Combined search for the standard model Higgs boson decaying to a $b\bar{b}$ pair using the full CDF data set


(1) Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China
(2) Argonne National Laboratory, Argonne, Illinois 60439, USA
(3) University of Athens, 157 71 Athens, Greece
(4) Institut de Fisica d'Altes Energies, ICREA, Universitat Autònoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain
(5) Baylor University, Waco, Texas 76798, USA
(6) Istituto Nazionale di Fisica Nucleare, Bologna, Italy
(7) University of California, Davis, Davis, California 95616, USA
(8) University of California, Los Angeles, Los Angeles, California 90024, USA
(9) Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain
(10) Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
(11) Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA
(12) Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia
(13) Joint Institute for Nuclear Research, RU-141980 Dubna, Russia
(14) Duke University, Durham, North Carolina 27708, USA
(15) Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA
(16) University of Florida, Gainesville, Florida 32611, USA
(17) Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy
(18) University of Geneva, CH-1211 Geneva 4, Switzerland
(19) Glasgow University, Glasgow G12 8QQ, United Kingdom
(20) Harvard University, Cambridge, Massachusetts 02138, USA
(21) Division of High Energy Physics, Department of Physics, University of Helsinki and Helsinki Institute of Physics, FIN-00014, Helsinki, Finland
(22) University of Illinois, Urbana, Illinois 61801, USA
(23) The Johns Hopkins University, Baltimore, Maryland 21218, USA
(24) Institut für Experimentelle Kernphysik, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany
(25) Center for High Energy Physics, Kyungpook National University, Daegu 702-701, Korea; Seoul National University, Seoul 151-742, Korea; Sungkyunkwan University, Suwon 440-746, Korea; Korea Institute of Science and Technology Information, Daejeon 305-806, Korea; Chonnam National University, Gwangju 500-757, Korea; Chonbuk National University, Jeonju 561-756, Korea
(26) Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
(27) University of Liverpool, Liverpool L69 7ZE, United Kingdom
(28) University College London, London WC1E 6BT, United Kingdom
(29) Centro de Investigaciones Energeticas Medioambientales y Tecnologicas, E-28040 Madrid, Spain
(30) Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
(31) Institute of Particle Physics: McGill University, Montréal, Québec, Canada H3A 2T8; Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6; University of Toronto, Toronto, Ontario, Canada M5S 1A7; and TRIUMF, Vancouver, British Columbia, Canada V6T 2A3
(32) University of Michigan, Ann Arbor, Michigan 48109, USA
(33) Michigan State University, East Lansing, Michigan 48824, USA
(34) Institution for Theoretical and Experimental Physics, ITEP, Moscow 117259, Russia
(35) University of New Mexico, Albuquerque, New Mexico 87131, USA
The mechanism of electroweak symmetry breaking in the standard model (SM) predicts the existence of a fundamental scalar boson, referred to as the Higgs boson ($H$). Although there is strong evidence of electroweak symmetry breaking, the Higgs boson has yet to be observed. The SM does not predict the mass of the Higgs boson, $m_H$, but the combination of precision electroweak measurements, including recent top quark and $W$ boson mass measurements from the Tevatron and LHC, constrains $m_H < 152 \text{ GeV}/c^2$ at the 95% confidence level. Direct searches at LEP2, the Tevatron, and the LHC exclude all possible masses of the SM Higgs boson at the 95% confidence level or the 95% credibility level (C.L.), except within the ranges $116.6 - 119.4 \text{ GeV}/c^2$ and $122.1 - 127 \text{ GeV}/c^2$. A SM Higgs boson in these mass ranges would be produced in $\sqrt{s} = 1.96 \text{ TeV}$ $pp$ collisions of the Tevatron and have a branching fraction to $bb$ greater than 50%.

While the most sensitive searches for the SM Higgs boson based on the full CDF Run II data set obtained from $\sqrt{s} = 1.96 \text{ TeV}$ $pp$ collisions at the Fermilab Tevatron corresponding to an integrated luminosity of $9.45 \text{ fb}^{-1}$. The searches are conducted for Higgs bosons that are produced in association with a $W$ or $Z$ boson, have masses in the range $90-150 \text{ GeV}/c^2$, and decay into $bb$ pairs. An excess of data is present that is inconsistent with the background prediction at the level of 2.5 standard deviations (the most significant local excess is 2.7 standard deviations).
boson at the LHC are those based on its decays into pairs of
gauge bosons, searches based on decays into pairs of
b quarks are the most sensitive at the Tevatron. The
searches at the LHC in the four-lepton and diphoton fi-
nal state offer precise measurements of the mass of the
Higgs boson, while the results presented here provide di-
rect information about the Higgs boson’s couplings to b
quarks and are therefore complementary to the primary
LHC search modes. In searches for the production of a
Higgs boson in association with a vector boson (WH or
ZH), leptonic decays of the vector boson provide effec-
tive discrimination between the expected signal and the
large, uncertain hadronic backgrounds. Previous Higgs
searches focused on these production and decay modes
have been performed at LEP2 [2] and the LHC [13, 14].
This Letter describes the combination of the results of
three CDF searches for a SM-like Higgs boson with a
mass in the range 90 < m_{H} < 150 GeV/c^{2}. These
searches are targeted at \( ZH \rightarrow \ell^{+}\ell^{-}bb \) [15], \( WH \rightarrow \ell\nu bb \) [16], and \( WH, ZH \rightarrow E_{T}bb \) [17].
The CDF II detector is described in detail else-
where [18, 19]. Calorimeter energy deposits are clustered
into jets using a cone algorithm with an opening angle of
\( \Delta R = \sqrt{(\Delta \phi)^{2} + (\Delta \eta)^{2}} = 0.4 \) [20]. High-pT electron candidates are identified by matching charged-particle
tracks in the tracking systems [21, 22] with energy de-
posits in the electromagnetic calorimeters [23]. Muon
candidates are identified by matching tracks with muon-
detector track segments [24]. The hermeticity of the
calorimeter allows for good reconstruction of the missing
transverse energy \( E_{T} \) [25]. Jets are identified as consis-
tent with the fragmentation of a b quark (b-tagged) using
three different algorithms described in Ref. [20], which
make use of track impact parameters, the presence of
displaced vertices, the presence of leptons near the
jet, and jet kinematic properties. The average tag
efficiency for a jet originating (not originating) from b
quark fragmentation is in the range 42–70% (0.9–8.9%),
depending on the properties of the jet.

Higgs boson signal events are simulated using
PYTHIA [27], with CTEQ5L [28] parton distribution func-
tions (PDFs) at leading order (LO). We normalize our
Higgs boson signal-rate predictions to the most recent
higher-order calculations available. The WH and ZH
cross section calculations are performed at next-to-next-
to-leading order (NNLO) precision in QCD and next-
to-leading-order (NLO) precision in the electroweak cor-
rections and are described in Ref. [10]. The branching
fractions for the Higgs boson decays are obtained from
Ref. [12]. These rely on calculations using HDECAY [29]
and PROPHETC4F [30]. Assuming the \( m_{H} = 125 \) GeV/c^{2}
hypothesis, we expect approximately 85 Higgs boson
events to pass our selections. We model SM and instru-
mental background processes using a mixture of Monte
Carlo (MC) and data-driven methods. Diboson (WW, WZ, ZZ)
MC samples are normalized using the NLO
calculations from MCFM [31]. For \( \bar{t}t \) we use a produc-
tion cross section of 7.04 ± 0.7 pb [32], which is based
on a top-quark mass of 173 GeV/c^{2} [4] and MSTW 2008
NNLO PDFs [33]. The single-top-quark production cross
section is taken to be 3.15 ± 0.31 pb [34]. The normali-
zaion of the ZH+jets and W+jets MC samples is taken from
ALPGEN [35] corrected for NLO effects, except in
the case of the WH → \ell\nubb search. The normalization
of the W+jets MC sample in the WH → \ell\nubb search,
and normalization of the instrumental and QCD multi-
jet samples in all searches, are constrained from data
samples where the expected signal is several orders of
magnitude smaller than in the search samples.

All searches use the same data sample, which corre-
sponds to 9.45 fb\(^{-1}\) of integrated luminosity [36].
The analysis channels select non-overlapping subsets of the
data. Exactly two, one, or zero charged leptons are
required by the \( ZH \rightarrow \ell^{+}\ell^{-}bb, WH \rightarrow \ell\nu bb \), and
\( WH, ZH \rightarrow E_{T}bb \) event selections, respectively, where \( \ell \) denotes a reconstructed electron or muon. Both the
WH → \ell\nu bb and WH, ZH → E_{T}bb event selections re-
quire large \( E_{T} \) to be consistent with the signature of one or more high-pT neutrinos escaping the detector. Events
in all searches are required to contain exactly two or
three reconstructed jets. To optimize the sensitivity, the
data in each search are further divided into independent
sub-channels composed of differing jet multiplicity, lep-
ton quality, b-tag multiplicity, and b-tag quality. There
are 16, 26, and 3 sub-channels for the \( ZH \rightarrow \ell^{+}\ell^{-}bb \),
WH → \ell\nu bb, and WH, ZH → E_{T}bb analyses, respec-
tively, totaling to 45 for the combination presented here.
For a pair of jets, the dijet mass resolution for signal
events at CDF is expected to be 10–15% of their mean
mass in the range 90–150 GeV/c^{2}.

For a pair of jets, the dijet mass resolution for signal
events at CDF is expected to be 10–15% of their mean
mass in the range 90–150 GeV/c^{2}.

\begin{equation}
\mathcal{L}(R, s, b|\vec{n}, \vec{\beta}) \times \pi(\vec{\beta}) = \prod_{i=1}^{N_{C}} \prod_{j=1}^{N_{\text{jets}}} \int \frac{d^{3}p_{ij}}{n_{ij}} \exp \left( -\frac{\xi_{ij}^{2}}{\sigma_{ij}^{2}} \right) \times \prod_{k=1}^{N_{\text{jets}}} \exp \left( -\frac{\xi_{k}^{2}}{\sigma_{k}^{2}} \right) / 2.
\end{equation}
In this expression, the first product is over the number of channels ($N_C$), and the second product is over histogram bins containing $n_{ij}$ events, binned in ranges of the final discriminant variables for the individual analyses. The predictions for the bin contents are $\mu_{ij} = R \times s_{ij}(\vec{\theta}) + b_{ij}(\vec{\theta})$ for channel $i$ and histogram bin $j$, where $s_{ij}$ and $b_{ij}$ represent the expected SM signal and background in the bin, and $R$ is a scaling factor applied to the signal. By scaling all signal contributions by the same factor, we assume that the relative contributions of the different processes are as given by the SM.

Systematic uncertainties are parametrized by the dependence of $s_{ij}$ and $b_{ij}$ on $\vec{\theta}$. Each of the $n_{sys}$ components of $\vec{\theta}$, $\theta_k$, corresponds to a single independent source of systematic uncertainty, and each parameter may have an impact on several sources of signal and background in different channels, thus accounting for correlations. Gaussian priors are assumed for the $\theta_k$, truncated so that no prediction is negative. The likelihood function, multiplied by the $\theta_k$ priors, $\pi(\theta_k)$, is then integrated over $\theta_k$ including correlations [38],

$$ \mathcal{L}'(R) = \int \mathcal{L}(R, \vec{s}, \vec{b}|\vec{\theta}, \theta) \pi(\vec{\theta}) d\vec{\theta}. \quad (2) $$

We assume a uniform prior in $R$ to obtain its posterior distribution. The observed 95% C.L. upper limit on $R$, $R_{95}^{obs}$, satisfies $0.95 = \int_{0}^{R_{95}^{exp}} \mathcal{L}'(R) dR$. The expected distribution of $R_{95}$ is computed in an ensemble of pseudo-experiments generated without signal. In each pseudoexperiment, random values of the nuisance parameters are drawn from their priors. The median expected value of $R_{95}$ in this ensemble is denoted $R_{95}^{exp}$. A combined measurement of the cross section for Higgs boson production assuming SM branching ratios in units of the SM production rates is given by $R_{95}^{exp}$, which is the value of $R$ that maximizes $\mathcal{L}'$. The 68% C.L. interval (one standard deviation) is quoted as the shortest interval containing 68% of the integral of the posterior.

Though many sources of systematic uncertainty differ among the analyses, all correlations are taken into account in the combined limits, cross sections, and $p$-values. The uncertainties on the signal production cross sections are estimated from the factorization and renormalization scale variations, which includes the impact of uncalculated higher-order corrections, uncertainties due to PDFs, and the dependence on the strong coupling constant, $\alpha_s$. The resulting uncertainties on the inclusive $WH$ and $ZH$ production rates are 5% [10]. We assign uncertainties to the Higgs boson decay branching ratios as calculated in Ref. [39]. These uncertainties arise from imperfect knowledge of the mass of the $b$ and $c$ quarks, $\alpha_s$, and theoretical uncertainties in the $b\bar{b}$ decay rates. The largest sources of uncertainty on the dominant backgrounds in the $b$-tagged channels are the rates of $V$+heavy flavor jets, where $V = W$ or $Z$, which are typically 30% of the predicted values. In this case, the posterior uncertainties on these rates are typically 8% or less. Because Higgs boson production and decay branching ratios range from 2% to 6% and are applied to both signal- and MC-based background predictions. The uncertainty on the integrated luminosity of 6% arises from uncertainties in the luminosity monitor acceptance and the inelastic $pp$ cross section [40], and is assumed to be correlated between the signal- and MC-based background predictions.

To validate our background modeling and search methods, we additionally perform a search for SM diboson production in the same final states used for the SM $H \rightarrow b\bar{b}$ searches. The NLO SM cross section for $VZ$ times the branching fraction of $Z \rightarrow b\bar{b}$ is $682 \pm 50$ fb, which is comparable to the $410 \pm 20$ fb cross section times the branching fraction of $V(H \rightarrow b\bar{b})$ for a 100 GeV/$c^2$ SM Higgs boson. The data sample, reconstruction, background models, uncertainties, and subchannel divisions are identical to those of the SM Higgs boson search, but the discriminant functions are trained specifically for the signal of SM diboson production. The measured cross section for $VZ$ is $4.1 \pm 1.3$ pb (stat+syst), which is consistent with the SM prediction of $4.4 \pm 0.3$ pb and corresponds to a diboson signal significance of $\sim 3.2$ standard deviations.

To better visualize the data, we combine the histograms of the final discriminants, adding the contents of bins with similar signal-to-background ratio (s/b). Figure 1 shows the signal expectation and the data with the background subtracted, as a function of the s/b of the collected bins, for the diboson analysis described above and for the combined Higgs boson search, assuming $m_H = 125$ GeV/$c^2$. The background model has been fit to the data, and the uncertainties on the background are those after the nuisance parameters have been constrained in the fit. An excess of Higgs boson candidate events in the highest s/b bins relative to the background-only expectation is observed in Fig. 1.

We extract limits on SM Higgs boson production in the $m_H$ range of 90–150 GeV/$c^2$. We present our results in terms of $R_{95}^{obs}$, the ratio of the limits obtained to the rate predicted by the SM, as a function of the
Higgs boson mass. We assume the SM ratio for $WH$ and $ZH$ production. A value of $R_{95}^{obs}$ less than or equal to one indicates a SM Higgs boson mass that is excluded at the 95% C.L. These limits are shown, together with the median expected values and distributions of individual experiments assuming a signal is absent in Fig. 2. We also compute the best-fit rate parameter $R_{95}^{fit}$, which, when multiplied by the SM prediction for the associated production cross section times the decay branching ratio $(\sigma_{WH} + \sigma_{ZH}) \times B(H \rightarrow bb)$, yields the best fit values for this product. We show our fitted $(\sigma_{WH} + \sigma_{ZH}) \times B(H \rightarrow bb)$ as a function of $m_H$, along with the SM prediction, in Fig. 3.

Significances of excesses in data over the background prediction are computed by calculating the local background-only $p$-value using $R_{95}^{fit}$ as the test statistic. This $p$-value is the probability that $R_{95}^{fit}$ is equal to or exceeds its observed value, assuming a signal is truly absent. The Look-Elsewhere Effect (LEE) \cite{41,42} accounts for the possibility of a background fluctuation affecting the local $p$-value anywhere in the tested $m_H$ range, here taken to be from 115 to 150 GeV/$c^2$, owing to the prior exclusion \cite{43}. In this mass range, the reconstructed mass resolution is approximately 15-20 GeV/$c^2$. We therefore estimate that two independent outcomes are possible in these searches (LEE factor \approx 2). The $p$-value is computed for each $m_H$ in the range 90–150 GeV/$c^2$, and is shown in Fig. 1. Also shown are the expected values of the $p$-value assuming a SM signal is present, testing each value of $m_H$ in turn. The maximum local significance corresponds to 2.7 standard deviations at $m_H = 135$ GeV/$c^2$. Correcting for the LEE yields a global significance of 2.5 standard deviations.

In summary, we present a combination of CDF searches...
for the SM Higgs boson decaying to $b\bar{b}$ pairs using the entire Run II data sample. We search for a Higgs boson with a mass between 90 and 150 GeV/$c^2$, and exclude Higgs bosons with masses smaller than 96 GeV/$c^2$. The observed credibility limits are higher than those expected in the background-only hypothesis in the mass range 115–150 GeV/$c^2$. Within the currently non-excluded mass ranges, the lowest local $p$-value is found for a Higgs boson mass of 125 GeV/$c^2$, where the local significance of this deviation with respect to the background-only hypothesis is 2.7 standard deviations. At the same mass hypothesis, we measure an associated production cross section times branching ratio of $(\sigma_{WH} + \sigma_{ZH}) \times B(H \rightarrow b\bar{b}) = 291^{+118}_{−113}$ (stat + sys) fb.

This result is of fundamental interest both because similar searches are difficult at the LHC and because verification of a Higgs-boson-like particle decaying to $b$ quarks would offer a measurement of the $b$-quark Yukawa coupling, further establishing the mechanism of electroweak symmetry breaking as the source of fermionic mass in the quark sector. 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 World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; the Academy of Finland; and the Australian Research Council (ARC).
define $E_T = |\vec{E}_T|$. The transverse momentum $p_T$ is defined to be $p_T \sin \theta$.

[36] The luminosity quoted here is that for data collected after reconstruction-specific detector requirements specific to each subchannel are applied.