Search for the Standard Model Higgs Boson in Associated WH Production in 9.7 fb^{-1} of pp[over ¯] Collisions with the D0 Detector

V. M. Abazov et al. (D0 Collaboration)

Phys. Rev. Lett. 109, 121804 — Published 20 September 2012

DOI: 10.1103/PhysRevLett.109.121804
Search for the standard model Higgs boson in associated $WH$ production in $9.7 \text{ fb}^{-1}$ of $pp$ collisions with the D0 detector

We present a search for the standard model Higgs boson in final states with a charged lepton (electron or muon), missing transverse energy, and two or three jets, at least one of which is identified as a b-quark jet. The search is primarily sensitive to $WH \rightarrow ℓνb\bar{b}$ final states, which is the dominant decay mode in this mass range. The experiments now exclude $111 < M_H < 129$ GeV [11, 12]. Both experiments have observed a resonance consistent with SM Higgs production at $M_H \approx 125$ GeV, primarily in the $γγ$ and $ZZ$ final states, above the 5 s.d. level [11, 12]. Demonstrating that the observed resonance is due to SM Higgs boson production requires also observing it in the $bb$ final state, which is the dominant decay mode in this mass range.

The dominant Higgs boson production process at the Tevatron Collider is gluon-gluon fusion. The associated production measurements of other electroweak parameters constrain $M_H$ to be less than 152 GeV [5–7]. The region $147 < M_H < 179$ GeV is excluded by the combined analysis of the CDF and D0 Collaborations [8]. The ATLAS and CMS Collaborations at the CERN Large Hadron Collider (LHC) have excluded much of the allowed mass range and reported excesses at the 2–3 standard deviation (s.d.) level for $M_H \approx 125$ GeV [9, 10]. The experiments now exclude $111 < M_H < 129$ GeV [11, 12]. Both experiments have observed a resonance consistent with SM Higgs production at $M_H \approx 125$ GeV, primarily in the $γγ$ and $ZZ$ final states, above the 5 s.d. level [11, 12]. Demonstrating that the observed resonance is due to SM Higgs boson production requires also observing it in the $bb$ final state, which is the dominant decay mode in this mass range.
production of a Higgs boson with a weak boson occurs at a rate about three times lower than the gluon-gluon fusion production process, but is of particular importance in Higgs boson searches. At masses below $M_H \approx 135$ GeV, $H \rightarrow b\bar{b}$ decays dominate but are difficult to distinguish from background when the Higgs boson is produced by gluon-gluon fusion. Instead, associated production of a Higgs boson and a $W$ boson is one of the most sensitive search channels at the Tevatron.

This Letter presents a search based on events with one charged lepton ($\ell = e$ or $\mu$), an imbalance in transverse energy ($\mathbf{E}_T$) that arises from the neutrino in the $W \rightarrow \ell\nu$ decay, and two or three jets, where one or more of these jets is selected as a candidate $b$ quark ("$b$-tagged") jet. The search is also sensitive to $ZH$ production when one of the charged leptons from the $Z \rightarrow \ell^+\ell^-$ decay is not identified. The analysis is optimized by subdividing into channels with different background compositions and signal to background ratios based on lepton flavor, jet multiplicity, and the number and quality of candidate $b$-quark jets.

Several searches for $WH \rightarrow \ell b\bar{b}$ production have already been reported at a $p\bar{p}$ center-of-mass energy of $\sqrt{s} = 1.96$ TeV, most recently by the CDF Collaboration [13]. Previous searches [14–18] by the D0 Collaboration use subsamples of the data presented in this Letter with integrated luminosities up to 5.3 fb$^{-1}$. We present an updated search using a multivariate approach with a full dataset which, after imposing data quality requirements, corresponds to an integrated luminosity of 9.7 fb$^{-1}$.

This analysis uses most of the major components of the D0 detector, described in detail in Refs. [19–22]. Events in the electron channel are selected with triggers requiring an electromagnetic (EM) object in the calorimeter or an EM object with additional jets. In the muon channel we use a mixture of single muon, muon plus jet, $\mathbf{E}_T$ plus jet, and multijet triggers. We correct simulated events for trigger efficiency using a method similar to that described in Ref. [18].

Several SM processes produce or can mimic a final state with a charged lepton, $\mathbf{E}_T$, and jets, including diboson ($WW$, $WZ$, and $ZZ$), $V+$jets ($V = W$ or $Z$), $tt$, single top quark, and multijet (MJ) production. We estimate the MJ background from data and other backgrounds from simulation. The $V+$jets and $t\bar{t}$ samples are simulated with the ALPGEN [23] Monte Carlo (MC) generator interfaced to PYTHIA [24] for parton showering and hadronization. ALPGEN samples are produced using the MLM parton-jet matching prescription [23]. The $V+$jets samples contain $V+jj$ (where $j = u, d, s$ or $g$) and $V+cj$ (together denoted as "$V+$light-flavor") processes, and $V+b\bar{b}$ and $V+c\bar{c}$ (together denoted as "$V+$heavy-flavor"), generated separately from $V+$light-flavor. PYTHIA is used to simulate the production of dibosons ($WW$, $WZ$, and $ZZ$), and all signal processes. Single top quark events are generated with the SINGLETOP event generator [25, 26] using PYTHIA for parton evolution and hadronization. Simulation of background and signal processes uses the CTEQ6L1 [27, 28] leading-order (LO) parton distribution functions. Events are processed through a full D0 detector simulation based on GEANT [29]. To account for multiple $p\bar{p}$ interactions, all generated events are overlaid with an event from a sample of random beam crossings with the same instantaneous luminosity profile as data. Further on, events are reconstructed using the same software as is used for data.

The signal cross sections and branching ratios are normalized to the SM predictions [8]. Next-to-LO (NLO) cross sections are used for single top quark [30] and diboson [31, 32] production, and approximate next-to-NLO (NNLO) for $t\bar{t}$ production [33]. The $V+$jets processes are normalized to the NNLO cross section [34] with MSTW2008 NNLO PDFs [35] The $V+$heavy-flavor events are corrected using the NLO to LO ratio obtained from mcfm [32, 36]. We compare data with the prediction for $V+$jets production and find a relative data/MC normalization factor of $1.0 \pm 0.1$, obtained after subtracting all other expected background processes and before $b$-tagging.

This analysis begins with the selection of events with exactly one charged lepton, either an electron with transverse momentum $p_T > 15$ GeV and pseudorapidity $|\eta| < 1.1$ or $1.5 < |\eta| < 2.5$, or a muon with $p_T > 15$ GeV and $|\eta| < 2.0$. Events are also required to have $\mathbf{E}_T > 15$ (20) GeV for the electron (muon) channel and two or three jets with $p_T > 20$ GeV (after calibration of the jet energy [38]) and $|\eta| < 2.5$. $\mathbf{E}_T$ is calculated from the energy deposits in the calorimeter cells and is corrected for the presence of muons [18].

Electron candidates are identified based on a multivariate discriminant that uses information from the central tracker, preshower detectors, and calorimeter. Muon candidates are identified from the hits in muon system that are matched to a central track and must be isolated from the energy deposits in calorimeter. Inefficiencies introduced by lepton identification and isolation criteria are determined from $Z \rightarrow \ell \ell$ data and used to correct the efficiency in simulated events to match that in data.

Jets are reconstructed using a midpoint cone algorithm [39] with a radius of $\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2} = 0.5$, where $y$ is the rapidity. Differences in efficiency for jet identification and jet energy resolution between data and simulation are applied as corrections to the MC [18].

Comparison of ALPGEN with other generators [40] and with data [41] shows discrepancies in distributions of lepton and jet $\eta$, dijet angular separations, and the $p_T$ of $W$ and $Z$ bosons for $V+$jets events. The data are therefore used to correct the ALPGEN $V+$jets MC events by weighting the simulated distributions of lepton $\eta$, leading and second-leading jet $\eta$, $\Delta R$ between the two leading jets, and the $W$ boson $p_T$ through the use of functions.
that bring the total simulated background into agreement with data before b-tagging, similar to the method employed in Ref. [18].

Multijet backgrounds are estimated from data [18]. Before applying b-tagging, we perform a fit to the distribution of the transverse mass [6] of the W boson candidate ($M_W^T$) to determine the normalization of the MJ and V+jets backgrounds simultaneously. To suppress MJ background, events with $M_W^T < (40 - 0.5 \times \mathcal{E}_T)$ are removed in both the electron and muon channels.

To further suppress the MJ background, we construct a multivariate discriminator (MJ-MVA) that exploits kinematic differences between the MJ background and signal. The MJ-MVA is a boosted decision tree (BDT) implemented in the tmva package [42]. The output distribution in data is well modeled by the total expected simulated and MJ backgrounds and is used as one of the inputs to the final signal discriminant.

The b-tagging algorithm for identifying jets originating from b quarks is based on a combination of variables sensitive to the presence of tracks or secondary vertices displaced significantly from the $pp$ interaction vertex. This algorithm provides improved performance over an earlier neural network algorithm [43]. The efficiency is determined for taggable jets, which contain at least two tracks with each having at least one hit in the silicon microstrip tracker. The efficiency for jets to satisfy the taggability and b-tagging requirements in the simulation is corrected to reproduce the data.

Events must have at least one b-tagged jet. If exactly one jet is b-tagged, the b-identification discriminant output of that jet must satisfy the tight selection threshold described below. Such events are classified as having one tight b-tag. Events with two or more b-tagged jets are assigned to either the two loose b-tags, two medium b-tags, or two tight b-tags category, depending on the value of the average b-identification discriminant of the two jets with the highest discriminant values. The operating point for the loose (medium, tight) threshold has an identification efficiency of 79% (57%, 47%) for individual b jets, averaged over selected jet $p_T$ and η distributions, with a b-tagging misidentification rate of 11% (0.6%, 0.15%) for light quark jets, calculated by the method described in Ref. [43].

After applying these selection criteria, the expected event yields for the backgrounds and for a Higgs boson with mass $M_H = 125$ GeV are compared to the observed number of events in Table I. Figure 1(a) shows the distribution of the dijet invariant mass, using the two jets with the highest b-identification output, for events with exactly two jets and all b-tagged categories. The data are well described by the predicted background in all b-tag categories.

To separate signal and background, we use final BDTs trained on the $WH \rightarrow l\nu b\bar{b}$ signal samples and all the SM processes as background. We train an independent final BDT, using an individually optimized set of inputs, for each lepton flavor, jet multiplicity, b-tag category, and $M_H$ value considered, with $M_H$ varying between 100 and 150 GeV in 5 GeV steps. When selecting input variables, we ensure that each is well modeled and displays good separation between signal and one or more backgrounds. Figure 1(b,c) shows the final BDT output distributions for the two medium and two tight b-tag channels in two-jet events with electron and muon channels combined.

Uncertainties on the normalization and shape of the final BDT output distributions affect our sensitivity to a potential signal. Theoretical uncertainties include uncertainties on the $t\bar{t}$ and single top quark production cross sections (each having a 7% uncertainty [30, 33]), an uncertainty on the diboson production cross section (6% [31]), $V$+light-flavor production (6%), and $V$+heavy-flavor production (20%, estimated from mcfm [32, 36]).

Uncertainties from modeling that affect both the shape and normalization of the final BDT distributions include uncertainties on trigger efficiency as derived from data (3–5%), lepton identification and reconstruction efficiency (5–6%), re-weighting of alpgen MC samples (2%), the MLM matching [23] applied to $V$+light-flavor events ($\approx 0.5$%). Uncertainties on the alpgen renormalization and factorization scales are evaluated by multiplying the nominal scale for each, simultaneously, by factors of 0.5 and 2.0 (2%), while uncertainties on the choice of PDFs (2%) are estimated using the prescription of Ref. [28, 44].

Experimental uncertainty that affects only the normalization of the expected signal and simulated backgrounds arise from the uncertainty on the integrated luminosity (6.1%) [45]. Those that affect the final BDT distribution shapes include jet taggability (3% per jet), b-tagging efficiency (2.5–3% per heavy-quark jet), the light-quark jet misidentification rate (10% per jet), jet identification efficiency (5%), jet energy calibration and resolution (vary-

| TABLE I: Summary of event yields for W+2 and W+3 jets final states. Number of events in data is compared with the expected number of background events. Signal contributions ($M_H = 125$ GeV) are shown for $WH$ and $ZH$ production with $H \rightarrow b\bar{b}$. All listed signal sources are considered when setting limits. Uncertainties include both statistical and systematic contributions, as described later in this Letter. |
|-----------------|-----------------|-----------------|
|                  | Pre-b-tag       | One tight b-tag  | Two b-tags      |
| $WH$             | 41.2 ± 3.2      | 125 ± 1.2       | 17.3 ± 1.7      |
| $ZH$             | 4.7 ± 0.4       | 1.4 ± 0.1       | 1.9 ± 0.1       |
| $VV$             | 6824 ± 678      | 648 ± 55        | 256 ± 18        |
| $V+1f$           | 206358 ± 18624  | 7149 ± 794      | 2527 ± 306      |
| $V+hf$           | 34068 ± 4447    | 6486 ± 1510     | 3164 ± 739      |
| Top              | 7222 ± 555      | 2413 ± 229      | 2437 ± 238      |
| Multijet         | 68366 ± 6668    | 4634 ± 473      | 2020 ± 192      |
| All bkg.         | 322838 ± 24756  | 21330 ± 2190    | 10404 ± 1059    |
| Data             | 322836          | 20684           | 10071           |
The Higgs boson signal is shown for $M_W$ cross sections in the same final state as $20\%$.

The MJ background model has a contribution from each channel according to the bins’ log $s/b$ values.

The solid lines represent the $\pm 1$ s.d. systematic uncertainty after constraining with data. The darker shaded region is the expected final BDT distribution for a SM Higgs signal for $M_H = 125$ GeV. Here we combine BDT discriminant bins from each channel according to the bins’ $\log_{10}(s/b)$ values.

In conclusion, we have performed a search for SM Higgs boson production in $\ell+\vec{E}_T+n$+jets final states using two or three jets and $b$-tagging with the full Run II data set of 9.7 fb$^{-1}$ of integrated luminosity from the DØ detector. The results are in agreement with the expected event.
yield, and we set upper limits on $\sigma(pp \rightarrow H + X) \times B(H \rightarrow bb)$ relative to the SM Higgs boson cross section $\sigma_{SM}$ for $M_H$ between 100 and 150 GeV, as summarized in Table II. For $M_H = 125$ GeV, the observed limit normalized to the SM prediction is 5.2 and the expected limit is 4.7.

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); MON, NRC KI and RFBR (Russia); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Co-
ciencias (Colombia); CONACyT (Mexico); NRF (Ko-
rea); FOM (The Netherlands); STFC and the Royal Soci-
ety (United Kingdom); MSMT and GACR (Czech Repub-
clic); BMBF and DFG (Germany); SFI (Ireland); The
Swedish Research Council (Sweden); and CAS and CNSF
(China).

![Log-likelihood ratio for the background-only model (LLR_B), with 1 s.d. and 2 s.d. uncertainty bands), signal+background model (LLR_{S+B}) and data (LLR_{obs}) versus M_H.](image)

**FIG. 3:** (color online). Log-likelihood ratio for the background-only model (LLR_B, with 1 s.d. and 2 s.d. uncertainty bands), signal+background model (LLR_{S+B}) and data (LLR_{obs}) versus $M_H$.

**TABLE II:** The ratio of the observed, $R_{obs}$, and expected, $R_{exp}$, 95% upper limit to the SM Higgs boson production cross section.

<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>$R_{exp}$</td>
<td>2.2</td>
<td>2.5</td>
<td>2.9</td>
<td>3.2</td>
<td>3.8</td>
<td>4.7</td>
<td>6.2</td>
<td>8.2</td>
<td>11.7</td>
<td>17.5</td>
<td>25.6</td>
</tr>
<tr>
<td>$R_{obs}$</td>
<td>2.8</td>
<td>2.6</td>
<td>2.9</td>
<td>3.7</td>
<td>5.0</td>
<td>5.2</td>
<td>6.8</td>
<td>8.9</td>
<td>15.1</td>
<td>18.8</td>
<td>21.8</td>
</tr>
</tbody>
</table>

[29] R. Brun and F. Carminati, GEANT Detector Descrip-
tion and Simulation Tool, CERN Program Library Long
[37] The pseudorapidity $\eta = \ln(1 + \tan\frac{\theta}{2})$, where $\theta$ is the polar angle as measured from the proton beam axis.