Search for Neutral Higgs Bosons in Multi-$b$-Jet Events in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV


0031-9007/08/101(22)/2218202 (2008) PHYSICAL REVIEW LETTERS week ending 28 NOVEMBER 2008

(D0 Collaboration)

1Universidad de Buenos Aires, Buenos Aires, Argentina
2LAFAEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil
3Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
4Universidade Federal do ABC, Santo André, Brazil
5Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil
6University of Alberta, Edmonton, Alberta, Canada,
7Simon Fraser University, Burnaby, British Columbia, Canada,
8York University, Toronto, Ontario, Canada,
9University of Wuppertal, Wuppertal, Germany
10Panjab University, Chandigarh, India
11III. Physikalisches Institut A, RWTH Aachen, Aachen, Germany
12Physikalisches Institut, Universität Bonn, Bonn, Germany
13Physikalisches Institut, Universität Freiburg, Freiburg, Germany
14Institut für Physik, Universität Mainz, Mainz, Germany
15Ludwig-Maximilians-Universität München, München, Germany
16Fachbereich Physik, Universität of Wuppertal, Wuppertal, Germany
17Delhi University, Delhi, India
18Tata Institute of Fundamental Research, Mumbai, India
19FOM-Institute NIKHEF and University of Amsterdam/NIKHEF, Amsterdam, The Netherlands
Data recorded by the D0 experiment at the Fermilab Tevatron Collider are analyzed to search for neutral Higgs bosons produced in association with b quarks. This production mode can be enhanced in the minimal supersymmetric standard model (MSSM). The search is performed in the three b quark channel using multijet triggered events corresponding to an integrated luminosity of $1 \text{fb}^{-1}$. No statistically significant excess of events with respect to the predicted background is observed and limits are set in the MSSM parameter space.

DOI: 10.1103/PhysRevLett.101.221802  PACS numbers: 14.80.Cp, 12.38.Qk, 12.60.Fr, 13.85.Rm
Supersymmetry (SUSY) [1] is a popular extension of the standard model (SM), requiring the presence of at least two Higgs doublets. In the minimal supersymmetric standard model (MSSM) five Higgs bosons remain after electroweak symmetry breaking: three neutral (h, H, and A, collectively denoted as \( \Phi \)) and two charged (\( H^{\pm} \)). The Higgs sector can be parameterized by \( \tan \beta \), the ratio of the vacuum expectation values of the two Higgs doublets, and \( m_A \), the mass of the pseudoscalar A. For large values of \( \tan \beta \) two of the three neutral Higgs bosons have approximately the same mass and couplings to down-type quarks, which are enhanced by a factor \( \tan \beta \) relative to the SM ones, while the couplings to up-type quarks are suppressed. More precisely, the three Higgs boson couplings to bottom quarks follow the sum rule \( g_{hbb}^2 + g_{Hbb}^2 + g_{Abb}^2 = 2\tan^2 \beta g_{hsm}^2 \) where \( g_{hsm}^2 \) is the respective SM coupling. As well as increased production, the \( \tan \beta \) enhancement means that the main decay mode is \( \Phi \rightarrow b\bar{b} \), with a branching fraction, \( \mathcal{B}(\Phi \rightarrow b\bar{b}) \), of \( \approx 90\% \). Consequently, in \( p\bar{p} \) collisions at \( \sqrt{s} = 1.96 \) TeV at the Fermilab Tevatron Collider, the final state with at least three \( b \) jets is an important channel in the search for MSSM Higgs bosons at large \( \tan \beta \). However, the very large multijet background at a hadron collider makes the search for this topology challenging.

LEP excluded \( m_{h,A} < 93 \) GeV/c\(^2 \) for all \( \tan \beta \) values [2]. CDF [3,4] and D0 [5,6] have extended the MSSM Higgs boson searches to higher masses for high \( \tan \beta \) values. The result presented in this Letter supersedes our previous published result [5]. In addition to including more data, this analysis benefits from improved signal and background modeling and an improved limit setting procedure, which uses only the shape, and not the normalization, of the final discriminating variable.

The D0 detector is described in Ref. [7]. Dedicated triggers designed to select events with at least three \( b \) jets are used in this analysis. Typical trigger requirements are at least two \( b \) jets with transverse momenta \( p_T > 25 \) GeV/c and at least one additional \( b \) jet with \( p_T > 15 \) GeV/c. Algorithms for identifying \( b \) jets at the trigger level are also employed in about 70% of the integrated luminosity used for this analysis. After data quality requirements the total data sample corresponds to \( 1.02 \pm 0.06 \) fb\(^{-1} \) [8].

Signal samples are generated for Higgs boson masses from 90–220 GeV/c\(^2 \) using the leading order PYTHIA event generator [9] to generate associated production of \( \Phi \) and a \( b \) quark in the 5-flavor scheme, \( gb \rightarrow \Phi b \). Weights, calculated with MCFM [10], are applied to the signal samples as a function of \( p_T \) and \( \eta \) of the highest \( p_T b \) jet which is not from the decay of the Higgs boson to correct the cross section and experimental acceptance to next-to-leading order. Multijet background events from the \( bb, b\bar{b}, b^j, b\bar{b}^j, c\bar{c}, c\bar{c}^j, c\bar{c}^j, b\bar{c}c \), and \( bb\bar{b}b \) processes (where \( j \) denotes a light parton: \( u, d, s \) quark or gluon) are generated with the ALPGEN [11] event generator. The contributions from other processes, such as \( t\bar{t}, Zb\bar{b}, \) and single top production, are found to be negligible. The ALPGEN samples are processed through PYTHIA for showering and hadronization. All samples are then processed through a GEANT-based [12] simulation of the D0 detector and the same reconstruction algorithms as the data. A parameterized trigger simulation is used to model the effects of the trigger requirements on the simulated events.

Jets are reconstructed from energy deposits in calorimeter towers using the midpoint cone algorithm [13] with radius = 0.5. Jet reconstruction and energy scale determination are described in detail in Ref. [14]. All calorimeter jets are required to pass a set of quality criteria and have at least two reconstructed tracks within \( \Delta R(\text{track}, \text{jet-axis}) = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} < 0.5 \) (where \( \eta \) is the pseudorapidity and \( \phi \) the azimuthal angle), for a total per-jet efficiency of 83%–93%.

We select signal events for which the \( p\bar{p} \) interaction vertex is reconstructed well within the geometric acceptance of the silicon detector. We further require at least three and at most five jets with \( p_T > 20 \) GeV/c and \( |\eta| < 2.5 \). A neural network based \( b \)-tagging algorithm [15], with lifetime based information involving the track impact parameters and secondary vertices as inputs, is used to identify \( b \) jets. Each event must have at least three \( b \) jets satisfying a tight \( b \)-tag neural network requirement. This requirement provides \( \approx 50\% \) efficiency for tagging a single \( b \)-jet at \( \approx 0.4\% \) mistag rate of light jets (\( u, d, s \) quarks or gluons). The events with at least two tight \( b \) tags are also kept and used to model the background, since the relative amount of signal is negligible in this double tagged sample. Simulated events are weighted based on their tagging and fake rate probabilities determined from data. Finally, the transverse momenta of the two highest \( p_T \) jets which are also \( b \) tagged are required to be above 25 GeV/c. To further increase the sensitivity, the analysis is split into separate three-, four-, and five-jet channels. After the event selection 3224 events remain in the exclusive three-jet sample, 2503 and 704 events in the four- and five-jet samples, respectively. The signal efficiencies for Higgs boson masses between 100 and 200 GeV/c\(^2 \) range from 0.3%–1.2% in the three-jet channel (0.2%–0.6% and 0.03%–0.12% in the four- and five-jet channels).

The background composition is determined separately for each jet multiplicity. By considering eight different \( b \)-tagging criteria, each applied to three subsamples (exactly one, two and three \( b \)-tagged jets, respectively), a linear system of equations can be solved to determine the relative contribution of the different processes since the \( b, c \) and light jet tagging efficiencies are known. The double \( b \)-tagged (2Tag) sample is found to be dominated by \( bbj \) while the triple \( b \)-tagged (3Tag) sample consists of a mix of \( \approx 50\% bb\bar{b}, \approx 30\% b\bar{b}j, \text{ and } \approx 20\% bcc + b\bar{c}c \). An alternative method to determine the background, based on
fitting simulated \( H_T = \sum p_{Tjet} \) shape templates to the data, confirms the composition of the background.

For every event the two jet pairs with the largest summed transverse momenta are considered as possible Higgs boson candidates. To remove discrepancies between data and simulation originating from gluon splitting \((g \to b\bar{b})\), only jet pairs with \( \Delta R > 1.0 \) are considered in the final analysis. The agreement between data and simulation was verified in several control samples after the selection criteria.

The following six variables separate the Higgs boson jet pair from the background jet pairs and are well modeled by the simulation: the difference in pseudorapidity between the two jets in the pair; the azimuthal angular difference between the two jets in the pair; the angle between the leading jet in the pair and the total momentum of the pair; \( |p_1 - p_2|/|p_1 + p_2| \), the momentum balance, where \( p_i \) \((i = 1, 2)\) are the momenta of the jets in the pair; the combined rapidity of the jet pair; and the event sphericity \([[16], \) defined as \( \sqrt{e_1 + e_2} \), where \( e_n \) is the \( n \)th eigenvalue of the normalized momentum tensor calculated using all jets in the event. Based on these kinematic variables, a likelihood discriminant \( D \), is calculated according to:

\[
D(x_1, \ldots, x_6) = \frac{\prod_{i=1}^{6} P_{i}^{\text{sig}}(x_i)}{\prod_{i=1}^{6} P_{i}^{\text{bkg}}(x_i) + \prod_{i=1}^{6} P_{i}^{\text{bkg}}(x_i)},
\]

where \( P_{i}^{\text{sig}} \) (\( P_{i}^{\text{bkg}} \)) refers to the signal (background) probability density function (PDF) for variable \( x_i \), and \((x_1, \ldots, x_6)\) is the set of measured kinematic variables for the jet pair. The PDFs are obtained from the 3Tag signal and background simulation. Two likelihoods are built combining simulated samples in the 90–130 GeV/\( c^2 \) (“low-mass”) and 130–220 GeV/\( c^2 \) (“high-mass”) mass ranges, providing discrimination at low and high masses, respectively. Studies show that this division of the mass range gives the best discrimination.

Several multijet processes contribute to the background and the uncertainty on the cross sections is large. The \( bbb \) component may also contain a contribution that is indistinguishable from a signal and cannot be normalized from the data. To model the background we therefore rely on a combination of data and simulation. The distribution of the expected background in the 3Tag sample is obtained, in the two-dimensional \( D \) and invariant mass \((M_{bb})\) plane, by:

\[
S_{\text{3Tag}}^{\text{exp}}(D, M_{bb}) = R^{\text{MC}}(D, M_{bb}) S_{\text{2Tag}}^{\text{data}}(D, M_{bb}),
\]

where \( R^{\text{MC}} = S_{\text{3Tag}}^{\text{MC}} / S_{\text{2Tag}}^{\text{MC}} \). \( S_{\text{2Tag}}^{\text{data}} \) is the 2Tag data shape. Many uncertainties affecting the simulation cancel in the ratio \( R^{\text{MC}} \). Figure 1 shows the low-mass likelihood for data and background in the three-jet channel.

The selection cuts on \( D \), \( b \) tagging, and the number of jet-pair combinations per event are optimized by maximizing the expected sensitivity. The optimal cuts for the likelihood vary between 0.25 and 0.60 depending on the jet multiplicity and Higgs boson mass. The agreement of the data and the background expectation is verified in a control region where the impact of any Higgs boson signal is limited, defined by \( D < 0.25 \). Figure 2 shows the invariant mass for the optimized high-mass likelihood cuts.

The number of signal events depends on several measurements which introduce a systematic uncertainty: integrated luminosity (6.1%), theoretical uncertainty (12%–14%), trigger efficiency (2%–5%), jet identification (0.3%–0.5%), jet energy calibration (3%–5%), jet energy resolution (0.1%–0.6%), and \( b \) jet identification (8%–9%). The size of these uncertainties depends on the Higgs boson mass and the number of jets. Several sources of systematic uncertainties affecting the background shape through the ratio \( R^{\text{MC}} \) in Eq. (2) are parametrized as a function of \( M_{bb} \).

The dominant uncertainty, due to the background composition, is estimated by varying the ratio of \( bbj \) and \( bbb \) events in the sample corresponding to the uncertainties from the background composition fit. The uncertainty from the kinematic dependence of the \( b \) tagging of jets is evaluated by varying the light, \( b \), and \( c \) jet tag efficiencies within their uncertainties. The uncertainty from the \( b \)-jet energy resolution is obtained by smearing the simulated \( b \) and \( c \) jets by an additional factor of half the uncertainty of the light jet energy resolution. The effect due to the uncertainties in the kinematic modeling of \( bbb \) and \( bbj \) is estimated by replacing the \( bbj \) Monte Carlo events with the 2Tag data and taking half the resulting shape difference as the uncertainty. Finally, the small shape difference between 3Tag and 2Tag data in the turn-on of the trigger level \( b \) tag is accounted for as a systematic uncertainty.

The modified frequentist method \([17] \) is used to estimate \( 1 - CL_b \), the probability for a sample of background only.
to appear more signal-like than the observed data, as well as to derive limits at the 95% confidence level on the cross section times branching ratio as a function of $m_A$. As the absolute level of the multijet background cannot be reliably determined, only the shapes (not the normalization) of the $M_{bb}$ distributions are used in the limit setting.

The systematic uncertainties on the signal and on the background shape are included in the limit calculation. Table I shows the limits and the $1 - CL_b$ values obtained versus the hypothesized Higgs boson mass, assuming the width of $\phi$ to be negligible relative to the experimental resolution ($\approx 20\%$). The low $1 - CL_b$ values around a Higgs boson mass of 180 GeV/$c^2$ are due to a slight excess over the expected SM background.

The results of this search can be used to set limits on the parameters of the MSSM. As a consequence of the enhanced couplings to $b$ quarks at large $\tan\beta$ the total width of the neutral Higgs bosons also increases with $\tan\beta$. This can have an impact on our search if the width is comparable to or larger than the experimental resolution of the reconstructed invariant mass of a di-jet system. To take this effect into account, the width of the Higgs boson is calculated with FEYNHIGGS [18] and included in the simulation as a function of the mass and $\tan\beta$ by convoluting a relativistic Breit-Wigner function with the next-to-leading order cross section. In the MSSM the masses and couplings of the Higgs bosons depend, in addition to the SUSY parameters through radiative corrections. Limits on $\tan\beta$ as a function of $m_A$ are derived for two particular scenarios assuming a CP-conserving Higgs sector [19]: the $m_h^{\text{max}}$ scenario [20] and the no-mixing scenario [21]. Since the results depend considerably upon the Higgs sector bilinear coupling $\mu$, its two possible signs are also probed.

Figure 3 shows the results obtained in the present analysis interpreted in these different MSSM scenarios. Substantial areas in the MSSM parameter phase space up to masses of 200 GeV/$c^2$ are excluded. No exclusion can be obtained for the $m_h^{\text{max}}$, $\mu > 0$ scenario, due to the

![FIG. 2 (color online). Invariant mass for the high-mass likelihood region for the exclusive (a) three-jet (b) four-jet, and (c) five-jet channels. Black crosses refer to data, the solid line shows the total background estimate, and the shaded region represents the heavy flavor component ($b\bar{b}$, $bc$, and $c\bar{c}b$). The lower panels show the difference between the data and the background expectation. Only statistical errors are shown.](image1)

![FIG. 3 (color online). 95% C.L. exclusion limits in the ($m_A$, $\tan\beta$) plane for $m_h^{\text{max}}$, $\mu = -200$ GeV, and no-mixing, $\mu = -200$ GeV and $\mu = +200$ GeV. The exclusions from LEP are also shown [2]. The width of $\phi$ is larger than 70% of $m_A$ above $\tan\beta = 100$ in the $m_h^{\text{max}}$, $\mu = -200$ GeV scenario.](image2)

<table>
<thead>
<tr>
<th>Mass (GeV/$c^2$)</th>
<th>$\sigma \times B$ (Expected pb)</th>
<th>$\sigma \times B$ (Observed pb)</th>
<th>$1 - CL_b$ (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>$170^{+72}_{-52}$</td>
<td>184</td>
<td>39</td>
</tr>
<tr>
<td>100</td>
<td>$117^{+148}_{-35}$</td>
<td>128</td>
<td>38</td>
</tr>
<tr>
<td>110</td>
<td>$71^{+29}_{-20}$</td>
<td>69</td>
<td>52</td>
</tr>
<tr>
<td>120</td>
<td>$41^{+18}_{-9}$</td>
<td>34</td>
<td>73</td>
</tr>
<tr>
<td>130</td>
<td>$28^{+12}_{-6}$</td>
<td>24</td>
<td>70</td>
</tr>
<tr>
<td>140</td>
<td>$25^{+11}_{-6}$</td>
<td>22</td>
<td>60</td>
</tr>
<tr>
<td>160</td>
<td>$17^{+8}_{-4}$</td>
<td>26</td>
<td>12</td>
</tr>
<tr>
<td>180</td>
<td>$13^{+5}_{-4}$</td>
<td>23</td>
<td>4.4</td>
</tr>
<tr>
<td>200</td>
<td>$9^{+4}_{-3}$</td>
<td>17</td>
<td>7.0</td>
</tr>
<tr>
<td>220</td>
<td>$7^{+3}_{-2}$</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>
decrease in $\sigma \times B$ for positive values of $\mu$ [19]. With the restriction of only using the shape of the discriminant variable, this analysis puts stringent limits in the $(m_\text{A}, \tan \beta)$ plane up to Higgs boson masses of 220 GeV/c$^2$, particularly in the $m_\text{h}^{\text{max}}$, $\mu < 0$ scenario.

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC (United Kingdom); MSMT and GACR (Czech Republic); CRC Program, CFI, NSERC and WestGrid Project (Canada); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); CAS and CNSF (China); and the Alexander von Humboldt Foundation.

*Visitor from Augustana College, Sioux Falls, SD, USA.
†Visitor from The University of Liverpool, Liverpool, UK.
‡Visitor from ICN-UNAM, Mexico City, Mexico.
§Visitor from II. Physikalisches Institut, Georg-August-University, Göttingen, Germany.
‖Visitor from Helsinki Institute of Physics, Helsinki, Finland.
¶Visitor from Universität Zürich, Zürich, Switzerland.
**Deceased.

[20] $M_{\text{SUSY}} = 1 \text{ TeV}$, $X_t = 2 \text{ TeV}$, $M_2 = 0.2 \text{ TeV}$ $|\mu| = 0.2 \text{ TeV}$, and $m_t = 0.8 \text{ TeV}$.
[21] $M_{\text{SUSY}} = 2 \text{ TeV}$, $X_t = 0 \text{ TeV}$, $M_2 = 0.2 \text{ TeV}$ $|\mu| = 0.2 \text{ TeV}$, and $m_t = 1.6 \text{ TeV}$.