

A Search for the Standard Model Higgs Boson in the Process $ZH \rightarrow \mu^+ \mu^- b\bar{b}$ Using a Loosened Muon Selection

The CDF Collaboration
<http://www-cdf.fnal.gov>
July 15, 2010

Abstract

We present a new Z reconstruction category used in the $ZH \rightarrow \mu^+ \mu^- b\bar{b}$ Higgs search channel. We loosen several muon selection criteria to increase acceptance. The loosened selection allows Z candidates formed from muon objects to be recovered, since the selected muons failed the previous selection criteria. We search for new events using this new muon selection using existing muon trigger paths, as well as using an orthogonal data sample from a missing energy trigger. Due to an increase in non- Z background rates for this new category, we employ a neural network to increase the signal purity for this category of events. We also use a new multivariate regression technique to determine the trigger efficiencies needed for modeling the detector acceptance of signal and background. We present the analysis methods used for this new category, as well as the resulting gain in acceptance for Z candidates. This new analysis category is applied to data corresponding to an integrated luminosity of up to 5.6 fb^{-1} . For a Higgs boson mass of $115 \text{ GeV}/c^2$, we observe (expect) a limit on the ZH production cross section of 20.3 (17.6) times the value predicted by the Standard Model.

Preliminary Results

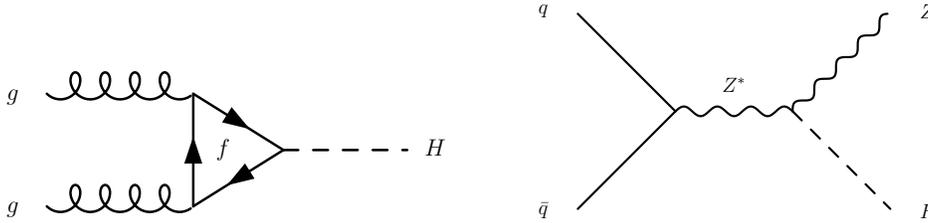


Figure 1: The Feynman diagrams for Higgs boson production through gluon fusion (left) and in association with a Z boson (right).

1 Introduction

The Standard Model (SM) is one of the most successful theories to date. The last remaining particle predicted by the SM but not yet observed is the Higgs boson, which results from the spontaneous breaking of the $SU(2) \times U(1)$ electroweak symmetry in the SM [1]. Through this spontaneous symmetry breaking, known as the Higgs mechanism, the W and Z electroweak gauge bosons acquire mass terms in the SM Lagrangian, while one massless degree of freedom remains (the photon).

Experiments at the LEP collider have excluded a Higgs boson with a mass below $114.4 \text{ GeV}/c^2$, at the 95 percent confidence level [2]. Searches for the SM Higgs boson at the Tevatron generally are performed over the mass range $100 < m_H < 150 \text{ GeV}/c^2$.

For searches in the low-mass range ($m_H < 135 \text{ GeV}/c^2$), the Higgs boson decays dominantly to a pair of b quarks. For the high-mass searches, the dominant decay of the Higgs is instead a pair of W bosons. Because the W bosons can decay leptonically, the direct production mechanism of the Higgs boson through gluon fusion, shown in Figure 1 is utilized best for the high-mass searches [3]. For the low-mass searches, the direct production mechanism is overwhelmed by QCD backgrounds, so the associated production mechanism shown in Figure 1 is best [3]. The additional lepton(s) from the decay of the vector boson is useful in reducing background contamination.

For this analysis, we search for events in the associated production process, with subsequent decays of $Z \rightarrow \mu^+ \mu^-$ and $H \rightarrow b\bar{b}$. This analysis is complementary to existing CDF analyses in the same channel which use both electrons and muons [4, 5]. In this note, we detail the event selection and additional analysis techniques applied. We also report our results in the form of limits on the SM Higgs production cross section times branching ratio for $p\bar{p} \rightarrow ZH \rightarrow \mu^+ \mu^- b\bar{b}$.

2 Event Selection

Since this analysis is complementary to the full $ZH \rightarrow \ell^+ \ell^- b\bar{b}$ analysis described in [4], we maintain many of the same analysis methods and techniques. We search for events using data collected with the CDF detector [6] corresponding to an integrated luminosity of 5.6 fb^{-1} . We utilize two main trigger paths to select events. The high P_T

muon trigger requires a muon candidate in the central region of the detector, $|\eta| < 1.1$, as well as a transverse momentum $P_T > 18$ GeV/ c . This allows events that failed the previous muon selection criteria to be recovered through the looser selection criteria of this analysis. In addition to the high P_T muon trigger, we also search for events using a missing energy (\cancel{E}_T) trigger, which requires at least 35 GeV of \cancel{E}_T in addition to one central jet ($|\eta| < 1.1$) and a second jet. Dimuon events can be found using this \cancel{E}_T trigger because of the minimum-ionizing deposits that give a large amount of \cancel{E}_T at the level where the trigger is applied (the \cancel{E}_T is later corrected for the muon P_T).

After requiring the trigger paths mentioned above, we search for events that contain at least two muons passing our loosened muon criteria. The reconstructed invariant mass of the highest P_T muon pair must fall within the Z selection window, $76 < m_{\mu\mu} < 106$ GeV/ c^2 . We also require that the two muons be oppositely charged. In addition to this selection, we have developed an artificial neural network to select the highest quality muons. We require both of the selected muons to pass this additional NN-based selection. More information on the muon identification NN can be found in a later section.

Apart from the muon selection, we require events to contain two or more jets, located in the region $|\eta| < 2.0$ of the detector. The lead jet has the energy requirement $E_T > 25$ GeV, while all other jets are required to have $E_T > 15$ GeV. We apply the same jet energy corrections used in previous ZH analyses [4] to improve the dijet mass resolution (this is one of the most important distributions for isolating Higgs-like events). Furthermore, we require a separation between the selected muons and the jets, requiring $\Delta R \equiv \sqrt{\Delta\phi^2 + \Delta\eta^2} > 0.4$.

Events passing both the muon selection (including the muon identification NN) and the jet selection criteria form the ‘pre-tag’ region of the analysis. This region is useful to validate our background model and final event discriminant, due its larger statistics.

From the pre-tag data, we apply a further selection to obtain the signal region used to produce the final results. We require one or more of the selected jets to be b -tagged, meaning the jet is likely to come from the production and resulting hadronization of a b quark. Two different b -tagging algorithms are used in this analysis. The secondary vertex (SecVtx) algorithm [7] assigns b -tags to jets by searching for tracks which have vertices displaced from the original interaction vertex. The SecVtx algorithm has two separate qualities - tight and loose, with the tight SecVtx b -tag being the better of the two. We also use the jet probability (JP) algorithm [8]. Instead of making a ‘yes’ or ‘no’ decision, the JP algorithm assigns a probability for the given jet to originate from the interaction vertex. Therefore, b -jets will have lower JP probabilities. In this analysis, we only choose JP-tagged jets if the probability is less than 0.05.

We use three separate categories of b -tagging to further isolate regions of high signal purity. First we search for two or more jets having a tight SecVtx tag (DT). If this is not the case, we search for one jet to have a loose SecVtx tag and another to have a JP tag (L+JP). The third and final tag category requires that only one of the jets have a tight SecVtx tag (ST). These three b -tagged categories of events make up the signal region for this analysis.

	Pre-tag Events
$Z + q\bar{q}$	1593.2 ± 336.0
$Z + c\bar{c}$	117.3 ± 47.6
$Z + b\bar{b}$	64.8 ± 26.3
$t\bar{t}$	12.2 ± 2.6
WW	0.5 ± 0.06
WZ	18.9 ± 2.5
ZZ	20.5 ± 2.7
Fakes	22 ± 11
ZH_{120} Signal	0.81 ± 0.07
Total Background	1849.4 ± 340.5
Data	1777

Table 1: Expected numbers of events for the individual background processes, as well as the total background expectation. The observed number of data events is well within uncertainties. The expected number of signal events is shown for $m_H = 120 \text{ GeV}/c^2$.

3 Background Models

The main background for this analysis consists of events consisting of a real Z boson with additional QCD multi-jet production (Z +jets). Feynman diagrams for examples of this background category are shown in Figure 3. To model these Z +jets backgrounds, we use Monte Carlo events generated with ALPGEN [9], while associated showering processes are described using PYTHIA. Smaller but significant sources of background events are the diboson processes ZZ , WZ , and WW , as well as $t\bar{t}$ production. We use PYTHIA [10] Monte Carlo samples to model these additional backgrounds.

Additionally, there are non- Z events (fakes) which contribute to the total background in this analysis. To determine the rate of these fake events, we apply the standard event selection, but instead select pairs of muons with identical charge ($\mu^+\mu^+$

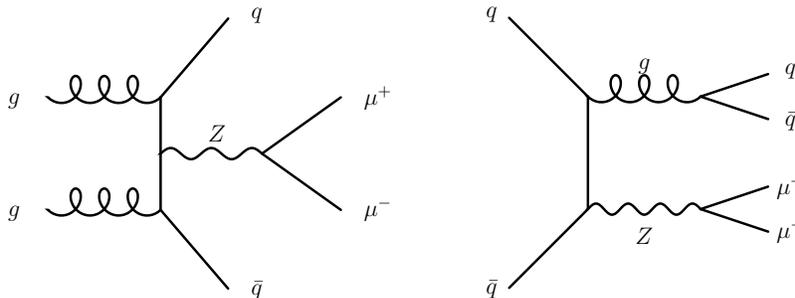


Figure 2: Two example Feynman diagrams for Z bosons produced with additional jets.

or $\mu^-\mu^-$). Because the fake rate is very small, $\sim 1\%$, we determine the kinematic shape of these fake events from a higher-statistics sample, specifically the sample of events which fail the muon identification NN requirements. This gives the various kinematic shapes of the fake contribution, while the rate comes from the total number of same-sign events passing all event selection.

We compare our data and background model for several different kinematic distributions, shown in Figures 3 and 4. The total numbers of observed and predicted events for the pre-tag region are shown in Table 1.

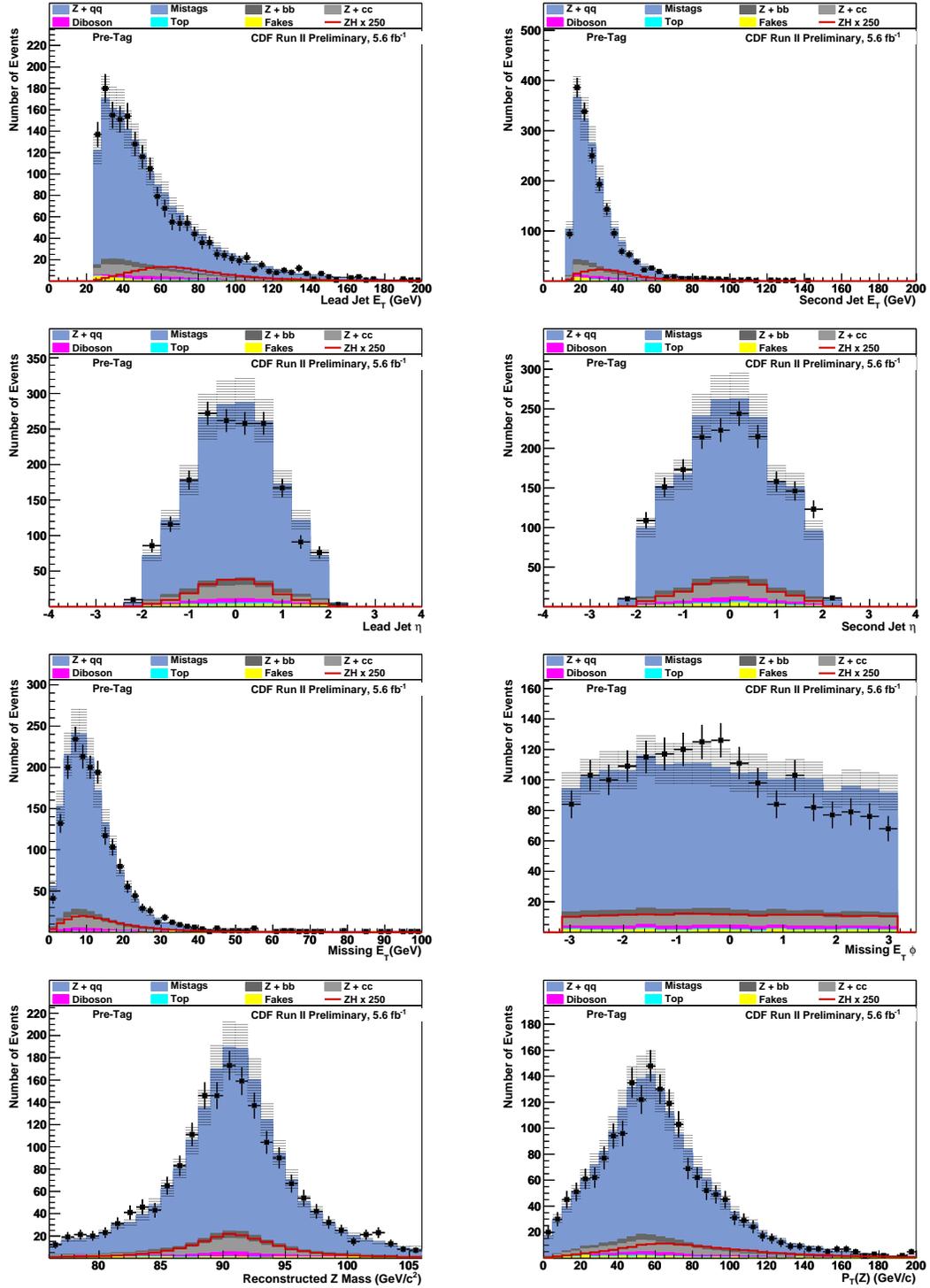


Figure 3: Several kinematic distributions to compare the background model used in this analysis with the observed data events.

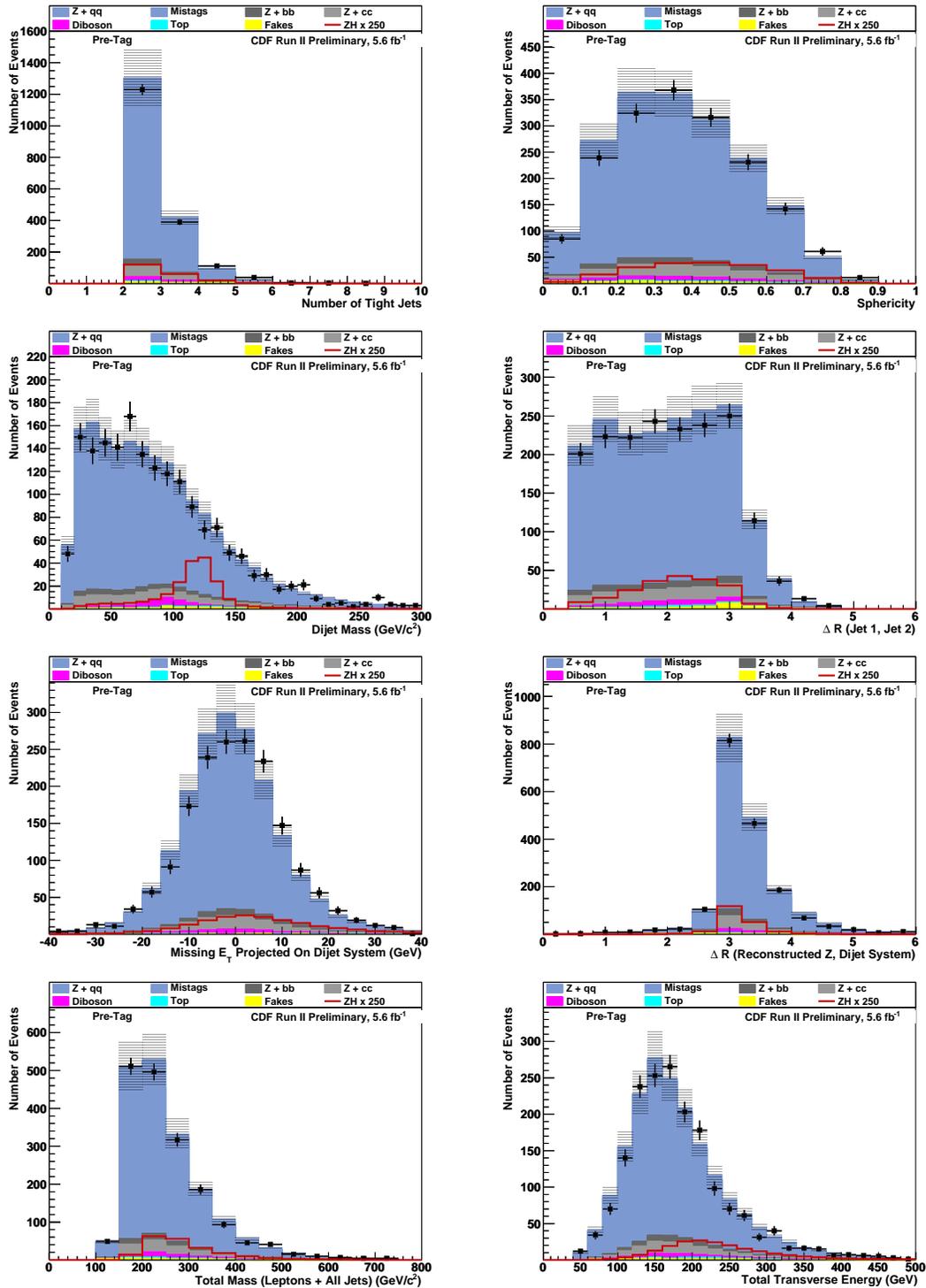


Figure 4: Several kinematic distributions to compare the background model used in this analysis with the observed data events.

4 Analysis Techniques

4.1 Muon Identification NN

Due to loosening the muon selection criteria, the non- Z background (fake) rate is considerably higher in this analysis. To deal with this increased contamination, we have developed an artificial neural network (NN) to select the highest quality muon candidates. This NN is trained on several muon kinematic quantities, and gives an output score ranging from 0 to 1 for the quality of a muon. High-quality muons are given scores close to 1. Events are included in this analysis if both of the selected muons pass this muon identification NN with a score greater than 0.1. Figure 4.1 shows the effect of applying this muon identification NN to the data sample. It can be seen that most of the underlying background distribution is rejected by the NN selection, and an isolated Z peak with little remaining background.

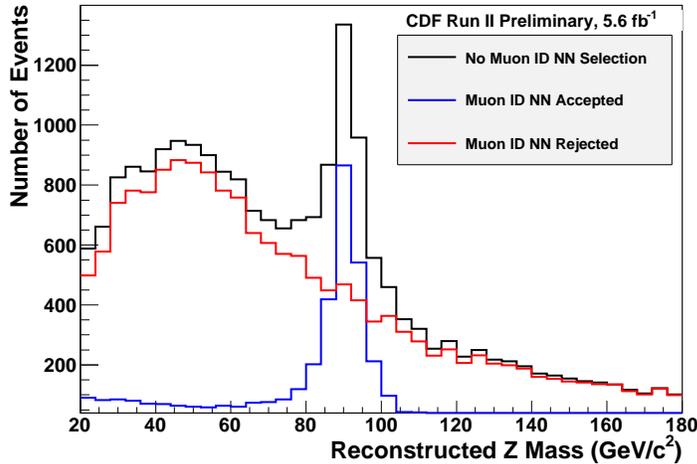


Figure 5: The effect of the muon identification NN on data events. The black line shows all data events, prior to any NN selection. The red curve shows events failing the muon ID NN selection, while the blue line shows events that pass the muon ID NN selection.

4.2 Final Event Discriminant

To better isolate the small expected signal from the larger background distributions, we employ an artificial neural network as a final event discriminant. The final NN is two-dimensional, and is trained to separate the two major classes of background events, top pair production and Z +jets production, from the ZH signal sample. We train the NN using a composition of $Z + b\bar{b}$, $Z + c\bar{c}$, and $Z + q\bar{q}$ Monte Carlo events

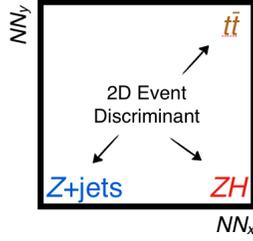


Figure 6: The three distinct output regions of the two-dimensional final event discriminant, which simultaneously separates Z +jets events and $t\bar{t}$ events from ZH (signal)-like events.

for the Z +jets training sample. The NN is trained on ZH signal Monte Carlo events with $m_H = 120 \text{ GeV}/c^2$.

The NN attempts to isolate each of these three classes of events to different corners of the two-dimensional output plane. Z +jets events are isolated to the $(NN_x, NN_y) = (0,0)$ corner of the plane, while the ZH signal events are isolated to the $(1,0)$ corner. $t\bar{t}$ events are separated both from Z +jets and ZH events simultaneously in the $(1,1)$ corner of the output plane. Figure 4.2 shows the three output regions graphically.

The two-dimensional NN accepts 11 different kinematic quantities as inputs, which are listed here:

- \cancel{E}_T — The missing transverse energy in the event, corrected for jets and muons.
- N_{jet} — The number of tight jets in the event.
- $\vec{\cancel{E}}_T \cdot (\vec{j}_1 + \vec{j}_2)$ — The projection of the \cancel{E}_T vector on to the vector sum of the two leading jets in the event.
- $P_T(Z)$ — The transverse momentum of the reconstructed Z boson.
- $\sum E_T$ — The scalar sum of the E_T of the Z and all tight jets in the event.
- m_{jj} — The dijet mass (or reconstructed Higgs boson mass).
- $\Delta R_{j_1, j_2}$ — The separation between the lead and secondary jet.
- $\Delta R_{Z, H}$ — The separation between the reconstructed Z and H bosons.
- Sphericity — An angular measure which takes into account the distribution of all reconstructed objects and their positions in the detector.
- m_{tot} — The total mass of all reconstructed objects (2 leptons and all jets).
- m_{Zjj} — The total reconstructed mass of the Z boson, lead, and secondary jets in the event.

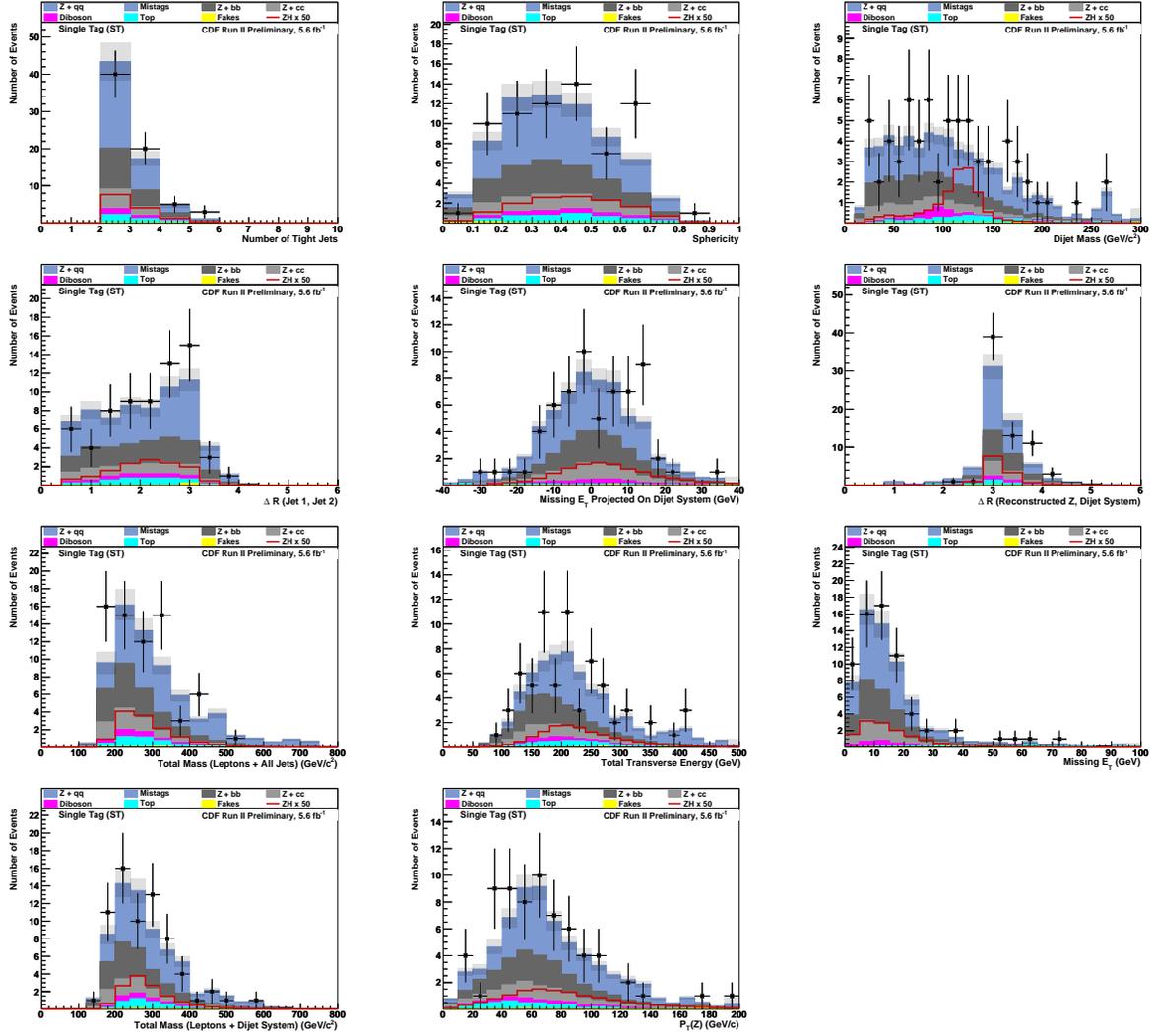


Figure 7: Final event discriminant inputs for the single tight SecVtx b -tagging category.

These kinematic distributions were shown at the pre-tag level in a previous section. We show the distributions for the b -tagging categories in Figures 7 and 8. Due to low statistics in the DT and L+JP categories, we have combined them into a single plot.

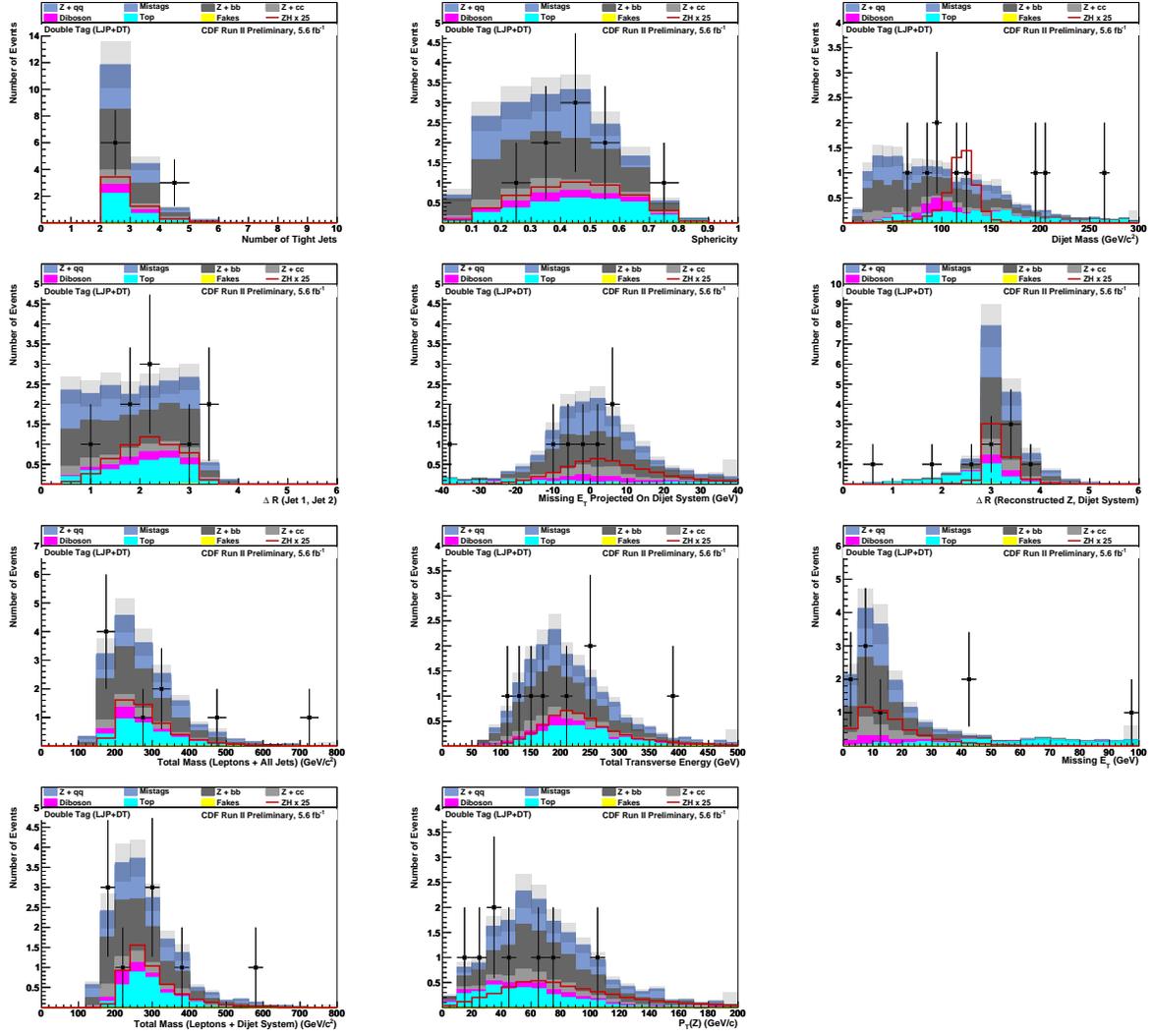


Figure 8: Final event discriminant inputs for the two double b -tag categories, L+JP and DT.

5 Systematic Uncertainties

We apply several systematic errors to our analysis to account for experimental and theoretical uncertainties. There are several rate uncertainties, such as theoretical uncertainties from process cross sections. The largest of these is the theoretical uncertainty on the $Z + b\bar{b}/c\bar{c}$ cross section, at 40%. Other errors on process cross sections include $t\bar{t}$ (20%), WW , WZ , and ZZ (11.5%), and ZH (5%).

There are also detector-related uncertainties due to reconstruction and measurement. We apply a systematic of 1% to account for our lepton identification efficiency, and an additional systematic of 1.5% to account for the electromagnetic calorimeter energy uncertainty. The error on our fake rate is set to 50%. A final uncertainty on the integrated luminosity is roughly 5%.

This analysis also includes a new method for estimating the trigger efficiencies for Monte Carlo events. It is based on a NN technique, and assigns a probability for each Monte Carlo event to pass the trigger requirements. We apply a systematic of 5% to account for the uncertainty in the choice of training conditions for this NN.

Due to uncertainties in b -tag efficiencies and calculation of scale factors, there are rate systematics for each of the three b -tag categories that make up the signal region. These range from 4% for the single SecVtx category to 11% for the loose SecVtx plus jet probability category.

Apart from the various rate uncertainties described above, there are additional systematics which affect the shape of the final event discriminants. These shape uncertainties include effects due to the jet energy scale, as well as production of initial or final state radiation. The mistag event weights also contribute an additional uncertainty.

Table 2 summarizes the various shape and rate systematics used in this analysis.

6 Results

After dividing our signal region into the three b -tagging categories, we use the final event discriminant to search for an excess of events representing a Higgs signal. We do not observe such an excess and instead set limits on the SM Higgs production cross section using the MCLimit tool [11]. This tool sets an upper limit at the 95% confidence level for the amount of ZH signal that could be consistent with the observed data and expected background events. Table 3 shows the expected number of background events and the observed number of data events for each of the three b -tagging categories in this analysis.

We validate the final event discriminant by comparing the Monte Carlo background estimation with the observed data events. We observe agreement in both the shape and expected number of events in the final event discriminant. Figure 9 shows the output of the final event discriminant for the pre-tag control region, as well as the single tag (ST) and both double tag categories (LJP and DT combined).

The expected and observed limits on the SM ZH production cross section are shown in Table 4. Figure 10 shows the same information in a graphical form.

Systematic Item	Uncertainty	Affected Samples
Luminosity		All MC
Tevatron	0.05	
CDF	0.04	
Process Cross Sections		
$Z + b\bar{b}, c\bar{c} \sigma$	0.40	$Z + b\bar{b}, Z + c\bar{c}$ MC
$t\bar{t} \sigma$	0.20	$t\bar{t}$ MC
$VV \sigma$	0.115	WW, WZ, ZZ MC
$ZH \sigma$	0.05	ZH MC
Reconstruction and ID		
Lepton Identification	0.01	All MC
NN-Derived Trigger Efficiency	0.05	All MC
EM Energy Scale	0.015	All MC
b -tag Scale Factors		
T+T	0.08	All T+T MC
L+JP	0.11	All L+JP MC
T Only	0.04	All T MC
Jet Energy Scale	SHAPE	All MC
Jet Mistags	SHAPE	Mistag Events
Initial/Final State Radiation	RATE	ZH MC
Fakes	0.50	$\mu^\pm \mu^\pm$ Events

Table 2: Systematic uncertainties applied to this analysis.

	ST Category	LJP Category	DT Category
$Z + q\bar{q}$ (Mistags)	33.8 ± 4.8	2.18 ± 0.8	0.22 ± 0.06
$Z + c\bar{c}$	8.5 ± 3.4	1.8 ± 0.7	0.20 ± 0.08
$Z + b\bar{b}$	17.1 ± 6.9	5.1 ± 2.1	2.29 ± 0.93
$t\bar{t}$	3.9 ± 0.8	2.3 ± 0.5	1.33 ± 0.28
WW	0.03 ± 0.004	0.01 ± 0.001	–
WZ	0.66 ± 0.09	0.08 ± 0.01	–
ZZ	1.82 ± 0.24	0.75 ± 0.10	0.32 ± 0.04
Fakes	0.24 ± 0.12	0.01 ± 0.005	–
ZH_{120}	0.25 ± 0.02	0.14 ± 0.01	0.08 ± 0.007
Total Background	66.1 ± 9.1	12.2 ± 2.4	4.36 ± 0.98
Data	68	5	4

Table 3: Comparison of background and signal expected numbers of events with observed number of data events for the three tagging categories.

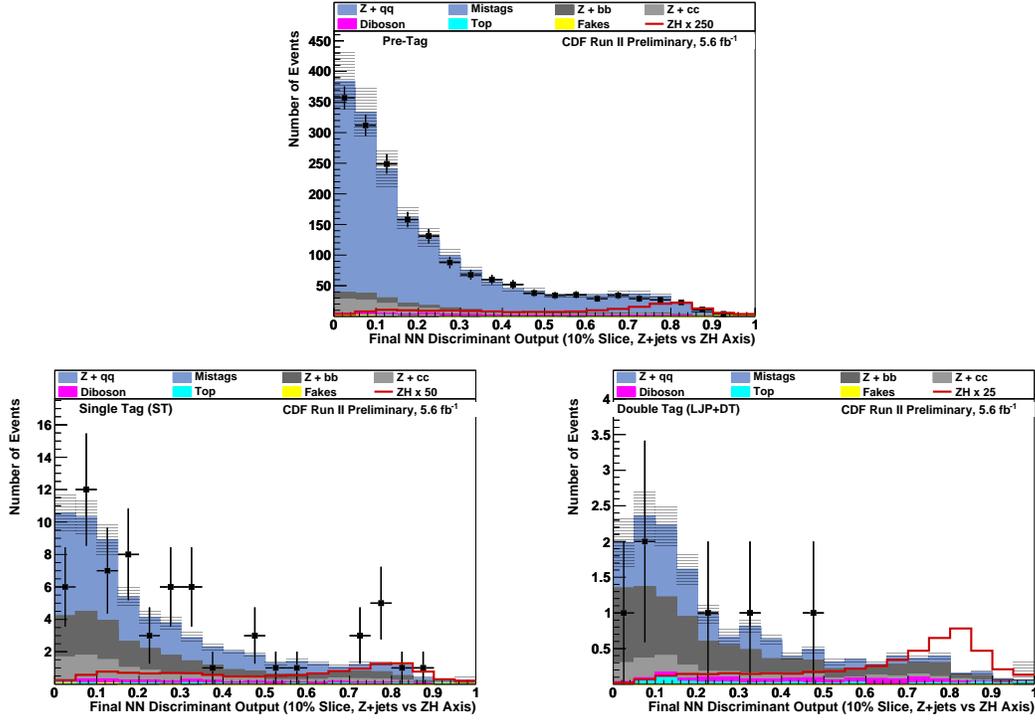


Figure 9: Final event discriminant outputs for the pre-tag region, as well as the single tag (ST) category and two double-tagged categories of events (LJP and DT, combined in a single plot).

m_H (GeV/ c^2)	Observed Limit	Expected Limits				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
100	14.76	7.06	9.64	13.44	19.08	28.81
105	16.32	7.85	10.48	14.48	20.98	30.31
110	18.30	8.79	11.29	15.94	23.52	36.64
115	20.28	9.08	12.45	17.60	25.37	36.17
120	27.65	13.58	18.05	25.22	36.39	51.48
125	29.97	14.46	19.05	27.03	40.50	57.91
130	37.49	17.31	23.32	33.93	51.17	76.42
135	48.00	25.42	33.28	49.03	73.03	118.05
140	66.92	32.19	46.89	68.73	110.53	164.27
145	97.40	49.22	69.73	109.51	170.37	266.12
150	142.27	65.78	99.82	161.78	268.73	412.37

Table 4: Observed and expected limits, for all tagging categories combined. The observed limits include the $\pm 1\sigma$ and $\pm 2\sigma$ error values as well.

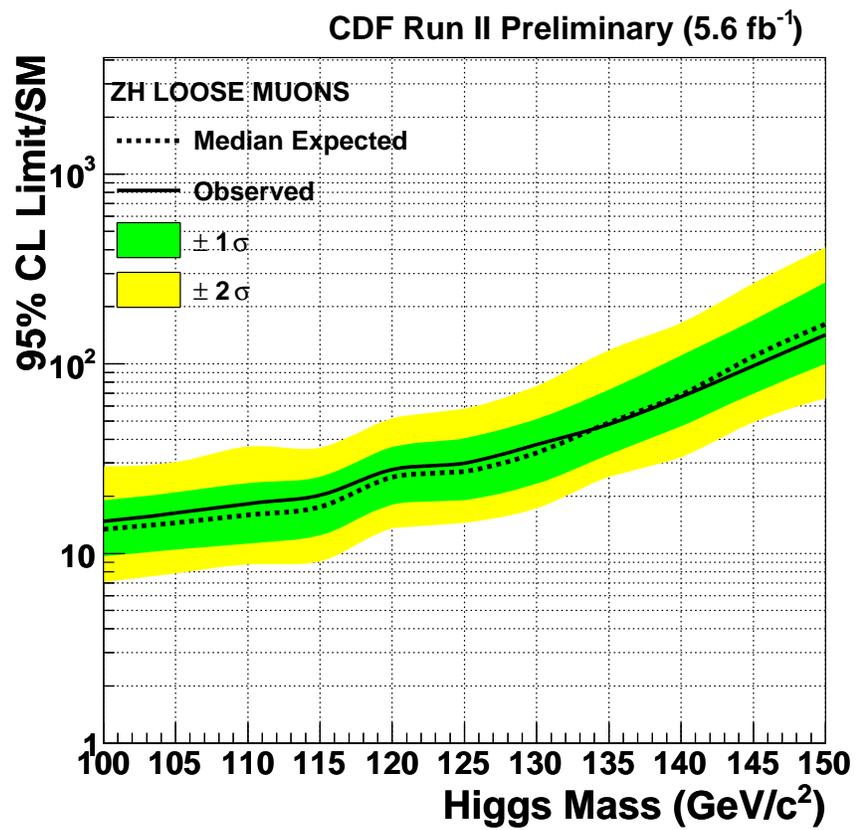


Figure 10: Observed and expected 95% confidence level limits on the ZH production cross section.

7 Conclusions

By loosening some of the muon selection criteria, we have increased acceptance for the $ZH \rightarrow \mu^+\mu^-b\bar{b}$ Higgs search channel. Developing an artificial neural network to help select these lower quality muons was key to obtaining a signal purity similar to previous ZH analyses. In addition to gaining additional events from the high P_T muon trigger, we have included the \cancel{E}_T +jets trigger for the first time in this analysis channel. We use a two-dimensional final event discriminant to isolate a high-purity signal region. From this discriminant, we are able to calculate 95% confidence level limits on the SM Higgs associated production cross section. The observed limits range from 14.8 to 142 times the SM predicted value for Higgs bosons in the mass range $100 < m_H < 150 \text{ GeV}/c^2$. This analysis will be combined with the full $ZH \rightarrow \ell^+\ell^-b\bar{b}$ analysis to increase sensitivity. This analysis will also be included in the latest Tevatron Higgs search results.

References

- [1] Peter W. Higgs, Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.* **13**, 508 (1964).
- [2] The LEP Working Group for Higgs Boson Searches, Search for the Standard Model Higgs boson at LEP. *Phys. Lett. B* **565**, 61-75 (2003).
- [3] Julien Baglio and Abdelhak Djouadi, Predictions for Higgs Production at the Tevatron and the associated uncertainties. [arXiv:1003.4266](https://arxiv.org/abs/1003.4266).
- [4] The CDF Collaboration, A Search for the Standard Model Higgs Boson in the Process $ZH \rightarrow \ell^+\ell^-b\bar{b}$ Using 4.1 fb^{-1} of CDF II Data. CDF Public Note 9889.
- [5] Richard Hughes, Ben Kilminster, Brandon Parks, Brian Winer, Rob Harr, Shalhout Shalhout, Bo Jayatilaka, Ashutosh Kotwal, Ravi Shekhar, Daniel Whiteson, A Search for $ZH \rightarrow l^+l^-b\bar{b}$ in 2.7 fb^{-1} using a Neural Network Discriminant. CDF Public Note 9665.
- [6] The CDF Collaboration, Measurements of inclusive W and Z cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$. *J. Phys. G* **34**, 2457 (2007).
- [7] The CDF Collaboration, Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$ using Lepton + Jets Events with Secondary Vertex b -tagging. CDF Public Note 7138.
- [8] Enrique Palencia, Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96 \text{ TeV}$ Using Lepton+Jets Events in the CDF Detector at Fermilab. Ph.D. Thesis, CDF Public Note 8772.

- [9] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa, ALPGEN, a generator for hard multiparton processes in hadronic collisions. [arXiv:hep-ph/0206293](#).
- [10] Torbjorn Sjostrand, Stephen Mrenna, Peter Skands, PYTHIA 6.4 Physics and Manual, [arXiv:hep-ph/0603175](#).
- [11] Thomas Junk, Confidence Level Computation for Combining Searches with Small Statistics. [arXiv:hep-ex/9902006](#).