Search for the standard model Higgs boson in associated $WH$ production in 9.7 fb$^{-1}$ of $pp$ collisions with the D0 detector


1 LAFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil
2 Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
3 Universidade Federal do ABC, Santo André, Brazil
4 University of Science and Technology of China, Hefei, People’s Republic of China
5 Universidad de los Andes, Bogotá, Colombia
6 Charles University, Faculty of Mathematics and Physics, Center for Particle Physics, Prague, Czech Republic
7 Czech Technical University in Prague, Prague, Czech Republic
8 Center for Particle Physics, Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
9 Universidad San Francisco de Quito, Quito, Ecuador
10 LPC, Université Blaise Pascal, CNRS/IN2P3, Clermont, France
11 LPSC, Université Joseph Fourier Grenoble 1, CNRS/IN2P3, Institut National Polytechnique de Grenoble, Grenoble, France
12 CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France
13 LAL, Université Paris-Sud, CNRS/IN2P3, Orsay, France
14 LPNHE, Universités Paris VI and VII, CNRS/IN2P3, Paris, France
15 CEA, Ifsa, SPP, Saclay, France
16 IPHC, Université de Strasbourg, CNRS/IN2P3, Strasbourg, France
17 IPNL, Université Lyon 1, CNRS/IN2P3, Villeurbanne, France and Université de Lyon, Lyon, France
18 III. Physikalisches Institut A, RWTH Aachen University, Aachen, Germany
19 Physikalisches Institut, Université Freiburg, Freiburg, Germany
20 II. Physikalisches Institut, Georg-August-Universität Göttingen, Göttingen, Germany
21 Institut für Physik, Universität Mainz, Mainz, Germany
22 Ludwig-Maximilians-Universität München, München, Germany
23 Fachbereich Physik, Bergische Universität Wuppertal, Wuppertal, Germany
24 Panjab University, Chandigarh, India
25 Delhi University, Delhi, India
26 Tata Institute of Fundamental Research, Mumbai, India
27 University College Dublin, Dublin, Ireland
28 Korea Detector Laboratory, Korea University, Seoul, Korea
29 CINVESTAV, Mexico City, Mexico
30 Nikhef, Science Park, Amsterdam, Netherlands
31 Radboud University Nijmegen, Nijmegen, Netherlands
32 Joint Institute for Nuclear Research, Dubna, Russia
33 Institute for Theoretical and Experimental Physics, Moscow, Russia
34 Moscow State University, Moscow, Russia
35 Institute for High Energy Physics, Protvino, Russia
36 Petersbury Nuclear Physics Institute, St. Petersburg, Russia
37 Institució Catalana de Recerca i Estudis Avançats (ICREA) and Institut de Física d’Altes Energies (IFAE), Barcelona, Spain
38 Uppsala University, Uppsala, Sweden
39 Lancaster University, Lancaster LA1 4YB, United Kingdom
40 Imperial College London, London SW7 2AZ, United Kingdom
41 The University of Manchester, Manchester M13 9PL, United Kingdom
42 University of Arizona, Tucson, Arizona 85721, USA
43 University of California Riverside, Riverside, California 92521, USA
44 Florida State University, Tallahassee, Florida 32306, USA
45 Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA
46 University of Illinois at Chicago, Chicago, Illinois 60607, USA
47 Northern Illinois University, DeKalb, Illinois 60115, USA
48 Northwestern University, Evanston, Illinois 60208, USA
49 Indiana University, Bloomington, Indiana 47405, USA
50 Purdue University Calumet, Hammond, Indiana 46323, USA
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 $W H \rightarrow \ell \nu b\bar{b}$ production and uses data corresponding to 9.7 fb$^{-1}$ of integrated luminosity collected with the D0 detector at the Fermilab Tevatron $p\bar{p}$ Collider at $\sqrt{s} = 1.96$ TeV. We observe agreement between the data and the expected background. For a Higgs boson mass of 125 GeV, we set a 95% C.L. upper limit on the production of a standard model Higgs boson of $5.2 \times \sigma_{SM}$, where $\sigma_{SM}$ is the standard model Higgs boson production cross section, while the expected limit is $4.7 \times \sigma_{SM}$.

PACS numbers: 14.80.Bn, 13.85.Rm

The Higgs boson is the only fundamental particle in the standard model (SM) predicted as a direct consequence of the Higgs mechanism describing spontaneous electroweak symmetry breaking.$^{1,3}$ The Higgs mechanism generates the masses of the weak gauge bosons and provides an explanation for the nonzero masses of fermions generated by their Yukawa couplings with the gauge bosons and provides an explanation for the nonzero symmetry breaking.$^{1–3}$. The mass of the Higgs boson ($M_H$) is a free parameter in the SM that must be constrained by experimental results. The direct searches at the CERN LEP Collider (LEP)$^{4}$ exclude $M_H < 114.4$ GeV at the 95% confidence level (C.L.) and precision measurements of other electroweak parameters constrain $M_H$ to be less than 152 GeV.$^{3,7}$ 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 $111 < M_H < 122$ GeV, $129 < M_H < 559$ GeV (ATLAS)$^{11}$, and $110 < M_H < 122$ GeV, $127 < M_H < 600$ GeV (CMS)$^{12}$. Both experiments have observed a resonance consistent with SM Higgs production at $M_H \approx 125$ GeV, primarily in the $\gamma\gamma$ 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 $b\bar{b}$ final state, which is the dominant decay mode in this mass range.

The dominant Higgs boson production process at the

$^{1}$with visitors from $^a$Augustana College, Sioux Falls, SD, USA, $^b$The University of Liverpool, Liverpool, UK, $^c$UPIITA-IPN, Mexico City, Mexico, $^d$DESY, Hamburg, Germany, $^e$SLAC, Menlo Park, CA, USA, $^f$University College London, London, UK, $^g$Centro de Investigacion en Computacion - IPN, Mexico City, Mexico, $^h$ECFM, Universidad Autonoma de Sinaloa, Culiacán, Mexico and $^i$Universidade Estadual Paulista, São Paulo, Brazil.
Tevatron Collider is gluon-gluon fusion. The associated production of a Higgs boson with a weak boson occurs at a rate about 3 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 ($E_T$) 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 pp 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 object in the calorimeter or an electromagnetic object with additional jets. In the muon channel we use a mixture of single muon, muon plus jet, $E_T$ plus jet, and multijet triggers. We correct simulated events for trigger efficiency by using a method similar to that described in Ref. [18].

Several SM processes produce or can mimic a final state with a charged lepton, $E_T$, and jets, including diboson ($WW$, $WZ$, and $ZZ$), $V+$jets ($V = W$ or $Z$), $t\bar{t}$, 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 [24] Monte Carlo (MC) generator interfaced to PYTHIA [24] for parton showering and hadronization. ALPGEN samples are produced by using the MLM parton-jet matching prescription [25]. The $V+$jets samples contain $V + j j$ (where $j = u, d, s, \text{ or } g$) and $V + c j$ (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 the data. Further on, events are reconstructed by using the same software as is used for the data.

The signal cross sections and branching fractions $B$ are normalized to the SM predictions $\mathcal{B}$. 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 parton distribution functions [35]. The $V+$heavy-flavor events are corrected by using the NLO to LO ratio obtained from the Monte Carlo program MCFM [32, 36]. We compare the data with the prediction for $V+$jets production and find a relative data to MC normalization factor of 1.0 ± 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 $E_T > 15$ (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 $|\eta| < 2.5$. $E_T$ 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 the muon system that are matched to a central track and must be isolated from the energy deposits in the 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 the data.

Jets are reconstructed by 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 the data and simulation are applied as corrections to the MC [18].

Comparison of ALPGEN with other generators [40] and with the 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 the data before $b$ tagging, similar to the method employed in Ref. [18].

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

To further suppress the MJ background, we construct a multivariate discriminator that exploits kinematic differences between the MJ background and signal. The multivariate discriminator is a boosted decision tree (BDT) implemented in the TMVA package [42]. The output distribution in the 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 $p\bar{p}$ 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 efficiency of 79% (57%, 47%) for individual $b$ jets, averaged over selected jet $p_T$ and $\eta$ 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.

<table>
<thead>
<tr>
<th></th>
<th>Pre-$b$-tag</th>
<th>One tight $b$-tag</th>
<th>Two $b$-tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>$WH$</td>
<td>41.2 ± 3.2</td>
<td>12.5 ± 1.2</td>
<td>17.3 ± 1.7</td>
</tr>
<tr>
<td>$ZH$</td>
<td>4.7 ± 0.4</td>
<td>1.4 ± 0.1</td>
<td>1.9 ± 0.1</td>
</tr>
<tr>
<td>$V\ell$</td>
<td>6824 ± 678</td>
<td>648 ± 55</td>
<td>256 ± 18</td>
</tr>
<tr>
<td>$V + 1f$</td>
<td>206 358 ± 18624</td>
<td>7149 ± 794</td>
<td>2527 ± 306</td>
</tr>
<tr>
<td>$V + hf$</td>
<td>34 068 ± 4447</td>
<td>6486 ± 1510</td>
<td>3164 ± 739</td>
</tr>
<tr>
<td>Top</td>
<td>7222 ± 555</td>
<td>2413 ± 229</td>
<td>2437 ± 238</td>
</tr>
<tr>
<td>Multijet</td>
<td>68 366 ± 6668</td>
<td>4634 ± 473</td>
<td>2020 ± 192</td>
</tr>
<tr>
<td>All bkg.</td>
<td>322 838 ± 24 756</td>
<td>21 330 ± 2100</td>
<td>10 404 ± 1059</td>
</tr>
<tr>
<td>Data</td>
<td>322 836</td>
<td>20 684</td>
<td>10 071</td>
</tr>
</tbody>
</table>

To separate the signal and background, we use final BDTs trained on the $WH \rightarrow Wb\bar{b}$ 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 the signal and one or more backgrounds. Figures 1(b) and 1(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 $tt$ 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 MC@NLO [32, 34]).

Uncertainties from modeling that affect both the shape and normalization of the final BDT distributions include uncertainties on trigger efficiency as derived from the data (3%-5%), lepton identification and reconstruction efficiency (5%-6%), reweighting of Alpgen MC samples (2%), and the MLM matching [22] 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 parton distribution functions (2%) are estimated by using the prescriptions of Ref. [28, 44].

Experimental uncertainty that affects only the normalization of the expected signal and simulated backgrounds arises from the uncertainty on the integrated luminosity (6.1%) [15]. 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%), and jet energy calibration and resolution (varying between 5% and 15%, depending on the process and channel). The MJ background model has a contribution from the statistical uncertainty of the data after tagging (10%–20%).

To demonstrate measurement of processes with small cross sections in the same final state as $WH$, we train a discriminant with $WZ$ and $ZZ$ production as the signal, using the same event selection and input variables. We observe a 1.0 s.d. excess in the data over the background expectation, and our expected sensitivity is 1.8 s.d. If interpreted as a cross section measurement, the resulting scale factor with respect to the predicted SM value \cite{31,32} of 4.4 ± 0.3 pb is 0.55 ± 0.36 (stat.) ± 0.37 (syst.).

In the search for the SM Higgs boson we observe no significant excess relative to the SM expectation and proceed to set upper limits on the SM Higgs boson production cross section. We calculate all limits at the 95% C.L. using the modified frequentist CLs approach with a Poisson log-likelihood ratio of the signal+background hypothesis to the background-only hypotheses as the test statistic \cite{40,48}. We treat systematic uncertainties as “nuisance parameters” constrained by their priors, and the best fits of these parameters are determined at each value of $M_H$ by maximizing the likelihood with respect to the data. We remove the $V$+jets normalization obtained from the $M_H^V$ distribution and allow the components to vary by the aforementioned uncertainties of 6% and 20% on $V$+light-flavor and $V$+heavy-flavor production, respectively. Independent fits are performed to the background-only and signal-plus-background hypotheses. All correlations are maintained among channels and between the signal and background. Figure 2 shows the background-subtracted data along with the best fit for the background-only model ±1 s.d. systematic uncertainties and the expected signal contribution for all channels combined, where we combine bins from each channel according to their bins’ log_{10}(s/b) values.

![FIG. 1: (color online). (a) The dijet mass distribution for all $b$-tag categories and two-jet exclusive events. (b) The final BDT output for two medium $b$-tagged events and (c) two tight $b$-tagged events. Electron and muon channels are combined. The Higgs boson signal is shown for $M_H = 125$ GeV. Signal events are scaled by a factor of 100 in (a) and 20 in (b) and (c).](image)

![FIG. 2: (color online). Distribution of the difference between the data and background expectation of the final BDT discriminant output for $M_H = 125$ GeV for the background-only model, shown with statistical uncertainties (points with error bars). The solid lines represent the ±1 s.d. systematic uncertainty after constraining with the 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.](image)
TABLE II: The ratio of the observed, $R_{obs}$, and expected, $R_{expt}$, 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>$R_{expt}$</td>
<td>2.2</td>
<td>2.5</td>
<td>2.9</td>
<td>3.2</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>
</tr>
<tr>
<td>$R_{obs}$</td>
<td>2.6</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>

In conclusion, we have performed a search for SM Higgs boson production in $\ell^{+}\ell^{-}+$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 D0 detector. The results are in agreement with the expected event yield, and we set upper limits on $\sigma(p\bar{p} \rightarrow H + X) \times B(H \rightarrow b\bar{b})$ 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, FAPE RJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); NRF (Korea); FOM (The Netherlands); STFC and the Royal Society (United Kingdom); MSMT and GACR (Czech Republic); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); and CAS and CNSF (China).

The pseudorapidity $\eta = -\ln \left( \tan \frac{\theta}{2} \right)$, where $\theta$ is the polar angle as measured from the proton beam axis.