CDF/ANAL/EXOTIC/PUBLIC/8442
August 3, 2006

Search for the Standard Model Higgs boson in the $\not\!\!E_T$ and B-jets Signature in $p\overline{p}$ Collisions at $\sqrt{s} = 1.96\ TeV$

The CDF Collaboration

Abstract

We present a search for the Standard Model Higgs boson produced in association with a Z or W boson in $p\overline{p}$ collisions at $\sqrt{s} = 1.96\ TeV$. We focus our search on the $H \rightarrow b\overline{b}$, and $Z \rightarrow \nu\overline{\nu}$ or $W \rightarrow \not{l}\nu$ (where the charged lepton escapes detection) decay modes. The selected events shall have exactly two jets in the final state, one or two b-tagged jet(s), and large missing transverse momentum due to the undetected leptons. The exclusive single and the double tag events are treated separately througout the analysis. We process 0.973 fb^{-1} data collected in CDF Run II. The observed number of events is consistent with the Standard Model background prediction within the errors. 95% C.L. upper limits are set on the SM Higgs production cross-section in the ZH, WH, and combined channels at various Higgs masses. The observed cross-section limits divided by the Standard Model cross-section range from 17.8 to 22.8 between 110 GeV/c^2 and 130 GeV/c^2 of Higgs masses. The observed limits are also consistent with the Monte Carlo expectations.

The Higgs mechanism is the cornerstone of the Standard Model (SM). The fact that the Standard Model has been proved successful in many instances supports the existence of the yet unobserved Higgs boson [1]. According to the latest results from LEP2, the lower bound on the mass of the Higgs is $m_H \geq 114\ GeV/c^2$ at a 95 % C.L. [2] The previous results of the CDF Higgs searches from Run I was published in [3]. In the Run I analysis, the available luminosity did not present enough sensitivity for observation or exclusion beyond the LEP limit. With the improved luminosity of Run II, we expect to have the sensitivity to observe or exclude the Higgs boson at the current lower mass limit [4, 5]. In this analysis, we use 0.973 fb^{-1} of data that was available in the beginning of 2006.

In this paper, we present the results of a search for the Higgs in the case when the Higgs is created in association with a Z or W boson, and when the decay modes of the two bosons are $H \to b\bar{b}$ and $Z \to \nu\bar{\nu}$ or $W \to \not{l}\nu$ (where the charged lepton escapes detection) respectively. These processes lead to two b-jets. Since the two leptons escape the detector without being observed, a large amount of transverse momentum appears missing from the signal events. This is quantified by the Missing Transverse Energy ($\not{E}_T$), the imbalance in the total electromagnetic and hadronic calorimeter energy projected to the transverse plane.

The analysis strategy is greatly influenced by the performance of the trigger system of the CDF detector. Based on the signal characteristics, all recorded events shall contain two jets above 20 GeV one of which is central and $\not{E}_T$ above 35 GeV. In the first step of the analysis, both the Monte Carlo and real data events are to pass a set of quality cuts to ensure that the possible beam and detector effects are removed from the data sample making it compatible with the simulation. The efficiency of the two-jet requirement is 97% if the offline transverse energy of the most energetic jet is above 35 GeV, the second most energetic jet is above 20 GeV, and one of the jets is within ± 1.1 pseudo-rapidity range (central jet). The $\not{E}_T$ selection is more complicated to handle, because the $\not{E}_T$ measured online depends on the physical process in the collision. In multi-jet events without isolated leptons, the $\not{E}_T$ is due to the mismeasurement of the transverse energy of the jets; whereas, in certain electroweak events, it is dominated by the missing transverse momentum of undetected particles. In the current analysis, we reduce this effect by requiring the offline corrected $\not{E}_T$ to be > 55 GeV. The overall efficiency of the online event selection is then parametrized by the offline corrected $\not{E}_T$ and applied on the Monte Carlo samples providing a proper scaling for the simulated events.

The final requirement imposed on the data and simulation before comparing them is the b-tag requirement. We use a standard b-tagging algorithm, called SECVTX [6], developed by the High-pt b-tag group in CDF. The algorithm searches for secondary vertices formed at the intersection of two or three displaced tracks, which are associated to jets originally reconstructed based on the calorimeters. The ratio of the displacement of the secondary vertex with respect to the primary to the measurement uncertainty of the displacement, the significance of the displacement of the secondary vertex, must be greater than 3. If such a secondary vertex is found, the associated jet is considered

to originate from a b-hadron, and thus becomes tagged. When the displacement is negative (positive), the tag is called negative (positive) tag. Negative tags do not carry physical meaning, they are due to detector resolution effects. The frequency of their occurance is assumed to be proportional to the rate of b-jet misidentification. In this analysis, we require at least one jet per event be positively tagged, have rapidity $|\eta| < 2.4$, and have corrected $E_T > 20$ GeV. Both jets are required to have an electromagnetic fraction greater than 0.1 and less than 0.9.
The summary of the basic selection cuts is as follows:

- Exactly two jets with $E_T^{Corr} > 20$ GeV, $|\eta| < 2.4$, and 0.1 <EM Fract.< 0.9
- One jet with $E_T^{Corr} > 35$ GeV, and one central jet ($|\eta| < 1.1$)
- ${\not\!\!E_T}^{Corr} > 55$ GeV
- 1 or 2 positive b-tag(s) ($E_T^{Corr} > 20$ GeV) in the single/double tag sample

The most significant background at this stage of the analysis is represented by flavor production via gluon splitting. Multi-jet (QCD) production has a large cross-section on the order of a μb, which is about 9 orders of magnitude greater than the signal. Although, these processes generally do not have intrinsic $\not\!\!E_T$ and heavy flavor jets, mismeasured jets can cause imbalance in the total transverse energy due to which these events can pass the basic selection cuts if one of the jets is mis-tagged. B-quark pair production yields taggable jets, furthermore a b-quark undergoing a semi-leptonic decay creates large $\not\!\!E_T$. In both cases, the missing transverse energy tends to be aligned parallel or anti-parallel to the first or second most energetic jet. Vetoing events with such topology is the most effective tool against the QCD background.

Due to the large cross-section, it is practically impossible to generate enough statistics to simulate all QCD processes. In this analysis, we simulate processes which yield real taggable objects, that is, when a b- or a c-quark pair is created. Events with light flavor jets, after requiring a positive tag in the event during the basic selections, are considered to be mistags and are estimated from the data. Consequently, a tagged jet in the simulation is considered to have a real tag only if there is a heavy flavor quark inside the R=0.4 radius cone of the jet to avoid double-counting the mistags. The QCD sample is scaled by a factor such that the total number of simulated and mistag events becomes equal to the number of data events passing the basic selection cuts. (Single and double-tag QCD events are scaled separately.)

Two classes of top-production are considered in this analysis. The pair-production [6, 7] and the single top-production in the t- and s-channels [8]. They both yield a significant contribution to the background in the signal region. In addition, the presence of the single top background becomes more pronounced after requiring only two hard jets in the event. Due to the large mass and the semi-leptonic decay of the top, these events are energetic, bear large $\not\!\!E_T$ and high jet multiplicity. The top quark production is estimated with Pythia Monte Carlo samples normalized to the NLO cross sections ($\sigma_{t\bar{t}} = 6.7^{+0.7}_{-0.9}pb$, $\sigma_{W^\star \to tb} = 0.88 \pm 0.11pb$, and $\sigma_{gW \to tb} = 1.98 \pm 0.25pb$) for a top quark mass of 175 GeV/c^2.

The electroweak backgrounds were all generated by Pythia. In the di-boson samples, the bosons' decays are inclusive. In the W/Z + jets samples, the bosons are forced to decay into leptons, or b-quarks. The electroweak backgrounds thus include the following processes: W to leptons + h.f., Z to leptons + h.f., Z to h.f. jets, WW/WZ/ZZ inclusive decays.

The mistagged light flavor multi-jet processes constitute the largest SM background. We estimate this component from the data [6]. The method is based on the assumption that positive mistags are due to resolution effects through the same mechanism as the negative tags; therefore, the positive mistag rate is proportional to the negative one. The number of negative tags are estimated from the jets in the data, the tag-rate of which is parametrized by their E_T, η and, φ in an event with a given $\sum E_T$ and N^{good}_{trk}. The proportionality constant between the positive mistags and the negative tags (asymmetry factor) was measured to be 1.36±0.23.

After defining two control regions in the events passing the basic selection criteria mentioned above, the simulated SM background is compared to the data. In the first control region all the leptons are vetoed from the events, and the azimuthal angular separation between the second leading jet and the $\not\!\!E_T$, $\varphi(2^{nd}jet, \not\!\!E_T)$, is less than 0.4. This control region is dominated by QCD multi-jet events, and is used to corroborate the shapes and normalization of the b-filtered Pythia Monte Carlo sample with respect to the data in the control region. The normalization is needed to correct for the NLO processes, and to compensate for the tagged c-jet events, which are not entirely included in the simulation. The second control region, by definition, contains at least one lepton or isolated track, and $\varphi(2^{nd}jet, \not\!\!E_T) > 0.4$. In this region, the distributions of the sum of all the background components are compared with the data, furthermore, the overall normalization of the Monte Carlo events is corroborated. (The electroweak and top events are normalized to the luminosity.) The double-tag events in this control region are dominated by the top processes, yielding an extra handle in the cross-check. After achieving a good agreement between the simulation and the data in the control regions (Table 1), a set of cuts are selected that are optimized for the signal Monte Carlo against the background prediction.
The optimization is composed of four consecutive cuts. The first three are independent of the simulated Higgs mass:

- Azimuthal separation between leading jet and $\not\!\!E_T$: $\Delta\phi(1^{st}jet, \not\!\!E_T) > 0.8$ rad
- $\frac{\not\!\!H_T}{H_T} > 0.45$, where H_T is calculated using the two most energetic jets in the event. (H_T is the scalar and $\not\!\!H_T$ is the magnitude of the vectorial sum of the p_T of the jets.)
- E_T of leading jet: $1^{st}jet\ E_T > 60$ GeV
- $\not\!\!E_T > 75$ GeV

Table 2 lists the number of predicted and observed events in the signal region after applying these cuts. Figure 1 shows the dijet invariant mass in the Signal Region for single and double-tag events.

Process	**CR 1**		**CR 2**	
	=1 Tag	**2 Tags**	**=1 Tag**	**2 Tags**
QCD multi-jet	9833 ± 1091	688 ± 68	58 ± 31	2.7 ± 1.9
Mistag	3081 ± 524	257 ± 44	135 ± 23	7.0 ± 1.2
TOP	3.8 ± 0.5	0.7 ± 0.2	100.2 ± 14.2	25.6 ± 5.2
W + h.f.	10.7 ± 2.5	0.8 ± 0.6	93 ± 20	6.8 ± 2.4
Z + h.f.	22.9 ± 5.3	5.3 ± 2.3	16.8 ± 3.5	1.5 ± 0.5
Diboson	0.9 ± 0.2	0.03 ± 0.02	12.6 ± 2	0.86 ± 0.2
Total Expected BG	12952 ± 1212	952 ± 81	416 ± 52	44.5 ± 7.3
Observed	13020	974	482	40

Table 1: Number of expected events in the control regions after basic selection cuts

Process	**Signal Region =1 Tag**	**Signal Region 2 Tags**
QCD multi-jet	104 ± 25	3.5 ± 1.7
Mistag	110 ± 19	4.8 ± 0.9
TOP	28 ± 4	5.5 ± 1.1
W + h.f.	36 ± 8	2.9 ± 1.4
Z + h.f.	24 ± 5	2.3 ± 0.6
Diboson	7.1 ± 1.1	0.6 ± 0.1
ZH (115 GeV)	0.77 ± 0.09	0.25 ± 0.05
WH (115 GeV)	0.64 ± 0.07	0.22 ± 0.04
Total Expected BG	310 ± 40	19.6 ± 3.1
Observed	333	24

Table 2: Number of expected events in the single and double tag signal regions after applying the optimized cuts

The systematic uncertainties are classified as correlated and uncorrelated errors considering the relations between the signal and the background processes. The correlated errors are taken into account separately for each processes in the limit calculation. The uncorrelated systematic uncertainties are: statistical error in negative tag estimate, negative-positive tag rate asymmetry factor (17%), NLO MC corrections (12%), QCD multi-jet Monte Carlo normalization (2%), MC statistical fluctuations. The correlated systematics are: luminosity (5.8%), b-tagging efficiency scale factor between data and Monte Carlo (8% for single and 16% for double tags), trigger effiency (3%), lepton veto efficiency (2%), PDF (2%), and jet energy scale. ISR/FSR systematic uncertainties (between 1% and 6%) are applied on the signal.

Higgs mass	**ZH Exp.**	**WH Exp.**	**Combined Exp.**	**Obs.**
110	25.5	34.8	14.9	17.8
115	28.4	34.0	15.4	16.0
120	31.7	35.1	16.8	15.6
125	40.5	39.2	20.0	21.4
130	50.0	41.5	22.6	22.8

Table 3: Expected limits in the ZH, WH, and combined channel wrt the Standard Model cross-sections

Since the number of events predicted in the Signal Region is consistent with the SM background without the Higgs, we have calculated the 95% C.L. exclusion limits. Table 3 shows the expected results in the ZH and WH production modes after combining the single and double-tag limits in both channels. The last two columns show the ZH+WH combined final result in the $\not\!\!E_T$ + b-jets channel. All the limits are calculated with respect to the Standard Model prediction, since the ratio of the observed and theoretical cross-section is the figure of merit with which the ZH, WH, and combined sensitivity can be compared to each other and to other channels. Figure 1 shows the dijet invariant mass distributions in the single and double tag samples, which were used as inputs to T. Junk's limit calculator [9].

We have performed a direct search for the Standard Model Higgs boson decaying into b-jet pairs in 0.973 fb^{-1} data accumulated in Run II of the CDF detector. We do not observe any significant excess over the background predicted by the Standard Model, thus we set a 95 % C.L. upper limit for the Higgs boson at varous masses. The observed cross-section limits agree with the expected ones. (See Figure 2.)

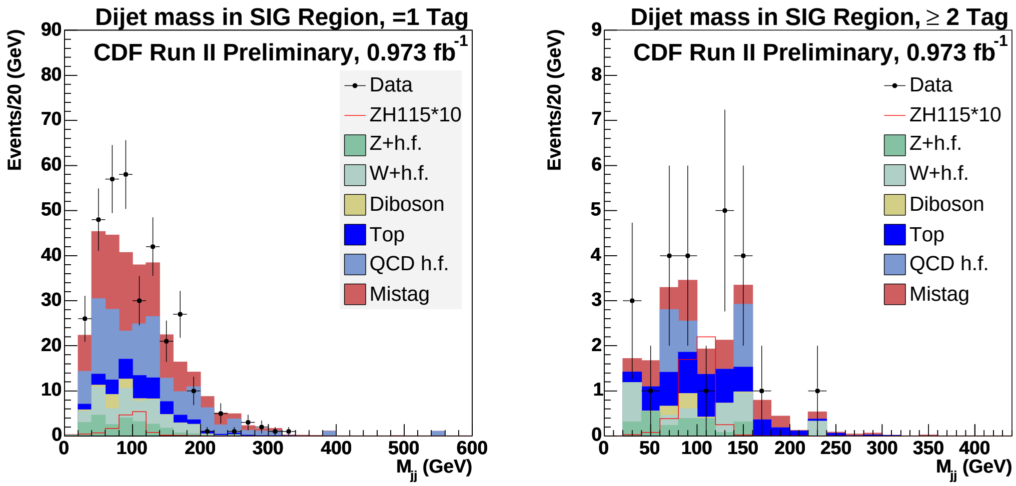

Figure 1: Invariant dijet mass in the single and double-tag signal region

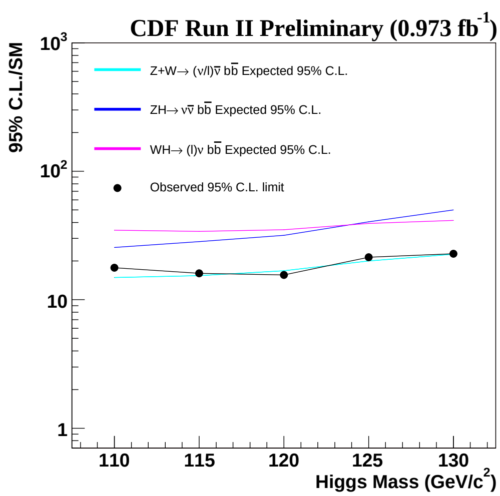

Figure 2: The 95% C.L. cross section upper limits

References

[1] J. Gunion et al., *The Higgs Hunter's Guide* (Addison-Wesley; New York, 1990). 2

[2] A. Heister et al., (The LEP Higgs Working Group), Phys. Lett. B **565**, 61 (2003). 2

[3] D. Acosta et al., FERMILAB-PUB-05-042-E, Accepted by PRL (2005). 2

[4] M. Carena et al., *Report of the Tevatron Higgs Working Group*, hep-ph/0010338. 2

[5] CDF and DØ Collaborations, *Results of the Tevatron Higgs Sensitivity Study*, FERMILAB-PUB-03/320-E. 2

[6] D. Acosta et al., Phys. Rev. D. **71**, 052003 (2005). 2, 3, 4

[7] D. Acosta et al., Phys. Rev. Lett. **93**, 142001 (2004). 3

[8] B. Harris, E. Laenen, L. Phaf, Z. Sullivan, S. Weinzierl, Phys. Rev. D **66**, 054024 (2002). 3

[9] T. Junk, *Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties*, CDF/DOC/STATISTICS/PUBLIC/8128,
see also http://www.hep.uiuc.edu/home/trj/cdfstats/mclimit_csm1/ 6