Search for the Higgs boson in the all-hadronic final state using the CDF II detector

We report on a search for the production of the Higgs boson decaying to two bottom quarks accompanied by two additional quarks. The data sample used corresponds to an integrated luminosity of 4 fb$^{-1}$ of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV recorded by the CDF II experiment. This search includes twice the data of the previous published result, uses new analysis techniques to distinguish jets originating from light flavor quarks and those from gluon radiation, and adds sensitivity to a Higgs boson produced by vector boson fusion. We find no evidence of the Higgs boson and place limits on the Higgs boson production cross-section for Higgs boson masses between 100 GeV/$c^2$ and 150 GeV/$c^2$ at 95% confidence level. For a Higgs boson mass of 120 GeV/$c^2$ the observed (expected) limit is 10.5 (20.0) times the predicted Standard Model cross-section.

The Higgs boson remains the only undiscovered particle of the standard model (SM) of particle physics. It is the physical manifestation of the mechanism which provides mass to fundamental particles [1, 2]. Direct searches at the LEP collider have excluded a Higgs boson mass $m_H < 114.4$ GeV/$c^2$ at 95% confidence level (CL) [3], while the Tevatron has excluded a Higgs boson mass between 163 GeV/$c^2$ and 166 GeV/$c^2$ at 95% CL [4]. The Tevatron has reported a preliminary update which extends the exclusion region for a Higgs boson mass between 158 and 175 GeV/$c^2$ [5]. Global fits to precision electroweak measurements set a one-sided 95% CL upper limit on $m_H$ at 157 GeV/$c^2$ [6].

This Letter presents the results of a search for the Higgs boson using an integrated luminosity of 4 fb$^{-1}$ of $p\bar{p}$ collision data at $\sqrt{s} = 1.96$ TeV recorded by the Collider Detector at Fermilab (CDF II). We search for a Higgs boson decaying to a pair of bottom-quark jets ($b\bar{b}$) accompanied by two additional quark jets (qq′). This search is most sensitive to a Higgs boson with low mass, $m_H < 135$ GeV/$c^2$, where the Higgs boson decay to $b\bar{b}$ is dominant [7]. The two production channels studied are associated production and vector boson fusion (VBF). The associated production channel is $p\bar{p} \rightarrow VH \rightarrow qq′b\bar{b}$, where $V$ is a W/Z vector boson, which decays to a pair of quarks. The hadronic branching ratio of $V$ to qq′ is $\approx 70\%$ [8]. In the VBF channel, $p\bar{p} \rightarrow qq′H \rightarrow qq′b\bar{b}$, the incoming partons each radiate a $V$ and the two $V$ fuse to form a Higgs boson.

Low-mass Higgs boson searches at CDF have concentrated on signatures that are a combination of jets, leptons and missing transverse energy which help to constrain the backgrounds but the signal yields are small [9, 11]. The hadronic modes used in this search exploit the larger branching ratio and thus have the largest signal yields among all the search channels at CDF. The major
challenge for this search is the modeling and suppression
of the large background from QCD multijets.

A previous Letter on the search for the Higgs boson
in the all-hadronic channel was published using an
integrated luminosity of $2 \text{fb}^{-1}$ [12]. This Letter has im-
proved the analysis sensitivity by a factor of two: a factor
of $\approx \sqrt{2}$ from doubling the analyzed data and a factor
of 1.4 from improvements to the analysis which are dis-
cussed in this Letter.

The CDF II detector is an azimuthally and forward-
backward symmetric detector designed to study $p \bar{p}$ col-
lisions and described in detail in [13–15] and references
therein. Jets are defined by a cluster of energy in the
calorimeter deposited inside a cone of radius $\Delta R = 
\sqrt{\Delta \phi^2 + \Delta \eta^2} = 0.4$ [16] as reconstructed by the JETCLU
algorithm [17]. Corrections are applied to the measured
jet energy to account for detector calibrations, multiple
interactions, underlying event and energy outside of the
jet cone [13].

The data for this search were selected by two multijet
triggers. The first 2.8 $\text{fb}^{-1}$ used a trigger which selected
four jet clusters with energy of at least 15 GeV and total
transverse energy $> 175 \text{GeV}$. This trigger was used in the
previous result [12]. The remaining 1.1 $\text{fb}^{-1}$ was recorded
with a new trigger which selected three jet clusters with
energy of at least 20 GeV and total transverse energy $> 
130 \text{GeV}$. The new trigger improves the acceptance for
low-mass Higgs by a factor of two.

After offline reconstruction, we select events with four
or five jets where each jet has $E_T > 15 \text{GeV}$ and $|\eta| < 2.4$.
The selected jets are ordered by descending jet-$E_T$ and
any fifth jet plays no further role in the search. The scalar
sum of the four leading jets is required to be $> 220 \text{GeV}$. 
Events with isolated leptons or missing transverse energy
significance $[19] > 6$ are removed to suppress the $t \bar{t}$ back-
ground.

The next stage of offline selection requires exactly two
of the four leading jets to be identified ("tagged") as
bottom-quark jets ($b$ jet). A $b$ jet is identified by its dis-
placed vertex, as defined by the SecVtx algorithm [14],
or by determining the tracks within the jet are unlikely to
have originated from the primary $p \bar{p}$ collision as defined
by the JetProb algorithm [20].

The two $b$-tagging algorithms are combined to form
two non-overlapping $b$-tagging categories: SS when both
jets are tagged by SecVtx, SJ when one jet is tagged by
SecVtx and the other by JetProb. For a jet tagged
by both algorithms, SecVtx takes precedence as it has
a lower rate of misidentifying a light flavor jet as a $b$
jet. The previous 2 $\text{fb}^{-1}$ search only included the SS cat-
gory [12]. The addition of the SJ category increased
the signal acceptance by 36%. Other $b$-tagging com-
binations, such as both $b$ jets selected by JetProb,
were not considered in this search as the relative gain in the
background is much larger than that for the signal.

The events which pass the offline selection are sepa-
rated into $VH$ or VBF signal regions. Both signal re-
regions are defined by the invariant masses of the two $b-
tagged$ jets, $m_{bb}$, and the remaining leading two $q$-jets,
$m_{qq}$. The division of events is based on the different kin-
ematics of the two processes. The $VH$ channel has two
mass resonances: $m_{bb}$ from the Higgs boson decay and
$m_{qq}$ from the $V$ decay. The $VH$ signal region is defined
as $75 < m_{bb} < 175 \text{GeV}/c^2$ and $50 < m_{qq} < 120 \text{GeV}/c^2$. 
The VBF channel shares the same $m_{bb}$ Higgs boson mass
resonance but there is no accompanying resonance for
$m_{qq}$. When compared to the background, a cut requir-
ing $m_{qq} > 120 \text{GeV}/c^2$ optimizes the VBF signal over
background ratio which exploits the large $\Delta \eta$ between
the $q$-jet pair. The VBF signal region is defined as
$75 < m_{bb} < 175 \text{GeV}/c^2$ and $m_{qq} > 120 \text{GeV}/c^2$. The
division of the events by their $b$-tagging categories and
their signal regions gives four analysis channels: $VH$-SS; 
$VH$-SJ; VBF$-$SS; VBF$-$SJ.

The $VH$ and VBF production are modeled with
PYTHIA [21] combined with a GEANT-based [22] simu-
lation of the CDF II detector [23]. The background is
dominated by QCD production of multi-jets which is
modeled by a data-driven technique developed in [12]
and described in more detail below. The non-QCD
backgrounds are modeled by simulation normalized to
next-to-leading order cross sections and are described
in detail in [12]. After applying the selection criteria
and correcting for the simulated trigger efficiency [12],
the expected yields for a Higgs boson of mass of
120 $\text{GeV}/c^2$ for the $VH$ and VBF signal regions are
7.8(SS)/2.9(SJ) and 3.2(SS)/1.2(SJ) events. The total
background for the $VH$ and VBF signal regions are
about 17000(SS)/9300(SJ) and 18000(SS)/9500(SJ)
events. The background composition is $\approx 98\%$ QCD and
the remaining 2% is a mixture of $t \bar{t}$, $Z$+jets, single-top,
$W$+jets and diboson events. We estimate the contribu-
tions of these backgrounds in the same way as in [12].

The large background precludes the use of simple vari-
ables, such as the dijet mass, to search for a Higgs boson
signal. We use an artificial neural network (NN) from
the TMVA package [24] to separate Higgs boson signal
events from the dominant QCD background and trained
at Higgs masses of 100, 120 and 140 $\text{GeV}/c^2$. As the kinematics
for $VH$ and VBF Higgs signals are different, a dedicated NN for each signal is trained. The input vari-
ables for the $VH$ neural net are $m_{bb}$, $m_{qq}$, the jet shapes
(explained below) of both $q$-jets, the cosine of the helicity
angle $\cos \theta^*_{q}\[25\]$, the cosine of the leading jet scattering
angle in the four jet rest frame $\cos \theta_{q}\[26\]$ and $\chi$ which is
a measure of whether both the $b$-jet pair and $q$-jet pair
are from a Higgs boson and $V$ decay respectively. $\chi$ is
defined as the minimum of $\chi_{W}$ and $\chi_{Z}$ where $\chi_{W}$ is
defined as $\chi_{W} = \sqrt{(M_W - M_{qq})^2 + (M_W - M_{bb})^2}$ and a similar expression exists for $\chi_{Z}$. For the VBF channel,
the neural net inputs are $m_{bb}$, $m_{qq}$ and the jet shape of
both $q$-jets. As the kinematics are not affected by the
different $b$-tagging categories, the neural net is trained with SS data as it has a better signal/background ratio. The NN distribution for the $VH$-SS channel, trained on simulated 120 GeV/$c^2$ Higgs boson events, is shown in Fig. 1. The NN returns a more negative (positive) score for background (signal) events.

![NN distribution for $VH$-SS](image)

**FIG. 1:** NN distribution for $VH$-SS trained on simulated 120 GeV/$c^2$ Higgs Boson events.

The QCD multi-jet background consists of a mixture of quark and gluon jets while the Higgs signal jets are mostly quark jets. One can make use of the fact that the gluon jets tend to be broader than light flavor quark jets to help separate the Higgs signal from the QCD multi-jet background [27]. The width of a jet is characterized by the jet $\phi$-moment $\langle \phi \rangle$ and jet $\eta$-moment $\langle \eta \rangle$ [28]. $\langle \phi \rangle$ is defined by Eq. (1) where the summation for the jet-moments are over the calorimeter towers used to form the jets and uses the tower-$E_T$ ($E_{t}^{\text{tower}}$), the jet-$E_T$ ($E_{t}^{\text{jet}}$), the tower’s $\phi$ position ($\phi_{\text{tower}}$) and the jet’s $\phi$ position ($\phi_{\text{jet}}$). A similar definition exists for $\langle \eta \rangle$.

$$\langle \phi \rangle = \frac{\sum_{\text{towers}} \frac{E_{t}^{\text{tower}}}{E_{t}^{\text{jet}}} (\Delta\phi(\phi_{\text{tower}}, \phi_{\text{jet}}))^2}{\sum_{\text{towers}} \frac{E_{t}^{\text{tower}}}{E_{t}^{\text{jet}}}}$$

(1)

The jet-width depends not only on the parton initiating the jet but also the jet-$E_T$, jet-$\eta$ and the number of reconstructed primary vertices. The variation of the jet-width measurement by these variables is removed by parameterizing their dependence and rescaling the jet-moment to a common reference value of jet-$E_T$=50 GeV, jet-$\eta$=0 and one reconstructed vertex.

The simulation of the jet shapes was verified using events from $tt \rightarrow b\ell\nu + bg’$, where $\ell$ are electrons or muons, as this provides a source of light-flavor quark jets in data. The event selection from [29] was used to select a data sample which is dominated by $tt$ ($\sim 86\%$) followed by $W+$jets, $Z+$jets, single top, diboson and backgrounds where light-flavor jets are misidentified as $b$ jets. The two highest $E_T$ untagged jets in the event whose dijet mass is consistent with a $W$ boson ($M_W = 80 \pm 30 \text{ GeV}/c^2$) are assumed to be quark jets. The jet shapes of these quark jets are compared to a simulation sample of $tt$ and $W+$jets in the same fraction as measured in data. An offset of $+0.0024 (+0.0015)$ was added to the simulation’s $\eta$-moment ($\phi$-moment) to agree with data. Half of the offset values were assigned as systematic uncertainties for the jet $\eta$ and $\phi$-moment. A cross-check using $Z+$jets simulation and data was performed and found to agree only after applying the corrections derived from $tt$.

The parameterizations used to rescale the jet-width to the common reference point for the simulation were cross-checked with the Higgs boson sample. The rescaled jet-widths for $VH$ were consistent with $tt$. However the rescaled jet-widths for VBF were lower than $tt$. This difference was considered as an additional systematic uncertainty for the VBF jet-width. Half of this maximum difference was used as a measure of the uncertainty: $\pm 0.0025 \pm 0.0010$ for the VBF jet $\langle \eta \rangle(\langle \phi \rangle)$.

A data-driven method, known as the tag rate function (TRF), is used to model the dominant QCD multijet background. The TRF is applied to events with at least one $b$-tagged jet (single tagged events) to predict the distribution of events with exactly two $b$-tagged jets (double tagged events). For each single-tagged event, the TRF gives the probability of each additional jet, called a probe jet, to be a second $b$-tagged jet. The TRF is described in detail in [12]. In this search the TRF was parameterized as a function of three variables: the probe jet-$E_T$, probe jet-$\eta$ and $\Delta R$ between the tagged $b$ jet and probe jet. The TRF is measured using jets in the TAG region (Fig. 2), defined as $m_{eq} < 45 \text{ GeV}/c^2$, $m_{bb} < 50 \text{ GeV}/c^2$ and $m_{bb} > 200 \text{ GeV}/c^2$. The TRF is then applied to the single tagged events in the $VH$ and VBF signal regions to predict the double tagged events.

![m_{bb} - m_{qq} plane](image)

**FIG. 2:** $m_{bb} - m_{qq}$ plane: This plane illustrates the regions used for the VH,VBF signal and the TRF.

The TRF predictions were verified by comparing the shapes of different variables constructed from single-
tagged events, after applying the TRF, and double tagged events from the TAG region. The TRF was able to model the shapes of all the NN training variables except $m_{qq}$ and the jet shapes $\phi$ and $\eta$ which required their own corrections. The prediction for these variables were scaled by the ratio of their observed shape to their prediction in the TAG region. These scaling corrections are applied in the $VH$ and VBF signal regions when predicting the shapes of double tagged events.

We consider two sources of systematic uncertainty on the shape of the NN output distributions for the multijet background. The interpolation uncertainty accounts for possible difference in the TRF between the regions where it was measured (TAG) and applied (SIGNAL). An alternative TRF was measured using events in the CONTROL region, as indicated in Fig. 2 which is still background-dominated. The difference in the shapes of the predicted background distribution in the NN output for the two TRFs is treated as a systematic uncertainty. The second source is due to the uncertainty in applying the $m_{qq}$ and jet shapes scaling factors measured in the TAG region to the SIGNAL region. Alternative scaling factors were derived using events in the CONTROL region.

The systematic uncertainties which affect the acceptance of the Higgs signal and non-QCD backgrounds are: ±7% jet energy scale uncertainty [18], ±2% parton distribution function uncertainty, b-tagging scale factor between simulation and data (±7.6% SS / ±9.7% SJ), ±2% $VH$ / ±3% VBF initial- and final-state radiation (ISR/FSR) uncertainties for the signal, ±4% trigger acceptance uncertainty and ±6% luminosity uncertainty [5]. The uncertainty on the cross-sections for the non-QCD backgrounds are: ±10% for $t\bar{t}$ and single-top [5], ±6% for diboson [5], ±50% for $W$+jets and $Z$+jets. In addition to changes to the acceptance, changes to the shape of the NN output were considered for the Higgs boson signal. The uncertainties which affected the shape of the NN output were jet-energy scale, ISR/FSR and jet-width. It should be noted the dominant systematic error for this analysis came from the uncertainties of the QCD prediction.

When the data are compared to the background prediction, we find no excess of events over the expected background and set upper limits on the excluded Higgs boson cross-section at 95% CL for $100 < m_H < 150$ GeV/$c^2$. The limits are calculated using a Bayesian likelihood method with a flat prior for the signal cross-section and Gaussian priors for the uncertainties on acceptances and backgrounds [20]. The normalization of the multijet background is a free parameter that is fit to the data. We combine the four search channels by taking the products of their likelihoods and simultaneously varying their systematic uncertainties.

For $m_H = 120$ GeV/$c^2$, the observed (expected) limit, normalized to the SM prediction, for the individual analysis channels are 11.9(25.6) for $VH$-SS, 43.4(51.8) for $VH$-SJ, 47.0(49.4) for VBF-SS, 93.7(132.3) for VBF-SJ, and 10.5(20.0) for the combination. The combined channel limits for Higgs boson masses in the range between 100 - 150 GeV/$c^2$ are shown Fig. 3.

The observed limits for the individual search channels are in agreement with their expected limits except for the $VH$-SS channel where there is a deficit in the data in the high signal region of the NN. As the $VH$-SS channel is the most sensitive, it has the strongest influence on the combined limit. Figure 4 shows the ratio of the data to background for the four analysis channels for a NN trained on a 120 GeV/$c^2$ Higgs boson simulated data. All four channels show a ratio $\approx 1$ over the whole NN output range, but the $VH$-SS channel has several points with a ratio of $\approx 0.9$ about a NN output of 0.5; the high signal region of NN output. The same TRF is used in the VBF-SS channel and the same NN is used in the $VH$-SJ channel, neither of which show such a feature. The low ratio for $VH$-SS is likely a statistical fluctuation rather than evidence of background mismodeling.

In summary, the measurement presented in this Letter shows a factor of two improvement over the previous 2 fb$^{-1}$ result for the all-hadronic Higgs search [12]. This Letter extends the 2 fb$^{-1}$ analysis by including the VBF channel, adding an additional algorithm to identify bottom-quark jets, adding an artificial neural network to separate signal from background which includes jet shapes to distinguish gluons from quark jets, and by doubling the analyzed data set. As the Tevatron continues to collect more data and further improvements to the analysis technique will extend the sensitivity of the all-hadronic Higgs search.

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
FIG. 4: Ratios of the data to background for \(VH\)-SS (a), \(VH\)-SJ (b), VBF-SS (c), and VBF-SJ (d) for the NN trained on 120 GeV/c^2 Higgs boson simulated events.

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 World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucléaire 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

[16] CDF uses a cylindrical coordinate system with the \(z\) axis aligned along the proton beam direction, \(\theta\) is the polar angle relative to the \(z\)-axis and \(\phi\) is the azimuthal angle. The pseudorapidity is \(\eta = -\ln(\tan(\theta/2))\). The transverse energy is \(E_T \equiv E \sin \theta\).
[19] Missing transverse energy significance is the ratio of the total missing transverse energy to the total transverse energy.
[25] \(\cos \theta_q\) is the cosine helicity angle of the leading non \(b\) jet \((q_1)\). The helicity angle \(\theta_q^\ast\) of the leading non \(b\) jet \(q_1\) is defined to be the angle between the momentum of \(q_1\) in the \(q_1 - q_2\) rest frame and the total momentum of \(q_1 - q_2\) in the lab frame.
[26] \(\cos \theta_3\) is defined in a three jet rest frame as the cosine of the leading jet scattering angle. We reduce from four jets to three jets by combining the two jets with the lowest two jet mass. Thus \(\cos \theta_3 = \rho_{AV}^\ast \rho_S^\ast / (\rho_{AV} |P_S|)\) where \(\rho_3\) is the third jet and \(P_{AV}\) is the vector sum of the three jets in the lab frame [31].