frac{N}{N_{ZH}}. \tag{9}$$

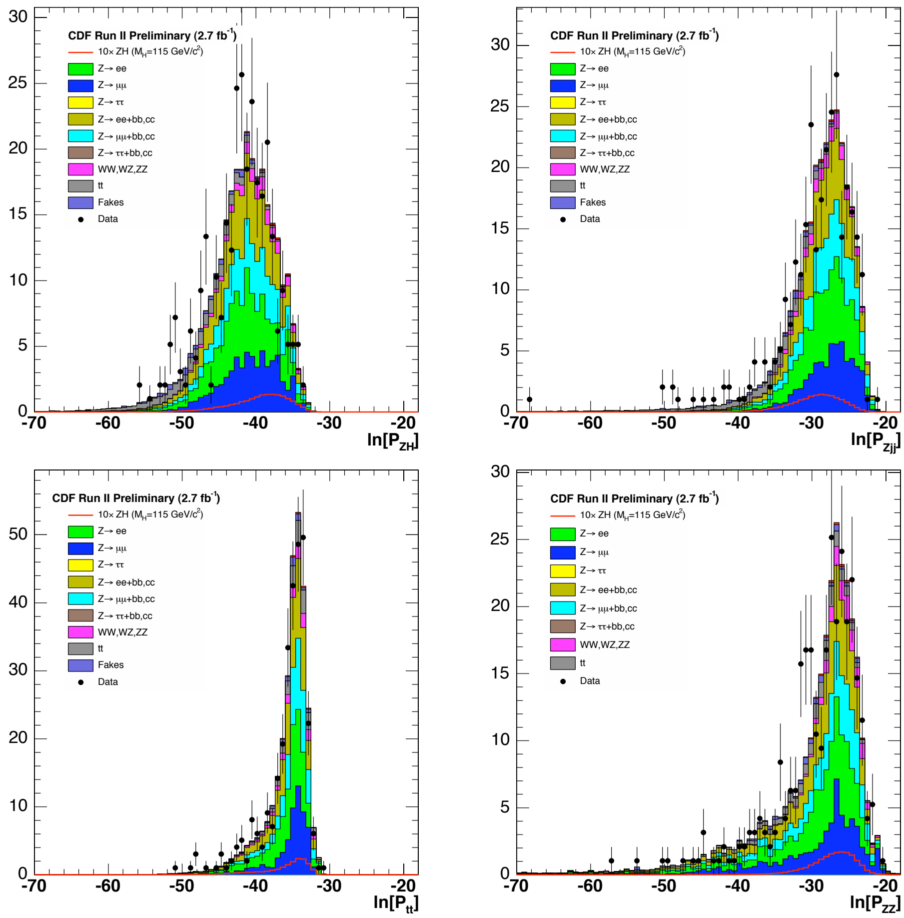

**FIG. 4:** The distributions of each of matrix-element probabilities log$P$, evaluated on the final selected sample of data events, and on simulated samples.

All the $P$ are defined as probability distributions functions, *i.e.*

$$\int P(\mathbf{x})d\mathbf{x} = 1. \tag{10}$$

### B. Limit Calculation

We utilize the Feldman-Cousins [5] prescription to extract a limit on $ZH \rightarrow l^- l^+ b\bar{b}$ production in data. We utilize the test statistic

$$R = \frac{\mathcal{L}(s_m|s_t)}{\mathcal{L}(s_m|s_t^{best})} \tag{11}$$

where $s_t^{best}$ is the most likely input signal fraction that results in a given measured signal fraction, $s_m$.

## IV. SYSTEMATIC UNCERTAINTY

We separate systematic uncertainties into shape-based and rate-based categories. Shape-based sources of systematics are estimated by performing pseudo-experiments with simulated events that vary the specific source of uncertainty. These sources of uncertainty are Jet Energy Scale (JES) uncertainty and Initial State (ISR) and Final State (FSR) Radiation. Rate-based sources of systematic uncertainty include heavy flavor scale factors ($k$-factors), luminosity, and tagging uncertainty. A summary of rate-based systematics and their magnitude is shown in Table IV.

| Source | Magnitude | Samples Affected |
|---|---|---|
| Luminosity | 6% | All |
| ALPGEN $k$-factor | 40% | $Z$+jets |
| PYTHIA $k$-factor | 20% | $t\bar{t}$, Diboson |
| Mistag uncertainty | 8% | Samples with $b$ quarks |
| | 16% | Samples with $c$ quarks |
| | 13% | Samples with no heavy flavor |
| Fake uncertainty | 50% | Fakes |

**TABLE III:** Summary of rate-based systematics

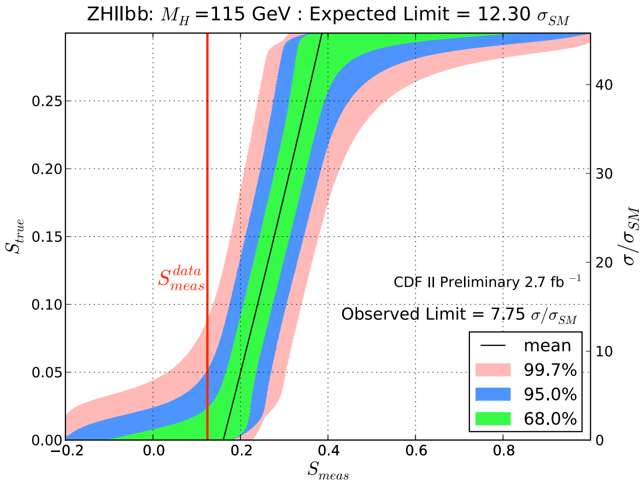


**FIG. 5:** Feldman-Cousins intervals for 68%, 95% and 99.7% coverage for $m_H = 115$ GeV/$c^2$.

## V. RESULTS

We run the data sample of the selected 290 events through the matrix-element likelihood fitter (Eqn. 4) and obtain the value $s_m^{data} = 0.124$, as shown in Fig. 6. For this value of $s_m^{data}$, the Feldman-Cousins band of Fig. 5 yields the limit $s_t^{data} < 0.049$, corresponding to the $ZH$ cross section limit $\sigma < 7.78 \times \sigma_{SM}$ at $m_H = 115$ GeV/$c^2$. The distribution of measured limits for background-only pseudo-experiments for $m_H = 115$ GeV/$c^2$ is shown in Fig. 7.

| $M_H$ (GeV/$c^2$) | $-2\sigma$ | $-1\sigma$ | **median** | $+1\sigma$ | $+2\sigma$ | **Observed** |
|---|---|---|---|---|---|---|
| 100 | 4.33 | 5.72 | **8.44** | 12.27 | 16.13 | 6.36 |
| 105 | 4.59 | 6.05 | **8.91** | 13.01 | 16.91 | 6.17 |
| 110 | 5.77 | 7.67 | **11.19** | 16.33 | 21.37 | 7.40 |
| 115 | 6.23 | 8.38 | **12.30** | 17.89 | 23.44 | 7.75 |
| 120 | 6.87 | 9.11 | **13.48** | 19.86 | 25.79 | 8.55 |
| 125 | 9.63 | 12.73 | **18.77** | 27.25 | 35.66 | 11.94 |
| 130 | 12.69 | 16.74 | **24.68** | 35.95 | 46.89 | 15.75 |
| 135 | 15.70 | 20.55 | **30.04** | 44.16 | 57.01 | 19.51 |
| 140 | 22.72 | 29.79 | **43.08** | 63.14 | 83.13 | 27.31 |
| 145 | 32.60 | 42.27 | **61.69** | 90.60 | 117.76 | 37.26 |
| 150 | 53.57 | 69.45 | **101.65** | 148.66 | 192.47 | 59.07 |

**TABLE IV:** Expected and observed limits for $m_H = [100, 150]$GeV/$c^2$ at 5 GeV/$c^2$ increments. Limits are in units of $\sigma_{SM}$.

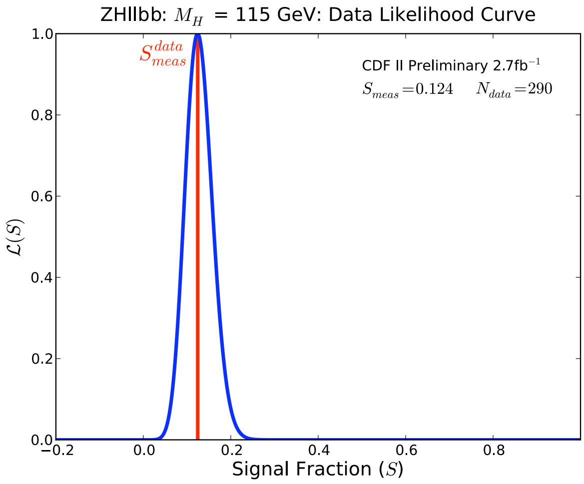


**FIG. 6:** The sample-likelihood curve for $m_H = 115$ GeV/$c^2$ for the data, with the maximum-likelihood estimate $s_m^{data}$ indicated by the red vertical line.

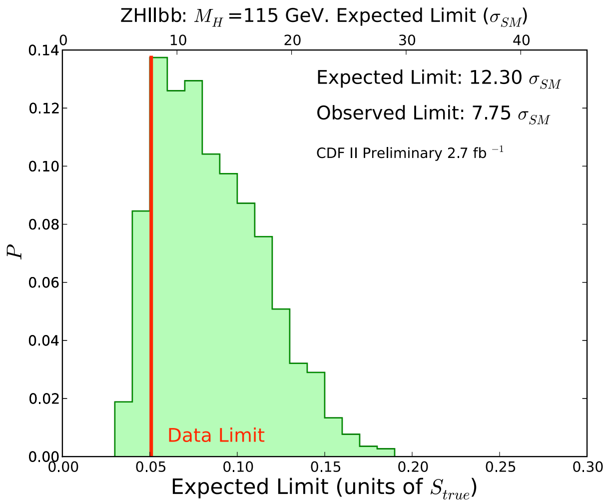


**FIG. 7:** The distribution of measured upper limits for background-only psuedo-experiments for $m_H = 115$ GeV. The measured limit in data is shown as the red line.

## VI. SUMMARY

We have searched for the SM Higgs process $ZH \rightarrow l\bar{l}b\bar{b}$ using 2.7 fb$^{-1}$ of Gen6 data, through period 17. We use standard model matrix elements to calculate probability distributions for event kinematics, which are then used as inputs to a maximum-likelihood fitter for the $ZH$ signal fraction in the data. We have used the Feldman-Cousins approach to convert the measured signal fraction into a $ZH$ cross section limit. The expected cross section limit is $12.30 \times \sigma_{SM}$ and the observed limit is $7.75 \times \sigma_{SM}$ for $m_H = 115$ GeV/$c^2$ at 95% CL.

### Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural

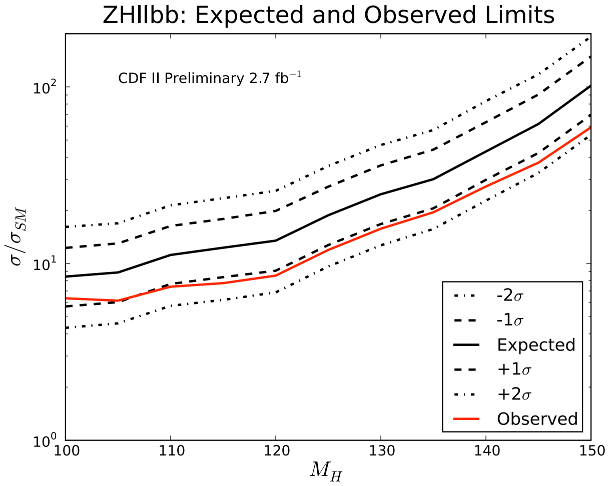


**FIG. 8:** Expected and observed limits for $m_H = [100, 150]$ GeV/$c^2$ at 5 GeV/$c^2$ increments.

Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; in part by the European Community's Human Potential Programme under contract HPRN-CT-2002-00292; and the Academy of Finland.

[1] The CDF Collaboration, CDF Note 9665
[2] D. Acosta, *et al.*, The CDF Collaboration, Phys. Rev. D **71**, 032001 (2005).
[3] T. Aaltonen, *et. al.*, The CDF Collaboration, Phys. Rev. Lett **101** 251803 (2008).
[4] D. Acosta, *et al.*, The CDF Collaboration, Phys. Rev. D **71** 052003 (2005).
[5] Gary J. Feldman and Robert D. Cousins, Phys. Rev. D57, 3873-3889 (1998).