Combined Search for the Standard Model Higgs Boson Decaying to $b\bar{b}$ Using the D0 Run II Data Set

We present the results of the combination of searches for the standard model Higgs boson produced in association with a W or Z boson and decaying into $b\bar{b}$ using the data sample collected with the D0 detector in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV at the Fermilab Tevatron Collider. We derive 95% C.L. upper limits on the Higgs boson cross section relative to the standard model prediction in the mass range $100$ GeV < $M_H$ < 125 GeV and have recently published results [10] confirming these excesses at the level of 5 standard deviations (s.d.), driven by searches for $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^{(\ell\ell')} \rightarrow \ell^+\ell^-\ell'^+\ell'^-$, where $\ell$ and $\ell'$ denote an electron or muon. These searches primarily exploit the gluon-gluon fusion production mechanism for the Higgs boson, $gg \rightarrow H$, mediated by a top-quark loop, while $H \rightarrow \gamma\gamma$ searches are also sensitive to vector $(V = W, Z)$ boson fusion, $q\bar{q} \rightarrow Hq\bar{q}$. In the allowed mass range, the Tevatron experiments are particularly sensitive to the SM Higgs boson produced in association with a vector boson, $VH$, and the Higgs boson decaying into $b\bar{b}$, the primary decay mode for a Higgs boson with $M_H < 135$ GeV. Searches at both hadron colliders have a high degree of complementarity, with the main search channels at the LHC being particularly sensitive to the Higgs boson mass and couplings to vector bosons, while searches at the Tevatron provide information on the Higgs boson coupling to $b$ quarks.

This Letter describes the combination of searches for $VH, H \rightarrow b\bar{b}$ production at the D0 experiment using the sample of $p\bar{p}$ collision data at $\sqrt{s} = 1.96$ TeV collected during run II of the Fermilab Tevatron Collider. These
searches are focused on leptonic \( W \) and \( Z \) boson decays that allow us to efficiently suppress the large multijet background present at a hadron collider and are restricted to the mass range \( 100 \text{ GeV} \leq M_H \leq 150 \text{ GeV} \). Therefore, the signal processes being targeted are \( WH \rightarrow \ell v b \bar{b} \) \cite{11}, \( ZH \rightarrow v b \bar{b} \) \cite{12}, and \( ZH \rightarrow \ell^+ \ell^- b \bar{b} \) \cite{13}. A similar combination of searches in the \( H \rightarrow b \bar{b} \) decay mode has recently been reported by the CDF Collaboration \cite{14} and previously by the ATLAS \cite{15}, CMS \cite{16}, and LEP \cite{3} Collaborations.

The D0 detector is described elsewhere \cite{17}. Details on the reconstruction and identification criteria for the physics objects used in these searches [electrons, muons, jets, and missing transverse energy (\( E_T \))] can be found elsewhere \cite{11–13,18}. Jets are identified as consistent with the fragmentation of a \( b \) quark (\( b \)-tagged) by a multivariate algorithm \cite{19} combining information from the impact parameter of displaced tracks and the topological properties of secondary vertices reconstructed in the jet.

The main backgrounds affecting these searches originate from \( W/Z + \) heavy-flavor jets (jets initiated by \( b \) and \( c \) quarks) and from top-quark pair (\( t \bar{t} \)) production. Smaller contributions arise from \( W/Z + \) light-flavor jets, single top-quark, diboson (\( WW, WZ, ZZ \)), and multijet production. Multijet events contribute to the selected samples via the misidentification of a jet or a photon as an electron, the presence of a nonprompt lepton from a semileptonic \( b \)- or \( c \)-hadron decay (\( WH \rightarrow \ell v b \bar{b} \) and \( ZH \rightarrow \ell^+ \ell^- b \bar{b} \) analyses), or jet energy mismeasurements resulting in apparent large \( E_T \) (\( ZH \rightarrow v b \bar{b} \) analysis). In all instances, the normalization and kinematic distributions of multijet events are estimated via data-driven methods. The remaining backgrounds, as well as the signal, are estimated with Monte Carlo simulation. Samples of \( W/Z + \) jets and \( t \bar{t} \) events are generated by using the \textsc{alpgen} \cite{20} tree-level matrix element generator, while samples of single top-quark and diboson events are generated by using the \textsc{singletop} \cite{21} and \textsc{pythia} \cite{22} leading-order (LO) generators, respectively. These samples are normalized to next-to-next-to-LO (NNLO) \cite{23}, approximate NNLO \cite{24,25}, and next-to-LO \cite{26} theoretical cross sections. Samples of \( WH \) and \( ZH \) signal events are generated by using the \textsc{pythia} generator for a range of masses, \( 100 \text{ GeV} \leq M_H \leq 150 \text{ GeV} \), in steps of 5 GeV and are normalized to the most recent theoretical predictions \cite{27–29}. All Monte Carlo samples are generated by using the \textsc{cteq6l1} parton distribution function set \cite{30} and processed through \textsc{pythia} to model parton showering and fragmentation. Signal and backgrounds samples are processed by a \textsc{geant3}-based \cite{31} simulation of the D0 detector and reconstructed by using the same algorithms applied to the collider data. Simulated events are corrected so that the object identification efficiencies, energy scales, and energy resolutions match those determined in data control samples. More details on the simulation and normalization of the signal and background samples can be found elsewhere \cite{11–13}.

In the case of the \( ZH \rightarrow v b \bar{b} \) analysis, the data were collected by using triggers requiring jets plus \( E_T \) and correspond to an integrated luminosity of 9.5 \( \text{fb}^{-1} \) \cite{32}. The \( ZH \rightarrow \ell^+ \ell^- b \bar{b} \) and \( WH \rightarrow \ell v b \bar{b} \) analyses use a logical OR of triggers dominated by single lepton, dilepton, lepton-plus-jets, and jet-plus \( E_T \) triggers, resulting in an integrated luminosity of 9.7 \( \text{fb}^{-1} \). The analyses select nonoverlapping subsets of data via different requirements on lepton multiplicity: (i) exactly two opposite-charge leptons (\( ZH \rightarrow \ell^+ \ell^- b \bar{b} \)), (ii) exactly one charged lepton and large \( E_T \) (\( WH \rightarrow \ell v b \bar{b} \)), and (iii) exactly zero charged leptons and large \( E_T \) (\( ZH \rightarrow v b \bar{b} \)). A significant fraction of signal events selected by the \( ZH \rightarrow v b \bar{b} \) analysis originate from \( WH \) production, where the charged lepton is not reconstructed. In addition, events are required to have two or three reconstructed jets, with the exception of the \( ZH \rightarrow v b \bar{b} \) analysis, which is restricted to events with exactly two jets. The signal-to-background ratio is significantly enhanced by requiring one or two \( b \)-tagged jets in an event. The sensitivity of the searches is maximized by categorizing events into different analysis subchannels depending on the flavor and quality of the charged leptons, jet multiplicity, \( b \)-tagged jet multiplicity, and \( b \)-tagged jet quality. The primary discriminating variable between the \( VH \) signal and the backgrounds is the dijet invariant mass, for which the signal shows a distinct resonant structure; however, by combining this variable with several other kinematic variables via a multivariate approach, the sensitivity of the searches is improved by approximately 25%. Therefore, the final observable for each of the subchannels in the different searches is a one-dimensional multivariate discriminant optimized for each hypothesized \( M_H \) value.

We interpret the result of the searches via the \( \text{CL}_s \) method \cite{33,34}, which employs a log-likelihood ratio \( \text{LLR} = -2 \ln(L_{s+b}/L_b) \) as a test statistic, where \( L_{s+b} \) (\( L_b \)) is a Poisson likelihood to observe the data under the signal-plus-background (background-only) hypothesis. Separate channels are combined by summing LLR values over all bins, thus maintaining the individual channel sensitivities. The per-bin signal and background predictions are parameterized in terms of nuisance parameters that describe the effect of systematic uncertainties. The impact of systematic uncertainties on the search sensitivity is reduced by maximizing both likelihood functions \( L_{s+b} \) and \( L_b \), with respect to these nuisance parameters, subject to Gaussian constraints of their prior values. \( \text{CL}_s \) is defined as the ratio of the confidence levels for the signal-plus-background (\( \text{CL}_{s+b} \)) and background-only (\( \text{CL}_b \)) hypotheses, which are each evaluated by integrating the corresponding LLR distributions populated by simulating outcomes via Poisson statistics. Systematic uncertainties are incorporated via Gaussian fluctuations on the expected number of signal and background events per bin, taking
into account correlations across processes and channels [35]. Signal cross sections resulting in $\text{CL}_s < 0.05$ are excluded at the 95% C.L.

The systematic uncertainties differ between analyses, but we summarize here the largest contributions. We account for the impact of these uncertainties both on the integrated signal and background yields and on the shapes of the final discriminants where relevant. The $ZH \rightarrow \nu \nu b \bar{b}$ and $WH \rightarrow \ell^+ \ell^- b \bar{b}$ analyses carry a correlated uncertainty on the integrated luminosity of 6.1% [32]. The $ZH \rightarrow e^+ e^- b \bar{b}$ analysis normalizes the predictions using the peak from $Z \rightarrow e^+ e^-$ decays from data and the corresponding NNLO cross section [23]. The $b$-tagging efficiency has an uncertainty of $\approx 1\%–15\%$, depending on the sample and $b$-tagging criteria. The uncertainty due to acceptance and energy measurement of jets is typically around 7%. Uncertainties due to acceptance and energy measurement of leptons range from 1% to 9%, depending on the final state. A significant source of uncertainty comes from the $V +$ jets background cross sections, which have uncertainties of 4%–10% for light flavor jets and $\approx 22\%$ for heavy flavor jets. These account for both the uncertainty on the theoretical cross section calculations and the uncertainties on the higher-order correction factors. The uncertainty on the expected multijet background is dominated by the statistics of the data sample from which it is estimated and is considered separately from the other cross section uncertainties. All analyses take into account the uncertainties on the theoretical production cross sections for the different signal processes due to parton distribution function and scale choice. In addition, analyses incorporate differential uncertainties on the dominant backgrounds to allow for potential variations of the final discriminants due to generator and background modeling uncertainties. The total impact of systematic uncertainties on the combined sensitivity is $\approx 20\%$.

To confirm the ability of these analyses to measure a signal and to validate the background modeling, we perform a measurement of the $VZ$ production cross section in the same final states. The only difference from the Higgs boson search is to use SM $WZ$ and $ZZ$ production as the signal instead of $WH$ and $ZH$, while the rest of the SM processes, including $WW$ production, are treated as backgrounds. Multivariate discriminants using the same input variables as in the Higgs boson searches are trained to separate the $VZ$ signal from the backgrounds, and the resulting distributions are fit to determine the $VZ$ cross section. The combination of all three analyses yields $\sigma(VZ) = 3.3 \pm 1.4$ pb, consistent with the SM prediction of $4.4 \pm 0.3$ pb [26]. The observed (expected) significance of the measured excess is $2.5$ ($3.4$) s.d.

The statistical analysis makes use of simultaneous fits to the individual final discriminants, but it is useful for presentation purposes to collect all of the inputs into a single distribution. This is done by reordering the bins from the input distributions according to their signal-to-background ratios ($s/b$), so that bins with similar $\log_{10}(s/b)$ are combined. Figure 1 shows this distribution for the $VZ$ cross section measurement and for the Higgs boson search with $M_H = 125$ GeV after subtracting the expected background from the data. The subtracted background corresponds to the maximum-likelihood fit of the nuisance parameters to the data, and the posterior uncertainty from that fit is also shown in the plot.

We derive limits on SM Higgs boson production $\sigma(VH) \times \text{BR}(H \rightarrow b\bar{b})$, with $\text{BR}$ the branching fraction, for Higgs boson masses in the range $100$ GeV $\leq M_H \leq 150$ GeV in steps of 5 GeV. We assume the relative contributions of the different production and decay modes as given by the SM prediction. We present our results in terms of the ratio of 95% C.L. upper cross section limits to the SM predicted cross section. The SM prediction for Higgs boson
production would therefore be considered excluded at 95% C.L. when this limit ratio falls below unity. Figure 2(a) shows the combined expected and observed 95% C.L. cross section limits as a ratio to the SM prediction as a function of $M_H$. These results are also summarized in Table I. The LLR distributions for the combination are shown in Fig. 2(b). Although consistent with the background-only hypothesis for $M_H < 115$ GeV, the observed LLR exhibits a signal-like excess at the level of $1\text{–}1.7$ s.d. for the mass range $120$ GeV $\leq M_H \leq 145$ GeV.

To understand the compatibility of this excess with the hypothesis of a SM Higgs boson, we obtain the best-fit cross section for the Higgs boson signal relative to the SM cross section as a function of $M_H$. This value is obtained by performing a maximum-likelihood fit over all search channels simultaneously, allowing the fit to vary all nuisance parameters within their priors and with the Higgs boson cross section as a free parameter. Figure 3 shows the measured $\sigma(VH) \times BR(H \to b\bar{b})$ as a function of $M_H$, including its $\pm 1$ s.d. uncertainty band, and compared with the SM prediction. At a mass of 125 GeV, the best-fit cross section is $\sigma(VH) \times BR(H \to b\bar{b}) = 140^{+140}_{-130}$ pb, which is $1.2^{+1.2}_{-1.1}$ times the SM prediction.

The significance of the data excess above the background prediction is estimated by computing the $p$ value under the background-only hypothesis using $R^{\text{obs}}$ as the test statistic for each value of $M_H$. This $p$ value represents the probability to have a value of $R^{\text{obs}}$ as large or larger than that observed in the data due to a background fluctuation. The smallest $p$ value is obtained at a mass of 135 GeV and corresponds to a significance of $1.7$ s.d. above the background-only prediction. This significance does not take into account the look-elsewhere effect [36], which accounts for the possibility of a background fluctuation in the local $p$ value anywhere in the tested mass range. By taking into account existing limits on $M_H$ in the $b\bar{b}$ decay mode [3], the search region becomes $115$ GeV $\leq M_H \leq 150$ GeV. Given the expected mass resolution of these searches of $\approx 16\%$, this translates into a look-elsewhere-effect factor of $\approx 1.6$ for a global significance of $1.5$ s.d. Also taking into account the existing SM Higgs boson exclusions from the LHC [8,9] experiments, there is no look-elsewhere effect, and we find an excess at $M_H = 125$ GeV with a significance of $1.1$ s.d.

In summary, we have presented a combination of searches for the SM Higgs boson produced in association with a vector boson and decaying into $b\bar{b}$, using the data sample collected with the D0 detector in run II of the Fermilab Tevatron Collider. We achieve a sensitivity that is competitive with other searches in this final state [14–16], deriving 95% C.L. upper limits on the Higgs boson cross section relative to the SM prediction in the mass range $100$ GeV $\leq M_H \leq 150$ GeV and excluding Higgs bosons with masses smaller than 102 GeV at the

![Graph showing 95% C.L. cross section upper limit ratios versus $M_H$.](image)

**TABLE I.** Expected (median) and observed 95% C.L. cross section upper limit ratios for the combined $VH, H \to b\bar{b}$ analyses over the $100$ GeV $\leq M_H \leq 150$ GeV mass range.

<table>
<thead>
<tr>
<th>$M_H$ (GeV)</th>
<th>100</th>
<th>105</th>
<th>110</th>
<th>115</th>
<th>120</th>
<th>125</th>
<th>130</th>
<th>135</th>
<th>140</th>
<th>145</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected</td>
<td>1.2</td>
<td>1.3</td>
<td>1.4</td>
<td>1.6</td>
<td>1.9</td>
<td>2.3</td>
<td>2.9</td>
<td>3.8</td>
<td>5.3</td>
<td>7.8</td>
<td>12</td>
</tr>
<tr>
<td>Observed</td>
<td>0.94</td>
<td>1.1</td>
<td>1.2</td>
<td>1.9</td>
<td>2.6</td>
<td>3.2</td>
<td>4.3</td>
<td>6.5</td>
<td>8.0</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>

FIG. 2 (color online). (a) The 95% C.L. cross section upper limit ratios versus $M_H$, and (b) LLR distribution versus $M_H$, for the combined $VH, H \to b\bar{b}$ analyses. The solid lines represent the observed values in the data. The short-dashed black (red) lines represent the median expected values under the background-only (signal-plus-background) hypothesis at each mass. The long-dashed blue lines show the expected outcome from injecting a SM Higgs boson signal with $M_H = 125$ GeV. The green and yellow shaded bands correspond to the regions enclosing 1 and 2 s.d. variations about the median expected values under the background-only hypothesis, respectively.
95% C.L. In the mass range 120 GeV ≤ M_H ≤ 145 GeV, the data exhibit an excess above the background prediction with a global significance of 1.5 s.d. and a magnitude consistent with that expected for the SM Higgs boson.

We thank the staffs at Fermilab and collaborating institutions and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); MON, NRF (Korea); FOM (Netherlands); STFC and the Royal Society (United Kingdom); MSMT and GACR (Czech Republic); BMBF and DFG (Germany); SFI (Ireland); and CAS and CNSF (China).

**Visitor from Augustana College, Sioux Falls, SD, USA.**
**Visitor from The University of Liverpool, Liverpool, United Kingdom.**
**Visitor from UPIITA-IPN, Mexico City, Mexico.**
**Visitor from DESY, Hamburg, Germany.**
**Visitor from SLAC, Menlo Park, CA, USA.**
**Visitor from University College London, London, United Kingdom.**
**Visitor from Centro de Investigacion en Computacion—IPN, Mexico City, Mexico.**
**Visitor from ECFM, Universidad Autonoma de Sinaloa, Culiacan, Mexico.**
**Visitor from Universidade Estadual Paulista, Sao Paulo, Brazil.**


[35] Sources of uncertainty common to multiple channels (e.g., $b$-tagging, jet energy scale and resolution, and theoretical uncertainties) are treated as fully correlated between those channels.