Search for a Standard Model Higgs Boson in WH→lvbb-bar in pp-bar Collisions at sqrt[s]=1.96 TeV

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>CDF Collaboration et al. “Search for a Standard Model Higgs Boson in WH→lvbb-bar in pp-bar Collisions at s=1.96 TeV.” Physical Review Letters 103.10 (2009): 101802. © 2009 The American Physical Society</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1103/PhysRevLett.103.101802">http://dx.doi.org/10.1103/PhysRevLett.103.101802</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Physical Society</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun Apr 07 05:45:26 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/51882">http://hdl.handle.net/1721.1/51882</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Search for a Standard Model Higgs Boson in $\ell\nu\bar{b}\bar{b}$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV


(CDF Collaboration)

1Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China
2Argonne National Laboratory, Argonne, Illinois 60439, USA
3University of Athens, 157 71 Athens, Greece
4Institut de Fisica d’Altes Energies, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain
5Baylor University, Waco, Texas 76798, USA
6aIstituto Nazionale di Fisica Nucleare Bologna, I-40127 Bologna, Italy
6bUniversity of Bologna, I-40127 Bologna, Italy
7Brandeis University, Waltham, Massachusetts 02254, USA
8University of California, Davis, California 95616, USA
9University of California, Los Angeles, Los Angeles, California 90024, USA
10University of California, San Diego, La Jolla, California 92093, USA
11University of California, Santa Barbara, Santa Barbara, California 93106, USA
12Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain
13Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
14Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA
15Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia
16Joint Institute for Nuclear Research, RU-141980 Dubna, Russia
17Duke University, Durham, North Carolina 27708, USA
18Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA
19University of Florida, Gainesville, Florida 32611, USA
20Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy
21University of Geneva, CH-1211 Geneva 4, Switzerland
22Glasgow University, Glasgow G12 8QQ, United Kingdom
23Harvard University, Cambridge, Massachusetts 02138, USA

PRL 103, 101802 (2009) PHYSICAL REVIEW LETTERS week ending 4 SEPTEMBER 2009

101802-2
We present a search for a standard model Higgs boson produced in association with a W boson using 2.7 fb$^{-1}$ of integrated luminosity of $p\bar{p}$ collision data taken at $\sqrt{s} = 1.96$ TeV. Limits on the Higgs boson production rate are obtained for masses between 100 and 150 GeV/c$^2$. Through the use of multivariate techniques, the analysis achieves an observed (expected) 95% confidence level upper limit of 5.6 (4.8).
times the theoretically expected production cross section for a standard model Higgs boson with a mass of 115 GeV/c².

DOI: 10.1103/PhysRevLett.103.101802 PACS numbers: 14.80.Bn, 13.85.Rm, 14.70.Fm

The standard model (SM) of particle physics has proven to be an extremely successful theory through its accurate predictions of many experimental results over the last few decades. In the SM, spontaneous electroweak symmetry breaking gives rise to the masses of the W and Z bosons. Although the Higgs mechanism [1–3] was proposed in the 1960s as the source of this symmetry breaking, the fundamental particle it predicts to exist, the Higgs boson, has yet to be discovered. The mass of the Higgs boson is a free parameter of the SM. However, direct limits from the LEP experiments exclude Higgs boson masses below 114.4 GeV/c² [4] at 95% confidence level (C.L.). Taking into account additional electroweak precision measurements places a 95% C.L. upper limit on the mass of a SM Higgs boson of 185 GeV/c² [5]. Recently, combined results from the CDF Collaboration and D0 Collaboration experiments have excluded at the 95% C.L. Higgs boson masses between 160 and 170 GeV/c² [6].

For Higgs boson masses below 135 GeV/c², b⟨b⟩ is the main decay mode [7]. In this decay, each b quark fragments into a jet of hadrons and the Higgs boson signal may be reconstructed as a peak in the invariant mass distribution of these two jets. At the Tevatron associated production with a W boson (WH), where the W boson decays into a lepton (ℓ) and a neutrino (ν), provides one of the most sensitive search channels in this mass range, since the requirements of a charged lepton candidate and of large missing transverse energy dramatically reduce the backgrounds from multijet processes [8]. Both Tevatron experiments, CDF and D0, have published search results for WH → ℓνbb [9–11]. Here we describe a new search for the Higgs boson in the WH → ℓνbb channel with increased signal acceptance that employs improved analysis technique and 2.7 fb⁻¹ of pp collision luminosity collected by the CDF experiment. Although we focus on the SM here, many plausible extensions, such as Ref. [12], predict a low-mass SM-like Higgs boson.

The CDF II apparatus [13,14] is a general-purpose detector located at the Tevatron collider at Fermilab. The detector consists of a solenoidal charged-particle spectrometer which includes a silicon microstrip detector array surrounded by a cylindrical drift chamber in a 1.4 T axial magnetic field. Outside the tracking chambers, the energies of electrons and jets are measured with segmented sampling calorimeters. Surrounding the calorimeters are layers of steel instrumented with planar drift chambers and scintillators used for muon identification.

Events are collected with energetic lepton triggers that require one of the following signatures [15]: a high-pT electron candidate, a high-pT muon candidate, or missing transverse energy (E_T from the neutrino escaping detection) with an energetic forward (|η| > 1.2) electromagnetic cluster (designed to accept forward electrons from the W boson decay). An additional trigger is included that does not explicitly require an identified lepton, but instead requires large E_T plus two well-separated jets in η − φ space [16]. For these events, the charged lepton from the W boson decay is reconstructed only as a high-pT isolated track. The addition of this nontriggered lepton category increases WH → ℓνbb signal acceptance by approximately 25% [17].

Candidate events are selected by requiring a lepton candidate (triggered lepton or isolated track) with p_T^l > 20 GeV/c, E_T > 20 GeV, and two jets with |η| < 2.0 and E_j > 20 GeV after correcting for instrumental effects [18]. At least one of the jets must have a displaced vertex (b tag) defined by the SECVTX algorithm [19] signaling that the jet likely originated from a b quark. An additional b-tagging algorithm that relies on high-impact-parameter tracks within jets, JETPROB [15], is used to increase the acceptance for double-tagged events. Vetoos are applied to remove events with more than one lepton and events without leptonic W boson decays [11].

The Higgs boson events are modeled with the PYTHIA [20] MC generator combined with a parametrized response of the CDF II detector [21,22] and tuned to the Tevatron underlying event data [23]. After basic event selection, the total expected signal event yield in the current data set is 5.1 ± 0.5 (3.5 ± 0.4) single (double)-tag events for a Higgs boson with a mass of 115 GeV/c² (see Table I for other masses).

Models for background processes are derived from a mixture of MC simulation and data-driven techniques [11]. Important backgrounds to WH → ℓνbb include events with a W or Z boson produced in association with jets. These processes may include true b jets as in W + b⟨b⟩, or other jets that have been misidentified as b jets like W + c⟨c⟩ and W + jj, where j refers to jets not originating from heavy-flavor quarks. Events with a top quark (tℓ and single top quark production), diboson events, and multijet events without W bosons also contribute to the sample composition.

After applying the event selection defined above, the background expectation (1896 ± 301 for single-tag and 316 ± 60 for double-tag events) is significantly larger than the expected number of Higgs boson signal events. We have indicated that the dijet invariant mass is a useful variable for separating the Higgs boson signal from the dominant backgrounds, however its usefulness is limited by jet energy resolution and large background rate. These
challenges require that we extract as much discrimination as possible from the full information available in each event. Multivariate techniques allow us to collect the discriminating power of many variables into a single output variable. We take advantage of the benefits from different techniques [24] by combining the discriminating power of two separate analyses that use the same event selection but follow different multivariate strategies. We validate the discriminants separately for each Higgs boson mass hypothesis, and construct the discriminants so that they are not sensitive to statistical fluctuations in the background and signal samples. We first summarize the two analyses, and then discuss their combined result.

The first analysis uses an artificial neural network (NN, [25]) trained to discriminate $WH \rightarrow \ell \nu b\bar{b}$ signal from the background using the information contained in the following kinematic variables: the invariant mass of the two jets plus an additional “loose” jet [26] if it lies close to one of the primary jets [angular separation $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$ less than 0.9]; the vector sum of the transverse energies ($\sum_{jets} \vec{E}_T + \vec{p}_T + \vec{\mu}_T$); the scalar sum of the lepton and jet transverse momenta minus the $\vec{E}_T$ ($\sum_{jets} E_T + p_T - \vec{E}_T$); the scalar sum of the loose jet transverse energy ($\sum_{jets} E_T^{loose}$) [26]; the minimum invariant mass of the lepton, $\vec{E}_T$, and one of the two jets [$\min(M_{\ell,j_1,j_2}, M_{\ell,\mu})$]; and $\Delta R$ between the lepton and the momentum of the neutrino [27]. The strongest discriminating variable of the NN is the dijet mass variable shown in Fig. 1(a).

The second analysis uses a boosted decision tree technique (MEBDT, [28,29]). The notation MEBDT underscores the use of inputs derived from the matrix-element approach developed in Refs. [30,31]. In the matrix-element method, probability densities are calculated for each event using the measured kinematic quantities. Some of the best discriminating inputs to the decision tree include ratios of the signal event probabilities to various combinations of

![Graphs and plots](image-url)
the background probabilities, and an event probability
discriminant (EPD) defined as the ratio of the signal event
probability to the sum of the signal and all background
event probabilities as in Ref. [30]. The EPD distributions
for signal and backgrounds are shown in Fig. 1(b).

The MEBDT analysis also uses the output of a neural
network that has been trained to separate jet flavors [32].
This network is based on secondary vertex tracking infor-
mation and provides a continuous variable which helps to
identify the portion of the background that does not contain
real b-quark jets. The MEBDT analysis also includes the
following inputs: the dijet mass, the $E_T$ of both jets and $\vec{E}_T$
of the event, the difference in azimuthal angles ($\Delta \phi$)
between the leading jet and the $\vec{E}_T$, the $\Delta \phi$ between
the lepton and the $\vec{E}_T$, the $p_T$ and the $\eta$ of the lepton, the scalar
sum of the transverse energies $H_T = \sum_{\text{jets}} E_T + p_T^\ell + \vec{E}_T$,
the cosine of the angle between the lepton and leading
jet, and the transverse mass of the $W$ boson $M_T(W) = \sqrt{2(p_T^\ell \cdot \vec{E}_T - \vec{p}_T^\ell \cdot \vec{E}_T)}$.

We performed the NN and the MEBDT analyses inde-
dependently (see Table I), the results of which are partially
correlated. The correlations between the discriminant out-
puts range between 50% and 75% for the major back-
ground and signal samples. These correlations, while
high, do suggest that a sensitivity gain can be obtained
by combining the two approaches. We combine the NN and
MEBDT discriminants using a superdiscriminant (SD)
technique first developed in the CDF single top quark
search [30]. Here, a neural network using the discriminant
outputs of the NN and MEBDT as inputs is optimized using
genetic algorithms [33–35]. Three separate neural net-
works (one for each b-tag category: single SECVTX,
SECVTX + JETPROB, and double SECVTX) are trained
to separate the $WH_b\rightarrow \ell \nu b\bar{b}$ signal from the backgrounds
for each Higgs boson mass using events from the signal
and background samples described above. The distributions
of the SD outputs of the neural network trained for a
Higgs boson mass of 115 GeV/$c^2$ are shown in Fig. 1(c)
for the combined double-tag categories and the single-tag
category. The SD analysis improves the sensitivity com-
pared to the best individual analysis by 5%-13% for the
Higgs boson masses studied.

Finding no evidence for a Higgs boson signal, we cal-
culate a Bayesian C.L. limit for each mass hypothesis
based on the combined binned likelihood of the SD output
distributions. The two lepton categories (triggered leptons
and isolated tracks) and three tag categories yield six
independent channels that are included in the likelihood.
Systematic uncertainties on the rate of signal and back-
ground production from jet energy scale, $b$-tagging effi-
ciencies, lepton identification and trigger efficiencies, the
amount of initial and final state radiation, and the parton
distribution functions are included in the limit calculation
(for details on systematic studies, see [11,17]).

Uncertainties on the discriminant output shapes were
studied but found to have a negligible impact on sensitivity.
A posterior density is obtained by multiplying this like-
lihood by Gaussian prior densities for the background
normalizations and systematic uncertainties leaving $\sigma \times \mathcal{B}(H \rightarrow b\bar{b})$ with a uniform prior density. A 95% C.L. limit
is then determined such that 95% of the posterior density
for $\sigma \times \mathcal{B}(H \rightarrow b\bar{b})$ falls below the limit [36]. Removing
systematic uncertainties completely from the limit calcu-
lation improves the expected limit by about 15%.

Table I shows the expected and observed limits calcu-
lated for different Higgs boson masses. The limits are
displayed graphically in Fig. 2. We find an observed (ex-
pected) 95% C.L. limit of 5.6 (4.8) times the SM prediction
of the production cross section for a Higgs boson mass of
115 GeV/$c^2$ (next-to-leading order theory predicts $\sigma \times \mathcal{B}(H \rightarrow b\bar{b}) = 136 \text{ fb}$ [37]). At this mass, the expected
limit has improved by a factor of 1.7 over the 1.9 fb$^{-1}$
result from CDF [11], which corresponds to a 40% im-
provement in sensitivity over what is expected from the
increased data set [38]. The additional gain comes from our
increased lepton acceptance through the inclusion of a
nontriggered lepton category, a continuous jet flavor sepa-
ator variable which improves discrimination of light-
quark jets mistakenly tagged as $b$ jets, and the use of
new multivariate techniques. The excess in the observed
limit at higher masses is due primarily to the slight excess
observed at 150 GeV/$c^2$ in the dijet mass variable [see
Fig. 1(a)] and is an indication of the large weight this
variable carries in the full multivariate analysis. The suc-
ful previous application of many of the techniques to
the CDF single top analysis [30,39], and the consistency of
results obtained with NN and MEBDT algorithms provide
further confidence in the robustness of the multivariate
techniques. The increasing Tevatron data set together
with future analysis improvements, a combination of re-

FIG. 2 (color online). The expected and observed 95% C.L.
upper limits on the Higgs boson production cross section relative
to the SM expectation as obtained from the SD combination as a
function of the Higgs boson mass.
results from all Higgs boson production and decay modes, as well as the combination with the results from the D0 experiment [6], will continue to provide improved levels of sensitivity to the SM Higgs boson searches at the Tevatron.

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 Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, U.K.; the Institut National de Physique Nucleaire 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.

aDeceased.
bVisitor from University of Massachusetts Amherst, Amherst, MA 01003, USA.
cVisitor from Universiteit Antwerpen, B-2610 Antwerp, Belgium.
dVisitor from University of Bristol, Bristol BS8 1TL, U.K.
eVisitor from Chinese Academy of Sciences, Beijing 100864, China.
fVisitor from Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09042 Monserrato (Cagliari), Italy.
gVisitor from University of California Irvine, Irvine, CA 92697, USA.
hVisitor from University of California Santa Cruz, Santa Cruz, CA 95064, USA.
iVisitor from Cornell University, Ithaca, NY 14853, USA.
jVisitor from University of Cyprus, Nicosia CY-1678, Cyprus.
kVisitor from University College Dublin, Dublin 4, Ireland.
lVisitor from Royal Society of Edinburgh, Edinburgh, EH2 2PQ, U.K.
mVisitor from University of Edinburgh, Edinburgh EH9 3JZ, U.K.
nVisitor from Universidad Iberoamericana, Mexico D.F., Mexico.
oVisitor from Queen Mary, University of London, London, E1 4NS, U.K.
pVisitor from University of Manchester, Manchester M13 9PL, U.K.
qVisitor from Nagasaki Institute of Applied Science, Nagasaki, Japan.
rVisitor from University of Notre Dame, Notre Dame, IN 46556, USA.
sVisitor from University de Oviedo, E-33007 Oviedo, Spain.
tVisitor from Texas Tech University, Lubbock, TX 79409, USA.
uVisitor from IFIC(CSIC-Universitat de Valencia), 46071 Valencia, Spain.
vVisitor from University of Virginia, Charlottesville, VA 22904, USA.
wOn leave from J. Stefan Institute, Ljubljana, Slovenia.

[14] We use a cylindrical coordinate system with the origin at the center of the CDF detector, $z$ pointing in the direction of the proton beam, $\theta$ and $\phi$ representing the polar and azimuthal angles, respectively, and pseudorapidity defined by $\eta = -\ln \tan(\theta/2)$. The transverse momentum $p_T$ (transverse energy $E_T$) is defined to be $p \sin \theta$ ($E \sin \theta$). The missing $E_T$ ($\vec{E}_T$) is defined by $\vec{E}_T = -\sum_i E_{ni} \hat{n}_i$, $\hat{n}_i$ is a unit vector perpendicular to the beam axis and pointing at the $i$th calorimeter tower ($\vec{E}_T = |\vec{E}_T|$).
“Loose” jets are defined to be exclusive to our primary jet definition ($|\eta| < 2.0$ and $E_T > 20 \text{ GeV}$) with $|\eta| < 2.4$ and $E_T > 12 \text{ GeV}$.

The neutrino $p_z$ is chosen as the solution of the W mass constraint equation that gives the largest $|p_z|$ for the neutrino.

We assume that the sensitivity would scale inversely to the square root of the integrated luminosity in the absence of analysis improvements.