CDF/ANAL/EXOTIC/PUBLIC/8422
Version 1.0
March 20, 2007

Search for $ZH \to \ell^+\ell^- b\bar{b}$ in 1 fb^{-1} of CDF Run 2 Data

Jonathan Efron, Ben Kilminster, Richard Hughes, Brandon Parks, Brian Winer
The Ohio State University (Columbus, OH)
Beate Heinemann University of California (Berkley, CA)
Andrew Mehta
University of Liverpool (UK)

Abstract

We search for the Higgs Boson in 1 fb^{-1} from the process $ZH \to l^+l^-b\bar{b}$ in both ee and $\mu\mu$ channels. These channels have a small background of mostly real $Z+jets$ events due to the requirement of two leptons and a Z mass constraint. We loosen the tight and loose electron and muon cuts from standard high P_T (i.e. top physics) analyses to improve statistics. We correct the two candidate Higgs jets with an Artificial Neural Network which assigns $\not\!\!E_{\rm T}$ to the jets according to their $\not\!\!E_{\rm T}$ projections and relative ϕ. To maintain signal efficiency and improve signal discrimination, we employ an additional neural network trained to discriminate event kinematics of ZH compared to the main $Z + jets$ background and the kinematically different $t\bar{t}$ background. At a Higgs Mass of 120 GeV/c^2, we observe a 95% CL upper limit. This represents an improvement over our previous analysis on the same dataset by a factor of 2 greater luminosity.

In the Standard Model, the Higgs Mechanism was introduced to explain electroweak symmetry breaking. As a result of this mechanism, the Higgs Boson was theorized, but has still yet to be discovered. Current measurements of the Standard Model have calculated the Mass of the Higgs to be 76^{+39}_{-24} GeV/c^2 and therefore sets a upper limit of 144 GeV/c^2 with a 95% confidance level. [1] The experiments at LEP2 have excluded the a Standard Model Higgs for $m_H \leq 114.4$ GeV/c^2. [2] For this analysis we study a 1.019 fb^{-1} sample for electrons and a 0.973 fb^{-1} sample for muons. Our data sample was accumulated using the CDF detector at the Tevatron, which has a $\sqrt{s} = 1.96$ GeV/c^2.

This note updates the search for $ZH \rightarrow l^+l^-b\bar{b}$ described in CDF 8422 [**?**]. The new analysis uses the same 1 fb$^-$1 dataset as before, but makes improvements in the jet energies using a dijet energy correction function, and also by splitting the events into single and double-tagged events. A new kinematic Neural Network, using a different set of variables, is optimized to take advantage of the jet energy improvements. In addition, we have made improvements in our mistag estimations as well as their systematic shape uncertainty which is now included in the fitting procedure.

In this analysis, the Higgs Boson is produced in association with a Z^0 boson. Specifically, the channel where $Z \rightarrow \ell^+\ell^-$ and $H \rightarrow b\bar{b}$. Consequently, the signature of this analysis are two charged leptons that can be resemble a Z boson and at least two hadronic jets, one of which can be labeled as coming from a b jet.

To recreate a Z boson we search for one "Tight" lepton and one "Loose" lepton. These designations are base on the requirements for the lepton, which include transverse energy, transverse momentum and other detector quantities. For a electron events, we require a "Tight" lepton with $E_T \geq 18$ GeV and a "loose" lepton with $E_T \geq 10$ GeV. For muon events, we require a "Tight" muon with $P_T \geq 20$ GeV and a "loose" muon with $P_T \geq 18$ GeV. We require the combined invariant mass to be 76 GeV/$c^2 \geq M_{\ell\ell} \geq 116$ GeV/c^2. For all muon events and for electron events where both lepton have an $\eta \leq 1.0$, that the two leptons have opposite charges.

To recreate Higgs candidates we require two or more jets. A jet is defined as cluster of energy in the calorimeter with a cone size of $\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ 0.4 We also require that these jets have an $|\eta| \leq 2$ and an $E_T \geq 15$ GeV for all jets and ≥ 25 GeV for the highest energetic

Selection
One high E_T lepton
A second high E_T lepton of same type
ΔZ_0 of leptons < 4 cm
Opposite Sign muons or central-central electrons
Z mass window 76 GeV - 106 GeV
2 or more Cone 0.4 jets L5 $E_T > 15$ GeV, $\|\eta\| < 2$
1 or more jets with L5 $E_T > 25$ GeV
2 or more loose SECVTX tags
OR 1 tight SECVTX tags

Table 1: Summary of event selection. See lepton selection tables.

jet.

To find jets that originated from a b hadron, the SecVtx algorithm is used. [[4]] For jets in an event, the CDF silicon detector can measure tracks that originate near the beam line. If the vertex of these tracks is displaced from the primary vertex of the event, the jet is said to be "tagged." If the displacement is in the direction of the jet, it is said to be a positive tag. Otherwise, it is a negative tag. A b hadron will be displaced from the primary vertex because of its longer lifetime. Other objects can be tagged due to detector resolution effects. The number of negative tags is on order of the number of faked positive tags. In this analysis, we first search for two "loose" tags in an event. If they are not found, we then search for events with one "tight" tag.

The background consists of three components:

1. Background with two genuine leptons and a genuine b- or c-jet. This category can be further broken down into three categories. A produced with heavy flavor jets, $t\bar{t}$, and diboson (ZZ and ZW) events. During the evaluation of our analysis, we tested several generators to model these shapes. For the Z+jets, we found that Alpgen + Herwig modeled the shapes we examined better. For the other backgrounds, we used a Pythia generator. However, we used both generators to evaluate systematics that arose from shape differences between the two.

2. Backgrounds with a fake electrons are estimated using lepton fake rates [6]. A 50%

uncertainty is assigned to these lepton fake rates. For muons events where both muons have the same charge are taken to estimated this contribution. This procedure was also used for the $Z + b$ cross section measurement [7].

3. Backgrounds with a mistagged light jet are estimated using the mistag matrix [8].

The number of expected and observed events is given at various selection stages in Tables 2 for muons and 3 for electrons.

Since the only sources for $\not\!\!E_{\rm T}$ in $ZH \to l^+l^-b\bar{b}$ events are jet energy mismeasurement, we employ a correction function to reassign missing energy to the jets. This correction function is derived from ZH Monte Carlo events, and is only applicable to this final state. It is essentially a parton-jet transfer function which makes use of the correlations between the jet energies and directions and that of the $\not\!\!E_{\rm T}$. These improved jet energies are used to calculate the input kinematics for our signal to background discriminant, including most importantly, the dijet mass which is a direct measure of the Higgs mass.

The function takes as inputs the $\not\!\!E_{\rm T}$, $\not\!\!E_{\rm T}\phi$, first and second jet transverse energies, their ηs and ϕs, and the transverse projections of the jets onto the $\not\!\!E_{\rm T}$ direction. The function produces two outputs which are scale factors to correct the first and second jet energies. The directions of the jets are not changed.

A Neural Network produces this function by training on the inputs mentioned above in ZH MC and attempting to reproduce the true parton energies available from the MC hepg banks. Training is done with a wide range of Higgs masses from 60 GeV/c^2 to 180 GeV/c^2 in steps of 10 Gev/c^2. The wide range of masses is used in order to prevent the Neural Network from overtraining on a particular mass which would cause the function to produce corrections biased towards that mass. To ensure that there is no bias, we do a linearity test of the NN function on independent MC events.

We also test this function on background events to make sure that background distributions are not sculpted such that an increased amount of background were contained within a region around the Higgs dijet mass peak. To do this, we construct 2 σ windows around the ZH dijet mass peak before and after the jet correction functions are applied, and we verify that the signal to background discrimination is improved in the latter catgegory. The

backgrounds considered for this study are Z+jets backgrounds which have the largest contribution to the background estimation, and dilepton $t\bar{t}$ which has real $\not\!\!E_{\rm T}$ due to the W decays. $S/\sqrt{B}$ is found to improve in the case of both backgrounds, and the shape of the backgrounds is not distorted.

The NN dijet correction function is applied in the same way to data and MC. We validate the function by comparing the event kinematics, such as jet E_Ts, $\not\!\!E_{\rm T}$, H_T, of our background models of pretagged and tagged data before and after the function is applied.

To further improve signal to background discrimination after event selection, we employ an artificial Neural Network (NN) trained on a variety of kinematic variables to distinguish ZH from backgrounds. Specifically, A 2-D NN, trained to discriminate $t\bar{t}$ vs ZH on one axis, and Z+jets vs ZH on the other axis. We use the OSU RootJetnet interface to the Jetnet neural network program.

We first define a set of variables we expect to have differences between ZH and the other backgrounds, the largest two being Z+jets and $t\bar{t}$. To optimize the NN, we use an iterative procedure to determine the configuration which best discriminates signal and background, and which uses a minimal number of input discriminants.

This is done by first determining the best one-variable NN of all the above variables, then keeping this variable as an input, we loop over all other the variables to determine the best two-variable NN. The best N-variable network is finally selected once the N+1-variable network shows less than a percent improvement. The criteria for comparing networks is the testing error defined by how often an NN with a given configuration correctly classifies several thousand signal and background events. Once the best N-variable NN is chosen, we then optimize the number of hidden layers, and choose a number of training epochs which prevents over-training.

Our final Neural Network configuration is 8 input variables, 17 hidden nodes, and 2 output nodes.

- Corrected $\not\!\!E_{\rm T}$
- H_{T_2} : sum of $\not\!\!E_{\rm T}$, first two jet E_Ts, and lepton E_Ts
- dijet mass: Dijet invariant mass
- dRj1Z: delta R between Z and jet 1

- dRj1Z: delta R between Z and jet 2
- dRj1Z: delta R between jet 1 and jet 2
- sphericity
- Jet 2 η

Sample	2 Leptons	Z Selected	$\geq$1 Jet	$\geq$ 2 Jets	2 loose	1 tight (!=2loose)
ZH_{120GeV/c^2}	0.62	0.58	0.56	0.49	0.10	0.19
$t\bar{t}$	29.3	7.0	6.9	6.00	1.32 ± 0.27	2.33 ± 0.48
ZW	29.8	26.5	17.7	12.4	0.017 ± 0.003	0.51 ± 0.10
ZZ	43.8	38.6	22.9	17.1	0.58 ± 0.11	1.82 ± 0.36
$WW2p$	24.4	5.89	3.30	1.49	0.008 ± 0.002	0.020 ± 0.004
$Z \rightarrow \tau\tau$	320.1	11.3	1.04	0.42	0.00	0.00
fakes	1693	343	53	20	0.1 ± 0.1	1 ± 1
$Z^0 \rightarrow \mu^+\mu^-$	52500	45790	4218	1028	3.25 ± 1.01	35.1 ± 7.0
(b events)	648	576	144	49.8	$2.46 \pm .99$	14.5 ± 5.8
(c events)				106	0.43 ± 0.17	8.2 ± 3.3
(mistags)	Predicted mistags from data:				$0.36 \pm .09$	12.4 ± 2.1
Total	54,600	46,200	4320	1090	5.06 ± 1.05	40.8 ± 7.1
Data (972 pb^{-1})	56,740	47,982	4128	1240	5	46
Negative Tags	Predicted:				0.27	11.02 ± 0.72
	Found				1	12

Table 2: Amount of expected and observed events in $\int Ldt = 0.97 fb^{-1}$ of muon data at various cuts in the event selection. Until better cross sections of the heavy flavor Alpgen + Herwig is attained, the Top group's pythia is used to find b and c content normalizations. The Higgs signal assumes a $\sigma(ZH) = .093$ and $BR(H \rightarrow b\bar{b}) = 0.68$. Errors shown are from systematics.

The largest background after the full selection is Z+ jets production followed by $t\bar{t}$ production. Combining both electrons and muons we anticipate 0.64 ZH events assuming the Standard Model cross section value.

A critical check of the validity of the Neural Net is the comparison of the Neural Net distribution itself for the pretag sample. The 1-dimensional projections onto the x and y-axes are shown in Fig. 1 for electrons and muons combined. The data are generally well modeled by the simulation, however not ideally. The difference between data and the simulation is accounted for in the systematic uncertainty that we include by using a simulation using a different generator.. The x-axis is used to discriminate between Z+jets and ZH, the y-axis is used to discriminate between $t\bar{t}$ and ZH. Good agreement is observed for both muons

Sample	2 Leptons	Z Selected	$\geq$1 Jet	$\geq$ 2 Jets	2 loose	1 tight(!=2loose)
ZH_{120GeV/c^2}	0.83	0.76	0.73	0.63	0.13	0.25
$t\bar{t}$	36.9	9.01	8.77	7.69	$1.52 \pm .29$	2.91 ± 0.56
ZW	46.3	40.8	25.89	18.0	0.024 ± 0.005	0.72 ± 0.14
ZZ	61.4	54.0	29.7	21.5	0.76 ± 0.15	2.16 ± 0.43
$WW2p$	33.5	7.93	4.36	2.05	0.006 ± 0.001	0.040 ± 0.008
$Z \rightarrow \tau\tau$	536	26.9	2.75	0.84	0.00	0.00
fakes	1191	322.8	42.0	15.7	0.134 ± 0.067	0.89 ± 0.44
$Z^0 \rightarrow e^+e^-$	96,440	84,940	6943	1607	5.02 ± 1.52	54.1 ± 10.7
(b events)	1019	908	209	69.5	3.81 ± 1.48	20.6 ± 8.76
(c events)				161	0.58 ± 0.22	13.6 ± 5.21
(mistags)	Predicted mistags from data:				0.63 ± 0.15	19.9 ± 3.38
Total	98,340	85,400	7082	1673	7.36 ± 1.56	60.8 ± 10.7
Data (1019 pb^-1)	102,820	88242	6423	1794	6	54
Negative Tags	Predicted:				0.44 ± 0.04	17.2 ± 1.38
	Found				0	11

Table 3: Amount of expected and found events in $\int Ldt = 1.02 fb^{-1}$ electron data at various cuts in the event selection. The ZH event expectations are based on $\sigma(ZH) = .093$ and $BR(H \rightarrow b\bar{b}) = 0.68$. Errors are from systematics.

and electrons for both projections.

We consider the following sources of systematic uncertainties on the normalization of the background sources:

- an uncertainty of 40% on the normalization of the $Z+b\bar{b}$ and $Z+c\bar{c}$ backgrounds due to the uncertainty on the NLO cross section. This is consistent also with the uncertainty on the W+heavy flavor backgrounds in the WH analysis.
- an uncertainty of 20% on the $t\bar{t}$, WZ and ZZ normalization. This uncertainty includes the cross section uncertainty.
- an 8% uncertainty on the normalization of the mistag background due to the mistag matrix and an average uncertainty of ± 0.17 on the average mistag asymmetry factor of 1.37. These numbers are actually funtions of raw jet E_T.
- an 8% uncertainty on the SecVtx tagging efficiency for b-jets and a 16% uncertainty on the SecVtx tagging efficiency for c-jets. This affects the ZH signal, and the $Z + b\bar{b}$, $Z + c\bar{c}$, $t\bar{t}$, WZ and ZZ backgrounds.
- a 1% uncertainty on the trigger efficiencies and lepton identification efficiencies. This is

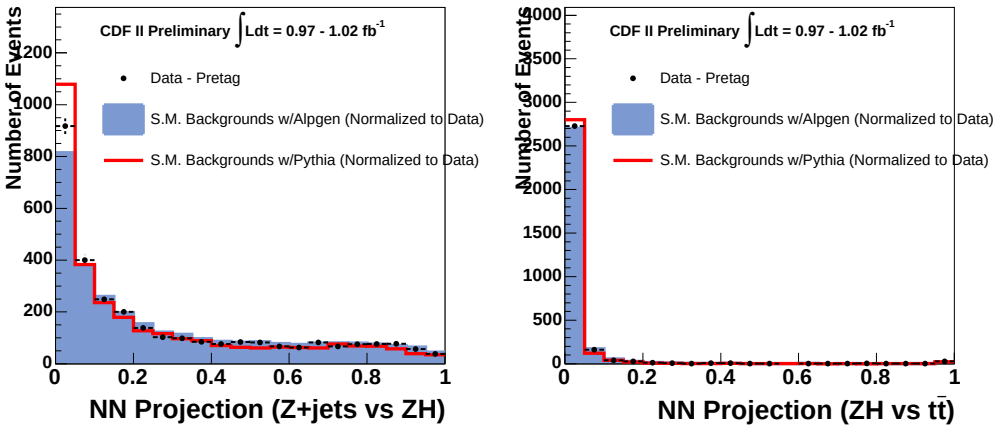

Figure 1: Pretagged data compared to Monte Carlo expectation: shown are the projections of the NN onto the x-axis (upper plot) and y-axis (lower plot). The data (points) are compared to the total background (blue). A second simulation of Z+jets is shown in red. The difference is taken into account in the systematic error.

correlated between the ZH signal, and the $Z+b\bar{b}$, $Z+c\bar{c}$, $t\bar{t}$, WZ and ZZ backgrounds.

- a 50% uncertainty on the fake Z background.
- a 6% uncertainty on the integrated luminosity. This affects all the MC samples.

In addition we consider the following uncertainties that change the shape of the NN distribution:

- the jet energy scale is applied to both signal and background samples to study the influence. It is treated to be correlated among all MC samples of signal and background.
- the difference between different generators for the $Z+$ heavy flavor backgrounds is taken as a systematic uncertainty. In the most signal like regions this is about a 30% effect.
- for signal we consider systematic uncertainties due to initial and final state QCD radiation and due to the parton distribution functions. We use the prescription from the Joint Physics web pages for both these effects.
- for $t\bar{t}$ production we estimate a systematic uncertainty using Herwig, and comparing that to Pythia. The contribution in the signal area of the NN distribution is very small though. So, it has a negligible effect.

We evaluate systematic uncertainties by doing pseudo-experiments where we assume our default likelihood function, and evaluate how our expected 95% CL upper limit changes when we input a different model of signal or background. For shape systematics, we choose events for our pseudo-experiments from systematically altered templates, and also take into consideration changes in expected event yields due to the effect of the systematic.

Our major systematic is the difference of modeling the Z+jets background using different MC generators. While Alpgen is in good agreement with the data for jet multiplicity, it tends to overestimate the Energy of jets and several other energy related distributions. Pythia, on the other hand, underestimates the energy of these distributions. Since Alpgen does a better job of modeling the NN output distribution, but the data falls between Pythia and Alpgen for the input kinematics, we assign a shape systematic due to the difference in modeling of the Z+jets background.

Other shape systematics include the jet energy scale which we determine by fluctuating the jet energy scale in all processes by 1 sigma, and the ISR and FSR systematics which we determine by comparing nominal signal MC using the "more or less" prescription of Pythia generation parameters outlined by Un-Ki Yang. Finally, we evaluate a further shape systematic due to the mismodeling of the $\not\!\!E_{\rm T}$ distribution. Electroweak Z Pythia with minimum bias events overlayed does a better job of predicting the $\not\!\!E_{\rm T}$ distribution, but we do not have such MC for the Z+ heavy flavor process, and so we model Z+jets with top MC not having extra interactions. We evaluate the systematic of this assumption using pseudo-experiments modeled with the electroweak Z+l.f. Pythia.

The table of systematics are listed in Table 4.

After verifying that the data are well modeled in both the pretag and the tagged sample the Neural Net is applied to the data. The resulting NN output distribution for the data is shown in Fig. 2. We observe the expected three contributions, the peak at $(1,1)$ is mostly due to $t\bar{t}$, the events at $(0,0)$ mostly due to Z+jets and the signal is expected near $(1,0)$.

The 1-dimensional projections along the x- and y-axes are shown in the top plot in Fig. 3 for single tag, and in Fig. 4 for double tagged events. for electrons and muons combined. The data agree well with the background predictions. In particular at the low and high values of the NN output the data confirm the $t\bar{t}$ and $Z+$ jets background estimates. The bottom plots

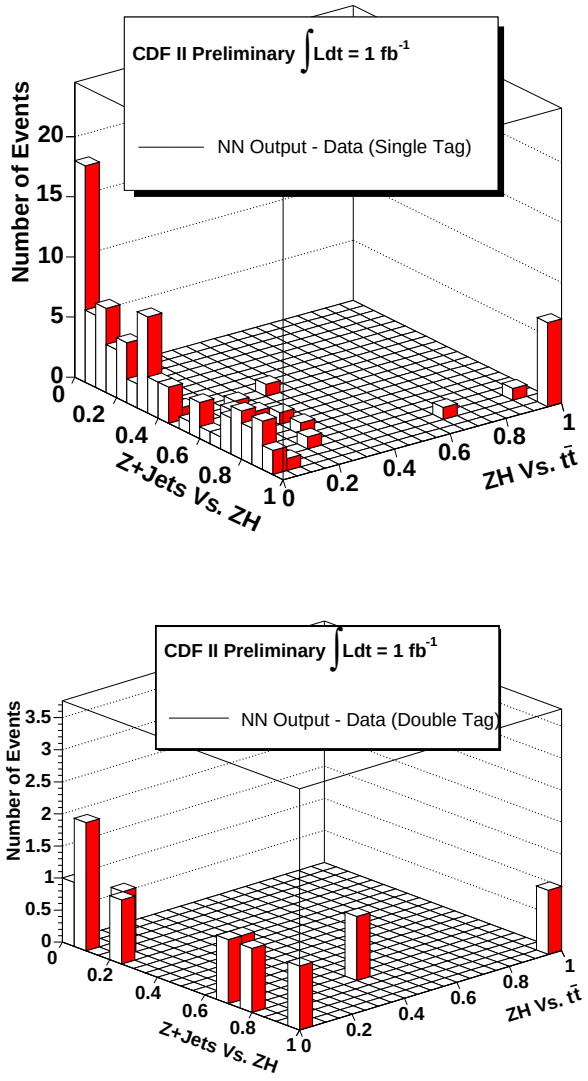

Figure 2: The NN output distribution for the single tagged (top) and double tagged (bottom) combined. The $t\bar{t}$ background is expected in the upper right corner, the $Z+$ jets in the lower left and the ZH signal in the lower right.

Systematic uncertainties and shift of σ in pb	
Background Shape (Pythia vs Alpgen)	0.17
Jet energy scale	0.08
ISR gluon radiation	0.02
FSR gluon radiation	0.04
PDF reweighting	0.00
Total systematic	0.19

Table 4: Summary of uncertainties in terms of shift of expected 95% CL upper limit exclusion limit.

show the NN projections after applying a cut on the other axes, i.e. for the x-axis projection a cut on $y_{NN} < 0.25$ is applied and for the y-axis projection a cut on $x_{NN} > 0.75$ is applied to clearly show how the data compare in the region most sensitive to ZH production. the bottom right bin shows an excess in the signal region.

Our 95% cross-section limit on σ_{ZH} is done using the technique described in CDF note 8276, Ref.[9]. This technique handles all our shape and normalization systematics. The final observed limits are shown in Table 5 and plotted in Figure 5. Also in the table is the comparison of our observed limit to the SM prediction, as well the expected limit based on pseudo-experiments.

References

[1] (The LEP Electroweak Working Group) http://lepewwg.web.cern.ch/LEPEWWG/

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

[3] J. Efron, et al. *Search for* $ZH \to \ell^+\ell^- b\bar{b}$ *in 1* fb^{-1} *of CDF Run 2 Data*, CDF note 8422.

[4] D. Acosta et al. (CDF), Phys Rev. D71, 052004 (2005), hep-ex/0410041

[5] B. Heinemann, *Summary of Most Commonly used Scale Factors for the 2006 Summer Conferences*, CDF note 8312 (and references therein)

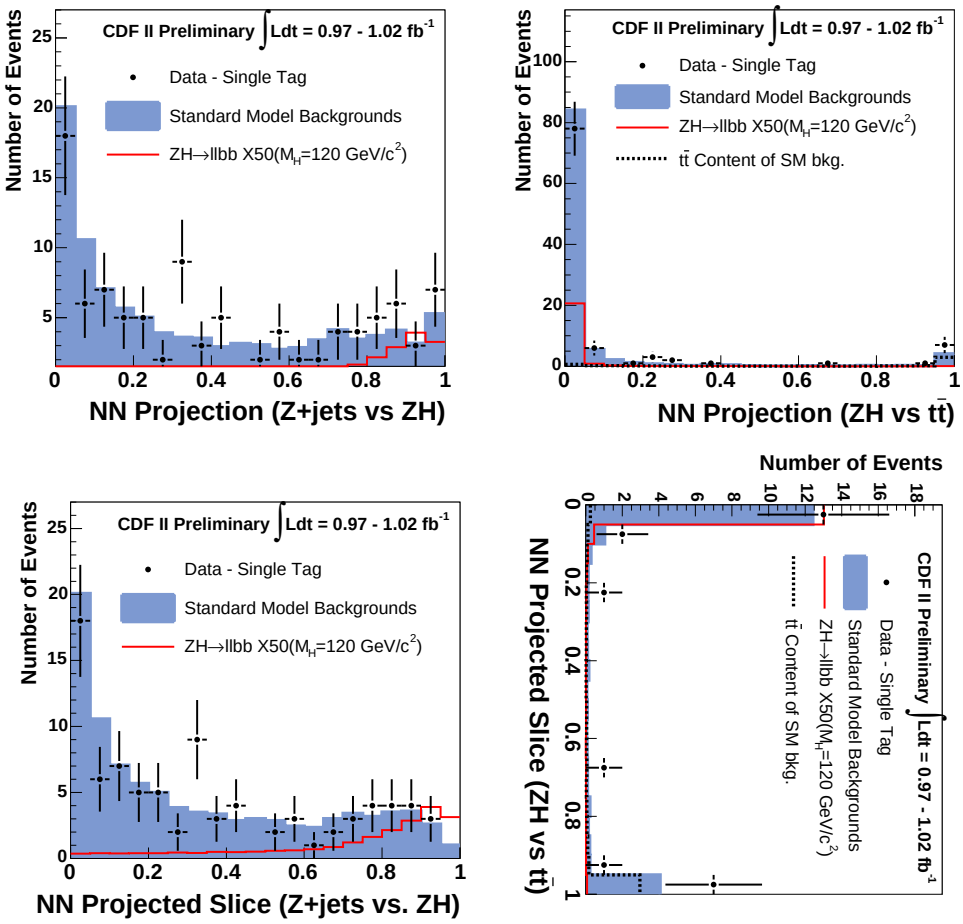

Figure 3: The NN single tagged output distribution projected onto the x-axis (top right plot) and the y-axis (top left plot) for the electron data.The NN output distribution projected onto the x-axis (bottom right plot) for $y_{NN} < 0.25$ and the y-axis (bottom left plot) for $x_{NN} > 0.75$ for the combined data. The data (points) are compared to the total Standard Model background (blue). The signal ZH shape is overlayed in red times a multiplication factor of 50.

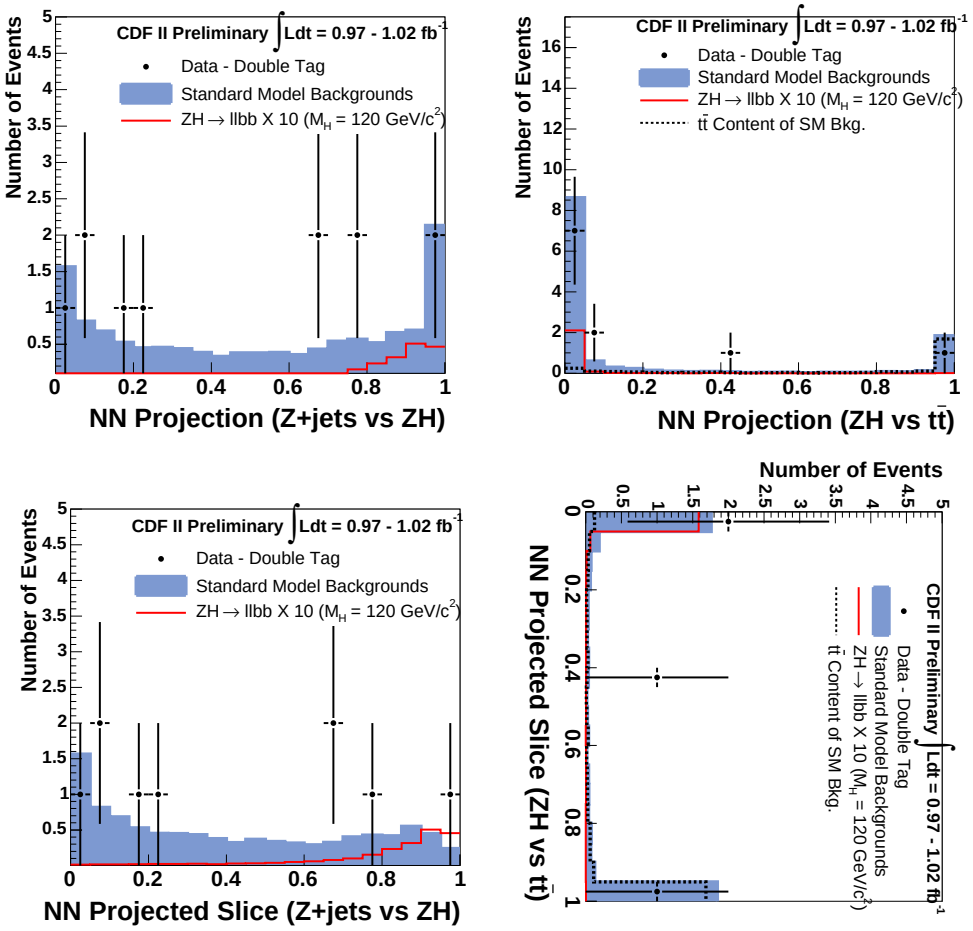

Figure 4: The NN double tagged output distribution projected onto the x-axis (top right plot) and the y-axis (top left plot) for the electron data.The NN output distribution projected onto the x-axis (bottom right plot) for $y_{NN} < 0.25$ and the y-axis (bottom left plot) for $x_{NN} > 0.75$ for the combined data. The data (points) are compared to the total Standard Model background (blue). The signal ZH shape is overlayed in red times a multiplication factor of 50.

Mass (GeV/c^2)	1-tag Obs. Limit (Exp.)	2-tag Obs. Limit (Exp.)	Comb. Obs.Limit(Exp.)
100	23 (23)	15 (17)	12 (13)
110	25 (24)	20 (19)	14 (14)
115	28 (27)	23 (22)	16 (16)
120	Old (Summer 2006 1 fb^{-1} single taggged only)		31 (26)
120	30 (31)	28 (26)	18 (18)
130	48 (49)	43 (39)	29 (28)
140	110 (96)	86 (75)	65 (54)
150	260 (250)	220 (200)	160 (140)

Table 5: Expected and observed limits from the data as a ratio compared to expected Standard Model cross-sections (95% CL XSec / SM). Single-tag and double-tag samples are shown separate and combined. The result obtained during the previous iteration of this analysis, over the same data set, is included for comparison. It was obtained by looking only in single tagged events.

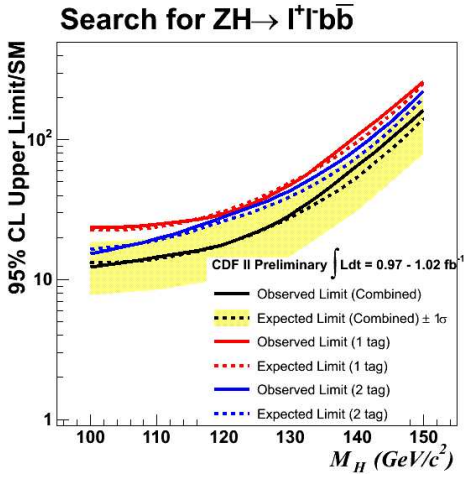

Figure 5: The 95% confidance leval upper limit on the production cross section of ZH multiplied by the branching ratio of $H \to b\bar{b}$ divided by the Standard Model prediction. The solid line represents the observed limit. The dashed limit is represents the mean expected limit predicted by psuedo experiments with the 1 *sigma* region represented by the yellow band.

[6] M. Griffiths, B. Heinemann, G. Manca, *Fake Rate For Low-p_T Leptons*, CDF note 7470

[7] A. Mehta, B. Heinemann, *Measurement of the b jet cross section for events with a Z^0 boson*, CDF note 7780

[8] S. Budd, C. Neu, *SECVTX Mistag Parameterization for Winter 2006*, CDF note 8072

[9] Y. Kusakabe, et. al, *"Combined Upper Limit on Standard Model Higgs Production at CDF"*, CDF note 8276