A Search for the Higgs Boson Produced in Association with $Z \to \ell^+\ell^-$ Using the Matrix Element Method at CDF II

We present a search 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 in $p\bar{p}$ collisions at the Fermilab Tevatron. We use event probabilities based on SM matrix elements to construct a likelihood function of the Higgs content of the data sample. In a CDF data sample corresponding to an integrated luminosity of 2.7 fb$^{-1}$, we see no evidence of a Higgs boson with a mass between 100 GeV and 150 GeV. We set 95% confidence level (C.L.) upper limits on the cross-section for $ZH$ production as a function of the Higgs boson mass $m_H$; the limit is 8.2 times the SM prediction at $m_H = 115$ GeV.

PACS numbers: 14.80.Bn, 13.38.Dg, 13.85.Qk, 14.70.Hp

In the standard model (SM), the Higgs mechanism is responsible for the observed breaking of the $SU(2)_L \otimes U(1)$ symmetry [1, 2], yet the Higgs boson remains the only SM particle that has not been directly observed. Direct searches

Iowa, Iowa City, IA 52242. 2 Queen Mary, University of London, London, E1 4NS, England. 3 University of Manchester, Manchester M13 9PL, England. 4 Nagasaki Institute of Applied Science, Nagasaki, Japan. 5 University of Notre Dame, Notre Dame, IN 46556. 6 University of Oviedo, E-33007 Oviedo, Spain. 7 Texas Tech University, Lubbock, TX 79409. 8 IFIC/CSCIC-Universitat de Valencia, 46071 Valencia, Spain. 9 University of Virginia, Charlottesville, VA 22904. 10 Bergische Universitat Wuppertal, 42097 Wuppertal, Germany. 11 On leave from J. Stefan Institute, Ljubljana, Slovenia.
have set a lower limit on the SM Higgs boson mass $m_H$ of 114.4 GeV/$c^2$ at 95% C.L. [3], while precision electroweak measurements indirectly constrain its mass to $m_H = 76^{+33}_{-24}$ GeV/$c^2$ [4]. At hadron colliders the dominant production process for the SM Higgs boson is $gg \rightarrow H$ while its decays are dominated by $H \rightarrow b\bar{b}$ for $m_H < 140$ GeV/$c^2$. However, the process $gg \rightarrow H \rightarrow b\bar{b}$ is dwarfed by multi-jet background, necessitating the search for Higgs bosons produced in association with a $W$ or $Z$ boson that decays leptonically. This article reports a search for the process $p\bar{p} \rightarrow ZH \rightarrow \ell^-\ell^+bb$ ($\ell = e, \mu$) in data with an integrated luminosity of 2.7 fb$^{-1}$ collected with the CDF II detector, nearly 3 times that of the previously reported analysis [5]. The study of Higgs boson production in association with a $W/Z$ gauge boson for low Higgs boson masses is further motivated by the fact that the signal to background ratio is more favorable at the Tevatron compared to the Large Hadron Collider.

For the first time in a $ZH \rightarrow \ell^-\ell^+bb$ search, we utilize a method based on leading-order matrix element calculations [6–8] convoluted with detector resolution functions [9] that form per-event likelihoods. This method, pioneered for use in top quark mass measurements [10, 11], has been recently used in Higgs boson searches in other decay channels [12] by forming a discriminating per-event variable. We extend the technique by expressing the event likelihoods as a function of the $ZH$ signal fraction and maximizing the joint likelihood for the data sample with respect to the signal fraction.

The CDF II detector [13, 14] is an azimuthally and forward-backward symmetric apparatus designed to study $p\bar{p}$ collisions at the Fermilab Tevatron. It consists of a magnetic spectrometer surrounded by calorimeters and muon chambers. The charged particle tracking system, consisting of a silicon detector and drift chamber, is immersed in a 1.4 T magnetic field parallel to the $p$ and $\bar{p}$ beams. Calorimeters segmented in $\eta$ and $\phi$ surround the tracking system and measure the energy of particles detected within them. The electromagnetic and hadronic calorimeters are lead-scintillator and iron-scintillator sampling devices, respectively. Drift chambers located outside the central hadron calorimeters detect muons. The data used in this analysis are collected with an online selection that requires events to have a lepton with $E_T > 18$ GeV (for an electron) or $p_T > 18$ GeV/$c$ (for a muon) [14].

The event selection used in this analysis closely follows that in Ref. [5]. Candidate events are required to have a pair of oppositely charged electrons or muons with invariant mass $76 < m_{\ell\ell} < 106$ GeV/$c^2$. Candidate events are also required to have one jet with $E_T > 25$ GeV and at least one additional jet with $E_T > 15$ GeV, both within $|\eta| < 2.0$. All jet energies are corrected for non-uniformities in calorimeter response, effects from multiple $p\bar{p}$ interactions and for the hadronic energy scale of the calorimeter [15]. Candidate events are required to have at least one jet with an associated displaced secondary vertex [16] ("$b$-tags", reconstructed using tracks with hits in the silicon detector), thus enriching the $b$-quark content of the sample.

### Table I: Expected and observed numbers of events with 1 or 2 $b$-tagged jets in 2.7 fb$^{-1}$ of data. The $ZH$ expectation is shown for $m_H = 115$ GeV/$c^2$ assuming the production cross section at $\sqrt{s} = 1.96$ TeV for $q\bar{q} \rightarrow Z^* \rightarrow ZH$ to be 1.04 pb [20] and the branching ratio $B(H \rightarrow b\bar{b})$ to be 73% [21].

<table>
<thead>
<tr>
<th>Source</th>
<th>1 tag</th>
<th>$\geq 2$ tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z \rightarrow \ell^+\ell^-+$light partons</td>
<td>129.6 $\pm$ 24.0</td>
<td>5.5 $\pm$ 0.9</td>
</tr>
<tr>
<td>$Z \rightarrow \ell^+\ell^- + b\bar{b}, c\bar{c}$</td>
<td>107.2 $\pm$ 14.0</td>
<td>19.5 $\pm$ 3.4</td>
</tr>
<tr>
<td>ZZ, WZ</td>
<td>11.6 $\pm$ 1.3</td>
<td>2.9 $\pm$ 0.4</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>13.9 $\pm$ 2.0</td>
<td>7.7 $\pm$ 1.1</td>
</tr>
<tr>
<td>Mis-ID lepton</td>
<td>15.9 $\pm$ 6.5</td>
<td>0.4 $\pm$ 0.2</td>
</tr>
<tr>
<td>$ZH$</td>
<td>1.3 $\pm$ 0.2</td>
<td>0.7 $\pm$ 0.1</td>
</tr>
<tr>
<td>Total expected</td>
<td>279.5 $\pm$ 28.6</td>
<td>36.3 $\pm$ 3.7</td>
</tr>
<tr>
<td>Data</td>
<td>258</td>
<td>32</td>
</tr>
</tbody>
</table>

The backgrounds for this analysis are dominated by events with real $Z$ bosons with additional contributions from $t\bar{t}$ and events where an object, such as a jet, is mis-identified as a lepton. We model the backgrounds with events generated with leading-order event generators, normalized to next-to-leading order cross-sections and simulated with a GEANT-based description of the CDF II detector [9]. $Z$+light-flavor jet contributions are modeled with the ALPGEN [17] simulation code matched with PYTHIA in the MLM scheme [17] for the hadronization and fragmentation. Heavy flavor contributions from $Z + b\bar{b}$ and $Z + c\bar{c}$ are modeled separately with ALPGEN and combined with the light-flavor jet samples. The $WZ$, $ZZ$ and $t\bar{t}$ processes are modeled using PYTHIA [18]. Events where a jet is mis-identified as a charged lepton are modeled using jet-enriched data samples [5, 19]. We model the kinematics of $ZH \rightarrow \ell^+\ell^-b\bar{b}$ events using PYTHIA for $m_H$ ranging from 100 GeV/$c^2$ to 150 GeV/$c^2$. The signal and background contributions expected in 2.7 fb$^{-1}$ and the number of observed events are given in Table I.
where \( M_{ZH} \) is the leading-order matrix element for the process \( q\bar{q} \rightarrow ZH \rightarrow t^+t^-b\bar{b} \) evaluated for a pair of incoming partons \( q \) and outgoing particles \( p, W(p_j, x_i) \) are transfer functions [22] linking the outgoing particle momenta \( p_j \) to measured quantities \( x_i \) and the \( f_{PDF} \) are parton density functions of the incoming partons. The factor \( 1/\sigma(m_H) \) ensures that the probability density satisfies the normalization condition, \( \int dx_i P_{ZH}(x_i|m_H) = 1. \)

The sample likelihood \( \mathcal{L} \) is obtained by taking the product over all events \( i \) in the sample

\[
\mathcal{L}(s|m_H) = \prod_i L(s, x_i|m_H).
\]

We enhance our statistical sensitivity by exploiting the expected difference in the rate of signal and background events with two \( b \)-tagged jets. We replace \( P_{ZH}(x_i|m_H) \) by \( P_{ZH}(x_i, n|m_H) \equiv P_{ZH}(x_i|m_H) \cdot P_{ZH}(n|m_H) \) and \( P_b(x_i) \) by \( P_b(x_i, n) \equiv P_b(x_i) \cdot P_b(n) \), where \( P_{ZH}(n|m_H) \) denotes the probability of tagging signal (background) events with \( n \) tags. Table II shows the expected tagging rates for simulated signal and background event samples.

The measured signal fraction \( S_{meas} \) is the value of \( s \) which maximizes \( \mathcal{L}(s|m_H) \). Using Eq. (1), we can define a per-event discriminant \( \Delta_i \equiv \partial \ln L/\partial s = (P_{ZH} - P_b)/L \) which increases (decreases) for more signal-like (background-like) events. The maximum-likelihood estimator for the measured signal fraction \( S_{meas} \) corresponds to \( \Sigma_i \Delta_i = S_{meas} = 0. \) The distribution of \( \tan^{-1}(\Delta_i) \) for \( s = S_{meas} \) for simulated events and data is shown in Fig. 1.

The dominant backgrounds in our data sample are due to \( Z+\text{jets}, \bar{t}t \) and \( ZZ \) processes, in the expected proportions denoted by \( \lambda_{Zjj}, \lambda_{tt} \) and \( \lambda_{ZZ} \) respectively. The background probability in Eq. (1) is given by

\[
P_b(x_i, n) = \lambda_{Zjj} P_{Zjj}(x_i, n) + \lambda_{tt} P_{tt}(x_i, n) + \lambda_{ZZ} P_{ZZ}(x_i, n),
\]

where \( P_{Zjj}(x_i, n), P_{tt}(x_i, n) \) and \( P_{ZZ}(x_i, n) \) are the respective probability densities (normalized to unit integral) for the \( Z+\text{jets}, \bar{t}t \), and \( ZZ \) background processes with \( n \) tags. Normalization of \( P_b \) is ensured by requiring \( \lambda_{Zjj} + \lambda_{tt} + \lambda_{ZZ} = 1. \)

We construct confidence intervals [23] for the test statistic \( R = \mathcal{L}(s_{\text{meas}}|S_{true})/\mathcal{L}(s_{\text{meas}}|S_{true}) \) by performing simulated experiments with the expected proportions of background and varying the amounts of signal, such that \( S_{true} \) is the true (input) signal fraction in the simulated experiment. \( S_{\text{best}}^{s_{true}} \) is the input signal fraction that has the highest likelihood for a given measured signal fraction, \( S_{meas} \). \( \mathcal{L}(s_{\text{meas}}|S_{true}) \) is given by Eq. (3) for the simulated experiment with the chosen value of \( S_{true} \) and \( m_H \). Since we are measuring the fractional signal content in the data sample, the number of events in each simulated experiment is held fixed at the value of 290 events observed in the data.

The methodology from Ref. [23] is used to construct confidence intervals in \( S_{meas} \) for each chosen value of \( S_{true} \) and \( m_H \). This method removes any bias resulting from imperfections in our modeling by relating \( S_{meas} \) to \( S_{true} \). The confidence intervals in \( S_{meas} \) obtained for \( m_H = 115 \text{ GeV}/c^2 \) and \( 0 < S_{true} < 0.25 \) are shown in Fig. 2. For a given value of \( S_{true} \) obtained from the data (or from an independent simulated experiment to evaluate the \textit{a priori} expectation), we extract the range of \( S_{true} \) for which the confidence intervals contain this value of \( S_{true} \). A feature of this method is that the resulting range of \( S_{true} \) can be quoted as an upper limit on \( S_{true} \) (if the lower bound is zero) or as a two-sided measurement of \( S_{true} \). As Fig. 2 shows, we obtain an upper limit on \( S_{true} \) given the data, which we convert to the equivalent upper limit on the signal cross section. This procedure is repeated for the range of Higgs boson masses 100 \( \leq m_H \leq 150 \text{ GeV}/c^2 \).

We evaluate systematic uncertainties by varying process rates and kinematic distributions in our simulated experiments. We apply a rate uncertainty of 40% for \( Z \) boson events and of 20% for diboson and \( t\bar{t} \) events. The uncertainty on the rate of heavy flavor production in association with a gauge boson is based on comparisons of data with theoretical predictions [19]. The uncertainty on the diboson and \( t\bar{t} \) contribution

\[
\begin{array}{c|c|c}
\hline
\text{Source} & \text{\( P(n = 1) \)} & \text{\( P(n \geq 2) \)} \\
\hline
\text{\( Z \rightarrow t^+t^-+\text{jets} \)} & 0.91 & 0.09 \\
\text{\( WZ, ZZ \)} & 0.80 & 0.20 \\
\text{\( \bar{t}t \)} & 0.74 & 0.26 \\
\text{\( ZH (m_H = 100 \text{ GeV}/c^2) \)} & 0.67 & 0.33 \\
\text{\( ZH (m_H = 125 \text{ GeV}/c^2) \)} & 0.65 & 0.35 \\
\text{\( ZH (m_H = 150 \text{ GeV}/c^2) \)} & 0.63 & 0.37 \\
\hline
\end{array}
\]
This result compares the observed Higgs boson masses against the SM prediction. The limit at 95% C.L. is shown in the simulated experiments. The signal cross section is evaluated in a range of true signal fraction values (along the y-axis) and the jet energy scale in simulated signal events. Uncertainties in the cross sections, selection efficiencies and the top quark mass are included. A rate uncertainty of 6% due to the luminosity uncertainty is applied to all events. The per-jet uncertainty on the jet energy scale in simulated signal and background events is weakly sensitive to uncertainties in the expected non-Higgs backgrounds. We set 95% C.L. upper limits on the cross section of this process for a range of Higgs boson masses between 100 GeV/c^2 and 200 GeV/c^2, and include statistical and systematic uncertainties.[15]. Our analysis is weakly sensitive to uncertainties in the expected total number of events passing our selection, since it relies only on the shapes of measured distributions. Uncertainties in the shapes of kinematic distributions are propagated by varying the amount of QCD radiation in simulated signal events and the jet energy scale in simulated signal and background events within their respective uncertainties.[15]. The intervals shown here are computed for a Higgs boson mass of m_H = 115 GeV/c^2, and include statistical and systematic uncertainties. The vertical dashed line indicates the value of S_{true} obtained from the data.

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 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 World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Instituto Nacional de Física Nucleares y Física de Partículas/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

We evaluate confidence intervals for a range of Higgs boson masses between 100 GeV/c^2 and 200 GeV/c^2. We evaluate a priori 95% C.L. upper limits on the cross section for the process p̅p → ZH → ℓ⁺ℓ⁻b̅b. We express these limits as a ratio with respect to the SM prediction. These expected limits along with those observed in the data are shown in Table III.

In conclusion, we have performed a search for the SM Higgs boson decaying to bb produced in association with a Z boson. This is the first analysis performed in this channel with a matrix element method. The data show no excess over expected non-Higgs backgrounds. We set 95% C.L. upper limits on the cross section of this process for a range of Higgs boson masses. The limit at m_H = 115 GeV/c^2 is 8.2 times greater than the SM prediction. This result improves by a factor of 2 over the previously published result in this channel.[5]. We are exploring further improvements in this technique by separating the leading-order and next-to-leading order contributions to the signal and backgrounds, as well as the use of matrix-element-based probabilities in conjunction with other multivariate discriminants.

We include the uncertainties in the cross sections, selection efficiencies and the top quark mass.[5]. A rate uncertainty of 50% is applied for mis-identified lepton events due to the uncertainty on the lepton misidentification probability.[5]. A rate uncertainty of 6% due to the luminosity uncertainty is applied to all events. The per-jet uncertainty on the b-tagging efficiency is 8% for events with b partons, 16% for events with c partons and 13% for events with no heavy flavor.[5]. Our analysis is weakly sensitive to uncertainties in the expected total number of events passing our selection, since it relies only on the shapes of measured distributions. Uncertainties in the shapes of kinematic distributions are propagated by varying the amount of QCD radiation in simulated signal events and the jet energy scale in simulated signal and background events within their respective uncertainties.[15].

We evaluate confidence intervals for a range of Higgs boson masses between 100 GeV/c^2 and 200 GeV/c^2. We evaluate a priori 95% C.L. upper limits on the cross section for the process p̅p → ZH → ℓ⁺ℓ⁻b̅b. We express these limits as a ratio with respect to the SM prediction. These expected limits along with those observed in the data are shown in Table III.

We include the uncertainties in the cross sections, selection efficiencies and the top quark mass.[5]. A rate uncertainty of 50% is applied for mis-identified lepton events due to the uncertainty on the lepton misidentification probability.[5]. A rate uncertainty of 6% due to the luminosity uncertainty is applied to all events. The per-jet uncertainty on the b-tagging efficiency is 8% for events with b partons, 16% for events with c partons and 13% for events with no heavy flavor.[5]. Our analysis is weakly sensitive to uncertainties in the expected total number of events passing our selection, since it relies only on the shapes of measured distributions. Uncertainties in the shapes of kinematic distributions are propagated by varying the amount of QCD radiation in simulated signal events and the jet energy scale in simulated signal and background events within their respective uncertainties.[15].

We evaluate confidence intervals for a range of Higgs boson masses between 100 GeV/c^2 and 200 GeV/c^2. We evaluate a priori 95% C.L. upper limits on the cross section for the process p̅p → ZH → ℓ⁺ℓ⁻b̅b. We express these limits as a ratio with respect to the SM prediction. These expected limits along with those observed in the data are shown in Table III.

In conclusion, we have performed a search for the SM Higgs boson decaying to bb produced in association with a Z boson. This is the first analysis performed in this channel with a matrix element method. The data show no excess over expected non-Higgs backgrounds. We set 95% C.L. upper limits on the cross section of this process for a range of Higgs boson masses. The limit at m_H = 115 GeV/c^2 is 8.2 times greater than the SM prediction. This result improves by a factor of 2 over the previously published result in this channel.[5]. We are exploring further improvements in this technique by separating the leading-order and next-to-leading order contributions to the signal and backgrounds, as well as the use of matrix-element-based probabilities in conjunction with other multivariate discriminants.

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 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 World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleares et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

TABLE III: Upper limits at 95% C.L. on the $ZH \rightarrow ℓ⁺ℓ⁻b̅b$ cross section, shown as a ratio to the SM cross section. The column labelled “Expected” shows the median of the limits obtained from simulated experiments containing no signal, and the columns labelled “±1σ” show the range containing 68% of the expected limits.

<table>
<thead>
<tr>
<th>$m_H$ [GeV/c²]</th>
<th>−1σ</th>
<th>Expected</th>
<th>+1σ</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>6.0</td>
<td>8.7</td>
<td>12.4</td>
<td>7.0</td>
</tr>
<tr>
<td>105</td>
<td>6.0</td>
<td>8.7</td>
<td>12.9</td>
<td>6.5</td>
</tr>
<tr>
<td>110</td>
<td>7.5</td>
<td>11.3</td>
<td>16.8</td>
<td>7.6</td>
</tr>
<tr>
<td>115</td>
<td>8.3</td>
<td>12.1</td>
<td>18.2</td>
<td>8.2</td>
</tr>
<tr>
<td>120</td>
<td>9.3</td>
<td>13.5</td>
<td>20.0</td>
<td>9.0</td>
</tr>
<tr>
<td>125</td>
<td>13.2</td>
<td>18.3</td>
<td>27.1</td>
<td>13.2</td>
</tr>
<tr>
<td>130</td>
<td>17.1</td>
<td>24.2</td>
<td>35.7</td>
<td>17.7</td>
</tr>
<tr>
<td>135</td>
<td>21.8</td>
<td>31.0</td>
<td>44.8</td>
<td>22.9</td>
</tr>
<tr>
<td>140</td>
<td>31.0</td>
<td>44.3</td>
<td>65.4</td>
<td>32.0</td>
</tr>
<tr>
<td>145</td>
<td>42.8</td>
<td>61.6</td>
<td>89.9</td>
<td>43.1</td>
</tr>
<tr>
<td>150</td>
<td>73.7</td>
<td>104</td>
<td>153</td>
<td>71.3</td>
</tr>
</tbody>
</table>

(1999).

[14] CDF uses a cylindrical coordinate system with the $z$ axis along the proton beam axis. Pseudorapidity is $\eta \equiv -\ln(\tan(\theta/2))$, where $\theta$ is the polar angle, and $\phi$ is the azimuthal angle relative to the proton beam direction, while $p_T = |p| \sin \theta$, $E_T = E \sin \theta$.