



## Search for the SM Higgs boson in the $\cancel{E}_T + b$ -jets signature with relaxed kinematic cuts in $9.45 \text{ fb}^{-1}$ of data at CDF.

The CDF Collaboration  
URL <http://www-cdf.fnal.gov>  
(Dated: March 13, 2012)

We present a search for the Higgs boson produced in association with a  $Z$  or  $W$  boson in the  $\cancel{E}_T + b$ -jets signature. We consider the scenario where  $Z \rightarrow \nu\nu$ , or  $W \rightarrow l\nu$  and the lepton escapes detection; the Higgs boson decays into a  $b\bar{b}$  pair. This analysis uses  $9.45 \text{ fb}^{-1}$  of data collected by the CDF II experiment at Fermilab. We analyze events with *relaxed kinematic* requirements, yielding an increase of 30-40% in acceptance to the  $WH/ZH$  signal. We collect events from three different triggers and parametrize the complex turn-on using an artificial neural network technique. To increase the sensitivity to the signal, we implement a NN to remove the huge instrumental background. An additional NN is used to discriminate the Higgs boson signal from the remaining background. We check the goodness of our background modeling by comparing data against backgrounds in many control regions, and find good agreement. Observing no significant excess in the data, we place 95% confidence level upper limits on the Higgs boson production cross section. For a mass of  $125 \text{ GeV}/c^2$  the expected (observed) limit is 3.6 (6.7) times the standard model prediction. Compared to the last iteration of this analysis, this result improves the significance by 10% throughout the 90-150  $\text{GeV}/c^2$  mass range, and is one of the most sensitive at the Tevatron for a low mass Higgs boson.

## I. INTRODUCTION

The search for the Higgs boson (H) is the most active areas of research at the Tevatron. The electroweak fits to standard model (SM) parameters, performed including the latest Tevatron top mass averaged measurements [1], point to the value  $m_H = 89_{-26}^{+35}$  GeV/ $c^2$ , or  $m_H < 158$  GeV/ $c^2$  [2]. In the mass region above 135 GeV/ $c^2$  the searches focus on  $gg \rightarrow H$  where  $H \rightarrow WW$ , because of the high cross section and the “low” backgrounds when the W’s decay leptonically. At low mass the searches focus on the production of  $H$  associated with either a  $Z$  or a  $W$  boson. It has to be noted that while both CDF and DØ have excluded the presence of the Higgs boson in the mass region  $158 < m_H < 173$  GeV/ $c^2$  [3], the low mass searches are harder because of higher backgrounds and lower signal efficiency. In fact, the individual searches in the various low mass channels have yet to reach sensitivity to the SM Higgs boson cross section. Nonetheless by combining these searches from CDF and DZero, the collaborations might have a chance to exclude or find a low mass Higgs boson.

This note describes a search for the SM Higgs boson production in association with a  $Z$  or  $W$  boson in  $\bar{p}p$  collisions at  $\sqrt{s} = 1.96$  TeV with the CDF detector at the Fermilab Tevatron; it updates the EPS2011 result using  $7.8 \text{ fb}^{-1}$  [4]. We consider a scenario where  $Z \rightarrow \nu\nu$ , or  $W \rightarrow l\nu$  and the electron or muon escape detection; the Higgs boson decays into a  $b\bar{b}$  pair. We improve the energy determination using a neural-network approach to correct the measured jet energies to the quark-energy level. We significantly relax the kinematic requirements and accept 30-40% more acceptance to the  $WH/ZH$  signal. Moreover, we collect events from three different triggers and parametrize the complex turn-on using a dedicated artificial neural network. We split the data sample into various control regions and a signal region. To avoid potential bias in the search, we test our understanding of the sample in control regions. The observed data in signal region is analyzed only after all background predictions and final event selection are determined.

The tools used in this analysis were used to measure the single top production cross-section for the first time in this channel [5]. This result was part of the observation of the single top quark by CDF [6]. We thus are at the stage where even the smallest backgrounds in this channel have been measured, apart from diboson production, which is as challenging as finding the Higgs boson, due to the low invariant mass and the low branching ratio;  $\mathcal{B}(Z \rightarrow b\bar{b}) = 0.2$  while  $\mathcal{B}(H \rightarrow b\bar{b}) = 0.75$ . Moreover, these tools, especially the QCD removing neural network, make this channel one of the most sensitive at low mass, and almost as sensitive as analyses requiring the presence of leptons.

The CDF detector is described in detail elsewhere [7].

## II. DATA SAMPLE & EVENT SELECTION

This analysis is based on an integrated luminosity of  $9.45 \text{ fb}^{-1}$ , collected with the CDF II detector between February 2002 and September 2011. The data are collected using two trigger paths firing on  $\cancel{E}_T$  plus two jets [8], and one path firing on  $\cancel{E}_T$  only. The trigger path used in past iterations of this analysis [9] has been kept for the first  $2.1 \text{ fb}^{-1}$  while we use the newer  $\cancel{E}_T$  plus two jet path for the more recent data. Additionally, we collect events with  $\cancel{E}_T > 45$  GeV in the first  $2.1 \text{ fb}^{-1}$ , and with  $\cancel{E}_T > 40$  GeV thereafter (following the upgrade of the Level-2 calorimeter trigger that can handle the higher trigger rates). These new triggers help collect 15% more data. When using multiple trigger paths, it is crucial to properly model their combined effect on the data collection, and propagate it to our simulations. For this purpose, we have developed a neural network parameterization, described below.

Jets are reconstructed from energy depositions in the calorimeter towers using a jet clustering cone algorithm with a cone size of radius  $R = \sqrt{(\Delta\varphi)^2 + (\Delta\eta)^2} = 0.4$ . Jet energies are corrected to account for effects causing mis-measurements in the jet energy such as non-linear calorimeter response, multiple beam interactions, or displacement of the event vertex from the nominal posi-

tion. We further correct the jet energies by reconstructing their four-momenta according to the H1 prescription [10]. Both the magnitude and the direction of  $\cancel{E}_T$  are recalculated after correcting the energies of jets.

The trigger efficiency is obtained from data and is used to scale the Monte-Carlo based signal and background samples to correct for event loss during data taking. The overall efficiency of the online event selection is parametrized by an artificial neural network, trained with the following inputs: the  $\cancel{E}_T$  in the event, and its direction ( $\varphi(\cancel{E}_T)$ ), three variables characterizing each jet in the event,  $E_T(j_i)$ ,  $\eta(j_i)$ , and  $\varphi(j_i)$ , and the separation in the  $\eta - \varphi$  plane between the jets. We thus have 9 (14) input variables for events with two (three) jets. This is necessary to properly apply the complex turn-on observed in data to the Monte Carlo, especially because relaxing the kinematic cuts has moved the trigger operation point away from the fully-efficient region.

From this inclusive dataset we select events offline with the following requirements:

- a  $\cancel{E}_T > 35$  GeV to avoid low trigger efficiencies and a too large increase in the backgrounds;
- the two leading jets within  $|\eta| < 2.0$ , and of which at least one is central, i.e. having  $|\eta| < 0.9$ ;
- the leading jet with a  $E_T > 25$  GeV, and the second jet with  $E_T > 20$  GeV;
- a separation in the  $\eta - \varphi$  plane between the two leading jets of  $\Delta R(1^{st} \text{ jet}, 2^{nd} \text{ jet}) > 0.8$ ;
- no events with 4 or more jets with  $E_T > 15$  GeV in the  $\eta < 2.4$  region.

The main motivation to accept events with three jets in this signature is to allow events where one of the  $b$  quarks coming from the Higgs boson radiates a gluon. In addition to that, we also accept  $WH$  events where the charged lepton coming from the  $W$  is reconstructed as a jet. The latter case happens when the  $W$  decays to  $e\nu$  and the electron fails the CDF electron identification algorithm, but is reconstructed as a jet; or when the  $W$  decays to  $\tau\nu$  and  $\tau \rightarrow \text{hadrons}$ . Table I shows the contributions in signal region from WH processes in 2 and 3 jet events.

Process	All events		$e, \tau$ matched jet	
	2 jet	3 jet	2 jet	3 jet
$W \rightarrow \tau\nu$	44%	61%	2.8%	33%
$W \rightarrow e\nu$	38%	25%	0.6%	4%
$W \rightarrow \mu\nu$	18%	14%	—	—

TABLE I: Contributions to 2/3jet events from different decay modes of the W-boson in WH events

The major drawback of accepting three-jet events lies in the increase of QCD multijet production and pair produced top background; the latter background is a secondary one at this point and can be dealt with at later stages in the analysis.

As a way to get a better estimate of the event true missing energy we calculate the  $\cancel{p}_T$ , which is defined as negative vector sum of charged particle track  $p_T$ 's. For true  $\cancel{E}_T$  events  $\cancel{p}_T$  is highly correlated with calorimeter  $\cancel{E}_T$ , while for QCD events with mis-measured jets it is not. Thus,  $\cancel{p}_T$  provides an additional handle to separate mis-measurements from real  $\cancel{E}_T$  events.

### A. Jet Energy Corrections

We reconstruct jets using a jet-energy algorithm where the calorimeter energies are adjusted according to the energy inferred from tracks within the jet cone (similar to the H1 algorithm).

Although such a method improved the jet energy resolution by accounting for energy leakage, various assumptions were made that led to less than optimal determination of the jet energy. We improve the energy determination using a neural-network approach to correct the measured jet energies to the quark-energy level.

The inputs to the jet energy neural network  $NN_{JER}$  are listed in table II. Each of the variables is well-validated in a dedicated control region. The neural network is trained using the two most energetic jets from  $ZH \rightarrow \ell^+ \ell^- b\bar{b}$ ,  $WH \rightarrow \ell \nu b\bar{b}$ , and  $ZH \rightarrow \nu \bar{\nu} b\bar{b}$  signal events for Higgs masses in the full low-mass Higgs range of  $100 \leq m_H \leq 150$  GeV/ $c^2$  at 5 GeV increments. To avoid correcting jet energies to an average characteristic energy, each training event receives a weight such that the cumulative jet  $E_T$  distribution is flat.

Variable	Description
Raw $E_T$	Uncorrected transverse jet energy
L5 $m_T$	Transverse jet mass corrected to hadronic level
H1 $E_T$	H1-corrected transverse jet energy
$\pi^0$ Energy	CES detector energies of $\pi^0$ candidates within jet cone
EM Fraction	Fraction of jet energy collected in EM calorimeter
Jet $\eta$	Jet pseudorapidity
Maximum Track $p_T$	Maximum transverse momentum of track within jet cone
Sum Track $p_T$	Linear sum of transverse momenta of tracks within jet cone

TABLE II: Description of the  $NN_{JER}$  input variables.

For this analysis, the  $NN_{JER}$  algorithm is implemented by correcting the H1-determined  $\hat{E}_T$  to an  $NN_{JER}$  determination. An analysis cut of  $\hat{E}_T(NN_{JER}) > 35$  GeV is applied to reduce the QCD multijet background (figure 2), which is the dominant contribution in this analysis. The  $NN_{JER}$  algorithm separates well the instrumental  $\hat{E}_T$  from the QCD multijet background and the real  $\hat{E}_T$  of the Higgs boson signal.

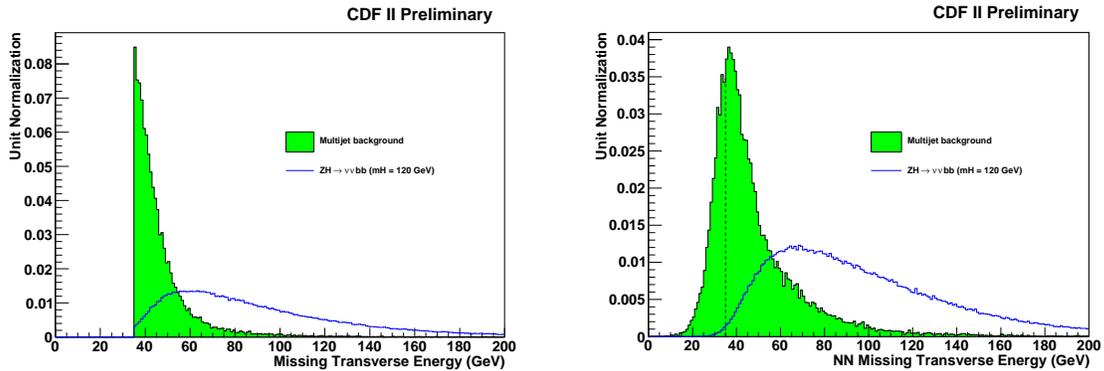


FIG. 1: H1 and  $NN_{JER}\hat{E}_T$  distributions for multijet background and  $ZH \rightarrow \nu \nu b\bar{b}$  events in the preselection region. The dashed vertical line at 35 GeV represents the additional analysis cut.

## B. Tagging Algorithms

In order to improve the signal to background further, we need to identify (tag) jets originating from a  $b$  quark. We do so by employing both the SECVTX [11] and JETPROB [12]  $b$ -tagging algorithms. The first identifies the displaced (secondary) vertex where the  $b$  quark decayed, while the second estimates the probability for a jet to not originate from the secondary vertex. When using the SECVTX (JETPROB) algorithm, we consider the “tight” operating point (a probability of less than 5% for the jet to originate from the primary vertex). We subdivide the sample into three orthogonal flavor categories: SS, where both jets are tagged by SECVTX; SJ, where one jet is tagged by SECVTX and the other by JETPROB; and 1S, where one jet – and only one – is tagged by SECVTX. The double-tagged samples provide the most sensitivity in this analysis. In addition to that the single-tagged sample adds  $\sim 10\%$  to the overall sensitivity.

## C. Neural network to remove QCD

The main background in this search is the QCD production of two or three jets. We investigated the dynamics of the events in the sample, using a QCD heavy flavor Monte-Carlo simulation. Looking at a large set of variables, we keep here only the ones for which QCD has a very different behavior with respect to the signal and the remaining backgrounds; the idea is that we will *remove* events that are not signal-like with an artificial neural network ( $NN$ ), and then use a second  $NN$  to *discriminate* signal from the surviving, more signal-like, backgrounds. The  $NN$  presented here is an improved version with that from the previous iterations, rejecting more background globally while still rejecting most of the QCD multijet and keeping high signal acceptance. This approach to remove QCD backgrounds has been successfully used in the search for other signals, such as single top [5].

We train a mixture of 50%  $WH$  events and 50%  $ZH$  events with  $m_H = 115 \text{ GeV}/c^2$  against pre-tagged data weighted by our *Tag-Rate-Function*. The latter is the first step of our data driven technique (described below) to estimate this type of background. The variables used in the training are shown in table III. We use the Multi Layer Perceptron (MLP) which is a simple feed-forward network, as implemented inside the TMVA [13] package (v4.1.0). We will refer to the output of this network as  $NN_{QCD}$ .

We use the output of this NN to define the final signal region. This cut reduces QCD by about one order of magnitude while keeping the signal efficiency between 90 and 95% (table IV).

Variable
Magnitude of $\vec{E}_T$
Magnitude of $\vec{\cancel{p}}_T$
$E_T / \sqrt{\sum E_T}$
$E_T / H_T$
$H_T / E_T$
$M(\vec{E}_T, \vec{j}_1, \vec{j}_2)$
$\Delta\varphi$ between $\vec{E}_T$ and $\vec{\cancel{p}}_T$
Maximum of $\Delta\varphi$ between any two jets
Maximum of $\Delta R$ between any two jets
Minimum of $\Delta\varphi$ between the $\vec{E}_T$ and $\vec{j}_i$
Minimum of $\Delta\varphi$ between the $\vec{\cancel{p}}_T$ and $\vec{j}_i$
$\Delta\varphi(\vec{j}_1, \vec{j}_2)$ in the 2-jet rest frame
Sphericity
Centrality

TABLE III: Input variables to the neural network devised to suppress the QCD background, and the background coming from production of light flavor jets.

Cat.	Sig. Acc.	Bkg. Rej.	QCD Rej.
1S	90.0%	72.6%	89.1%
SS	94.9%	48.2%	87.0%
SJ	93.6%	64.8%	87.4%

TABLE IV: Performance of the  $NN_{QCD}$  when discarding events with  $NN_{QCD} \leq 0.45$ .

### III. SIGNAL AND BACKGROUND MODELING

#### A. Signal Modeling

The signal Monte Carlo samples are generated with PYTHIA [14]. The  $ZH/WH$  processes were generated for Higgs boson masses ranging from 100 GeV/ $c^2$  to 150 GeV/ $c^2$  in 5 GeV/ $c^2$  steps. The cross-sections are corrected for NNLO effects by a k-Factor of 0.99 in case of  $ZH$  production and 0.96 for  $WH$  production [15]. In these samples the Higgs boson is forced to decay into  $b$ -jet pairs, the  $Z$  boson to neutrinos or a pair of charged leptons, and the  $W$  decays to leptons. We use  $\mathcal{B}(Z \rightarrow \nu\nu) = 0.200$ ,  $\mathcal{B}(Z \rightarrow \ell\ell) = 0.03$  and  $\mathcal{B}(W \rightarrow \ell\nu) = 0.324$ .

#### B. Background Modeling

In the signal events the Higgs boson decays into two  $b$ -jets, the  $Z$  boson into two neutrinos, and the  $W$  to leptons. The most important characteristics of these events are the large intrinsic missing transverse energy, relatively low jet multiplicity, and the lack of (detectable) isolated leptons. There are numerous standard model processes that can produce this signature.

The most significant background at the first stage of the analysis is the QCD multijet processes. QCD jet production has a large cross-section ( $\sim \mu\text{b}$ ), which is about 9 orders of magnitude greater than the signal before requiring the first  $b$ -tag. Although, these processes generally do not have

intrinsic  $\cancel{E}_T$ , mis-measured jets do cause imbalance in the total transverse energy by which the QCD events can pass the basic selection cuts if one of the jets is mis-tagged. Furthermore, QCD  $b$ -quark pair production yields taggable jets and if one  $b$  undergoes a semi-leptonic decay large  $\cancel{E}_T$ . In both cases, the missing transverse energy tends to be aligned parallel or anti-parallel to the first or second most energetic jet. This topology provides us one of the most effective devices against the QCD background.

To estimate the QCD background from data, we have developed a *Tag-Rate-Matrix* (TRM) method, which allows us to estimate heavy flavor QCD production. We use the *pre-tag* data sample to estimate the probability for an event to be in one of our three flavor categories. This is a major change since the last iteration, allowed by the much larger amount of collected data: instead of using a *per-jet* parameterization with the problem of dealing with the correlation among the jets [19], we now define *per-event* matrices. The tag rate probabilities are parameterized as a function of the scalar sum of all jets in the event ( $H_T$ ), and the transverse momentum imbalance ( $\cancel{p}_T$ ), and the ratio of the sum of the  $p_T$  of good SECVTX tracks to the jet  $p_T$  ( $\sum p_{T_{chgd}}^{j_1}/p_{T}^{j_1}$  and  $\sum p_{T_{chgd}}^{j_2}/p_{T}^{j_2}$ ). We define the matrix from a subsample of the  $\cancel{E}_T$ +jets dataset which is orthogonal to the final signal sample: we select events with  $35 < \cancel{E}_T < 70$  GeV, an azimuthal angular separation between the direction of the second leading jet and that of the  $\cancel{E}_T$ ,  $\varphi(j_2, \cancel{E}_T) \leq 0.4$ , and no identified leptons (vetoed loose lepton identification).

Two classes of top-production are considered in this analysis: the pair-production (PYTHIA) and the electroweak single top-production in the s- and t-channels (POWHEG). They both yield a significant contribution to the background in the signal region. Due to the large mass and the semi-leptonic decay of the top, these events are energetic, bear large  $\cancel{E}_T$  and high jet multiplicity. In the diboson samples (PYTHIA), the bosons' decays are inclusive. In the  $W/Z$  + jets samples (ALPGEN), the bosons are forced to decay into leptons, or  $b$ -quarks. The parton showering is done by PYTHIA.

In fact, the multijet component is a combination of different processes, which only share a same reconstructed state. On one side are the QCD-produced events, and on the other are electroweak events. Both types of events either do or do not contain a heavy flavor quark (and hence a  $b$ -jet). The data-driven model described above describes in fact all these four components together. To avoid double-counting events that we model using Monte Carlo, we subtract from our data-driven prediction the pre-tag Monte Carlo events weighted by our matrix. By doing so, we only retain the QCD part of the multijet events. But we also have removed events containing no  $b$  quark but identified as possessing a  $b$ -jet (mis-tags). We re-introduce these by weighing the same Monte Carlo by the probability for a mis-tagging to occur. The latter is obtained from the estimation of the negative tag rate, assumed to be solely due to resolution effects (the positive tag rate also includes a contribution from real heavy flavor jets) [12]. This approach allows us to model the QCD component of the multijet independently from the light flavor electroweak events, which is beneficial in the sense that the region in which we derive the TRM has a different ration between these two categories of events than the signal region.

We check our modeling of the data sample for all flavor categories in three control regions which are defined below and observe excellent agreement.

### C. Multijet Background Normalization

In order to estimate the backgrounds originating from QCD heavy flavor multijet production, we use the TRM method described above. This method provides an excellent model to describe the shapes of the background. Though better than the jet-based approach used in the past, the normalization of the background must be determined in a kinematic region closer to the final signal region than the region the *Tag-Rate-Matrix* was derived in.

In order to constrain the expected rates of these backgrounds we utilize the lower region of our  $NN_{QCD}$  output ( $NN_{QCD} < 0.1$ ). This region is rich in QCD events and has sufficient statistics.

Once we are confident that the shapes are well reproduced by the matrix, we extract the normalization factor used in the final measurement by scaling the multi-jet prediction to data minus MC background minus mis-tags. The scale factor is close to – and compatible with – 1.0.

#### IV. CONTROL REGIONS

We check our ability to predict the multijet backgrounds in two control regions (CR). QCD CR 1 is a high statistics region where we check the data-based model and evaluate the systematic uncertainties on the shapes of the various kinematic variables. In order to test our data-driven estimation in a more signal-like region, we define QCD CR 2. This region is defined by reversing the cut on  $NN_{QCD}$  to remain blind to the signal region. One part of this region,  $NN_{QCD} < 0.1$ , is used to extract the multijet normalization as described above. The region with  $0.1 < NN_{QCD} < 0.45$  serves as a medium statistics cross check of the multijet normalization with non-negligible electroweak contribution.

Since in the signal region we expect backgrounds originating from events with real high  $\cancel{E}_T$ , such as  $W/Z$ +jets,  $t\bar{t}$ , single top production and diboson production, we test our ability to predict these in an other control region, EWK CR. To remain unbiased to our final region, we test electroweak/top backgrounds in the kinematic region similar to signal region, with the exception of requiring at least one lepton in the event (all events with leptons are rejected from the signal region).

Table V summarizes the information regarding each control region and the final signal region. Table VI lists the *prior* expected and observed event yields in signal region. The posterior yields are very close to the prior expected yields. This is an indication of the quality of our modeling and its ability to adequately represent the data.

Comparisons of kinematic distributions in all control regions and in the signal region in all flavor categories are shown at CDF public web-page, accessible from:

<http://www-cdf.fnal.gov/physics/new/hdg/hdg.html>

QCD CR1	EWK CR	QCD CR2	Signal region
No lepton	At least one lepton	No lepton	No lepton
$\cancel{E}_T > 70$ GeV	$\cancel{E}_T > 35$ GeV	$\cancel{E}_T > 35$ GeV	$\cancel{E}_T > 35$ GeV
$\varphi(j_2, \cancel{E}_T) \leq 0.4$	$\varphi(j_2, \cancel{E}_T) > 0.4$	$\varphi(j_2, \cancel{E}_T) > 0.4$	$\varphi(j_2, \cancel{E}_T) > 0.4$
	$\varphi(j_3, \cancel{E}_T) > 0.4$	$\varphi(j_3, \cancel{E}_T) > 0.4$	$\varphi(j_3, \cancel{E}_T) > 0.4$
	$\varphi(j_1, \cancel{E}_T) > 1.5$	$\varphi(j_1, \cancel{E}_T) > 1.5$	$\varphi(j_1, \cancel{E}_T) > 1.5$
		$NN_{QCD} < 0.45$	$NN_{QCD} > 0.45$

TABLE V: Main kinematic selection requirements for each of the control regions and the signal region.

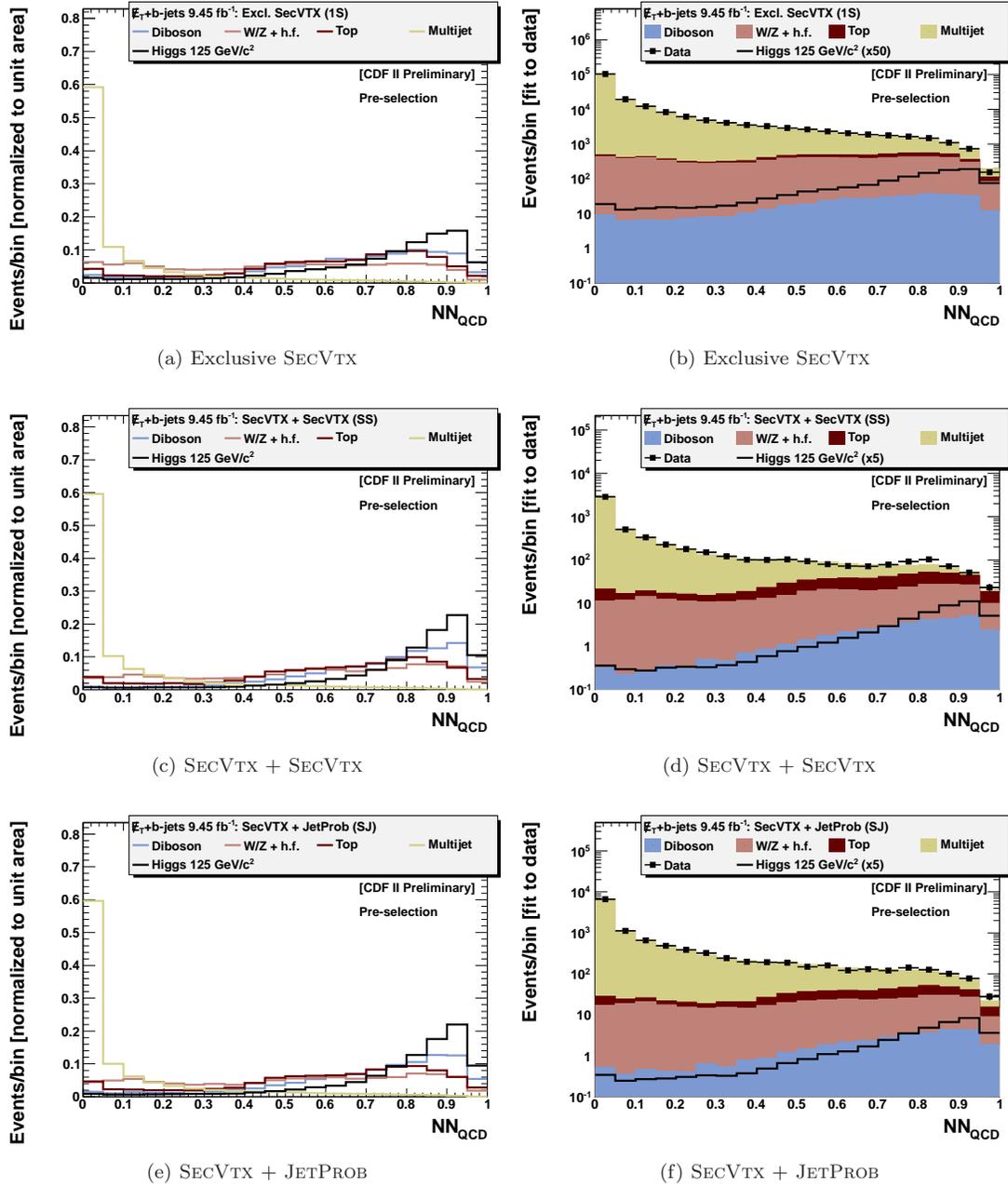


FIG. 2: QCD rejection neural network output. We retain events having  $NN_{QCD} > 0.45$  for the signal extraction.

$\cancel{E}_T + b\text{-jets } 9.45 \text{ fb}^{-1} \text{ [CDF II Preliminary]}$									
	1S			SS			SJ		
$WW$	158.8	$\pm$	17.2	0.8	$\pm$	0.1	2.5	$\pm$	0.3
$WZ/ZZ$	133.9	$\pm$	14.3	31.3	$\pm$	3.9	27.6	$\pm$	3.2
Single Top	273.6	$\pm$	35.2	48.3	$\pm$	7	41	$\pm$	5.6
Top Pair	741.5	$\pm$	93.1	147.2	$\pm$	21.1	133.3	$\pm$	18
$Z + h.f.$	812.2	$\pm$	146.2	73.6	$\pm$	14.2	72.4	$\pm$	13.5
$W + h.f.$	2868.1	$\pm$	528.8	123.5	$\pm$	23.9	154.4	$\pm$	29.3
QCD Multijet	10824.6	$\pm$	177.3	376.9	$\pm$	11.9	923.1	$\pm$	19.2
EWK Mistags	2287.8	$\pm$	283	16.5	$\pm$	5.4	38.4	$\pm$	20.3
Total	18100.6	$\pm$	1295.1	818	$\pm$	87.5	1392.7	$\pm$	109.5
Data	18165			807			1310		
$m_H = 90 \text{ GeV}/c^2$	49			20.6			15.9		
$m_H = 95 \text{ GeV}/c^2$	45			19.2			14.9		
$m_H = 100 \text{ GeV}/c^2$	41.3			17.6			13.6		
$m_H = 105 \text{ GeV}/c^2$	36.6			15.6			12.1		
$m_H = 110 \text{ GeV}/c^2$	32.2			13.8			10.7		
$m_H = 115 \text{ GeV}/c^2$	31.1			13.5			10.4		
$m_H = 120 \text{ GeV}/c^2$	25.8			11.2			8.6		
$m_H = 125 \text{ GeV}/c^2$	21			9.1			7		
$m_H = 130 \text{ GeV}/c^2$	16.1			7.1			5.4		
$m_H = 135 \text{ GeV}/c^2$	11.9			5.2			4		
$m_H = 140 \text{ GeV}/c^2$	8.4			3.6			2.8		
$m_H = 145 \text{ GeV}/c^2$	5.5			2.4			1.9		
$m_H = 150 \text{ GeV}/c^2$	3.4			1.5			1.2		

TABLE VI: Number of expected and observed events in the signal region in all flavor categories.

## V. THE SEARCH FOR THE SIGNAL

As mentioned above, we selected the signal region to maximize signal significance keeping high signal efficiency. The biggest background rejected is QCD events faking high  $\cancel{E}_T$ . The dominating backgrounds at this point are QCD, mis-tags,  $W/Z$ +jets and  $t\bar{t}$  in similar proportions. We study the dynamic of those events to develop a NN with the goal of discriminating the surviving backgrounds from the interesting signal.

### A. A second NN to discriminate the signal from the backgrounds

Since the background composition is different in events with 2 or 3 jets, we train separate networks in each category. The outputs of these networks are combined in the end, when searching for the signal. For the NN training of 2-jet (3-jet) events we use a background sample made of 75% (50%) of  $\cancel{E}_T$ +jets untagged data (none of the jets in the event are  $b$ -tagged) and 25% (50%) of  $t\bar{t}$  events. The Higgs boson signal used for the training is a mixture of 50%  $WH$  events and 50%  $ZH$  events. We train one network for each of the ten mass points we probe (from 100 to 150 GeV/ $c^2$  in steps of 5 GeV/ $c^2$ ). In order to increase the separating power of the NN, we implement a track-based discriminant, TRACKMET, which was trained to optimize the separation of both  $ZH$  and  $WH$  events from QCD and  $t\bar{t}$  backgrounds. A detailed description of the method can be found in [17].

The neural network chosen here is once more the Multi Layer Perceptron (MLP). The 7 input variables are presented in Table VII. Figure 4 shows the NN output which we will use to scan for the presence of a signal.

Variable
Invariant mass of the two leading jets in the event ( $M_{jj}$ )
Invariant mass of $\vec{E}_T$ , $\vec{j}_1$ and $\vec{j}_2$
Difference between the scalar sum of transverse energy of the jets ( $H_T$ ) and $\cancel{E}_T$
Difference between the vector sum of transverse energy of the jets ( $\vec{H}_T$ ) and $\vec{E}_T$
The output of the TRACKMET neural network
Maximum of the difference in the $\eta - \phi$ space between the directions of two jets, taking two jets at the time
The output of $NN_{QCD}$

TABLE VII: Input variables to the final discriminant neural network.

### B. Systematic Uncertainties

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: QCD multijet normalization (0.7% in single tagged, 2.5% in SECVTX+SECVTX, 1.6% in SECVTX+JETPROB samples), MC statistical fluctuations. Additionally, the statistical variations in TRM, which is used to estimate the multijet background, can also modify the distributions. It is taken into account by varying the TRM probability in each bin of the matrix by  $\pm 1\sigma$ , and the alternative shapes are used in the limit calculation. The correlated systematics are: luminosity (6.0%),  $b$ -tagging efficiency scale factor between data and Monte Carlo (5.2% for single and 10.4% for SECVTX+SECVTX, 8.3% for SECVTX+JETPROB samples), trigger efficiency (<3%), lepton veto

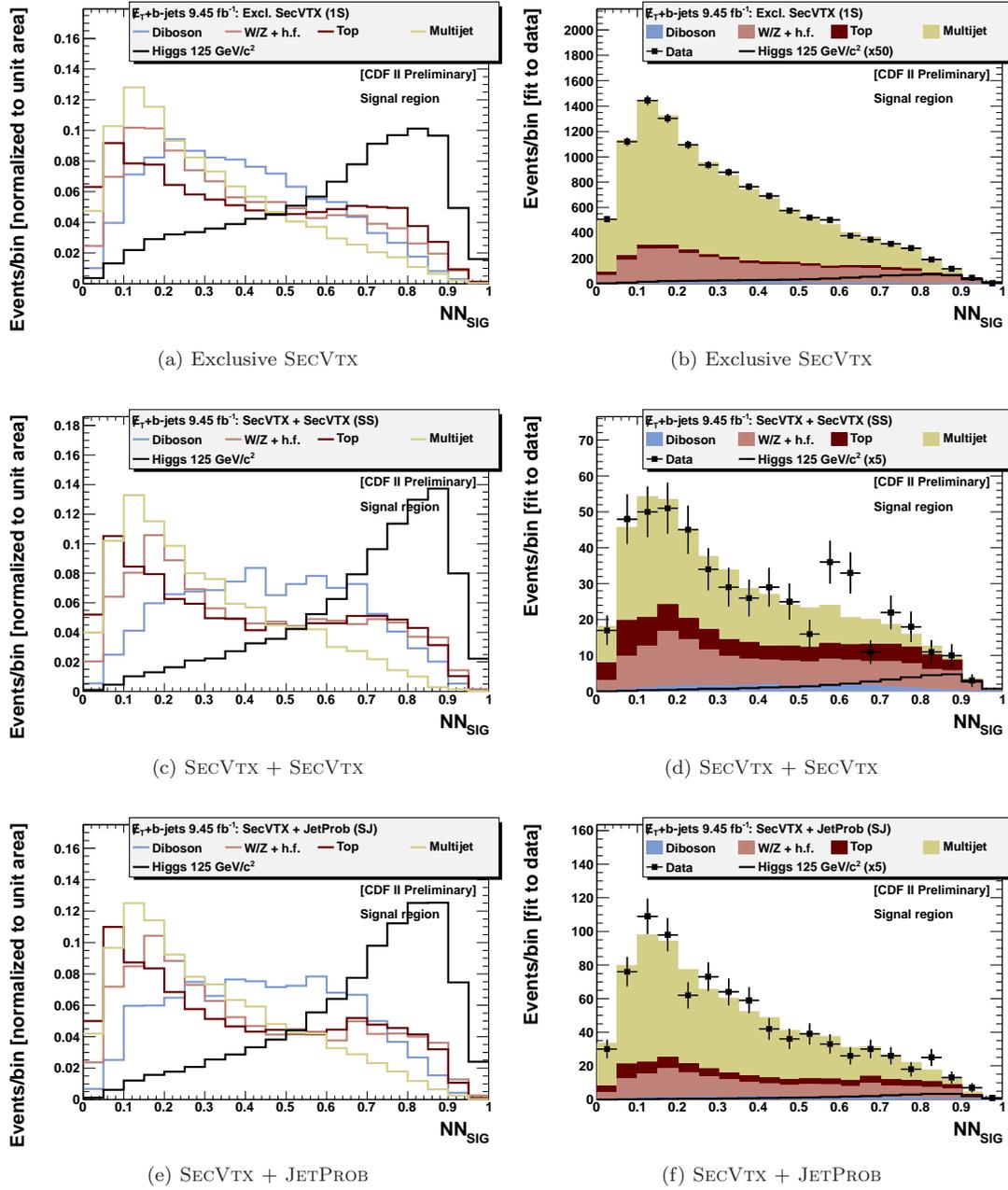


FIG. 3: Final discriminant  $NN_{\text{SIG}}$  output distribution in signal region. with the binning used to perform the likelihood fit.

efficiency (2%), PDF uncertainty (3%) and jet energy scale (with un-correlated uncertainties for quark- and gluon-jets[20]). ISR/FSR systematic uncertainties (between 2% and 3%) are applied on the signal.

$\cancel{E}_T+b\text{-jets } 9.45 \text{ fb}^{-1} \text{ [CDF II Preliminary]}$		
$m_H$	Expected	Observed
90	$1.83^{+0.79}_{-0.53}$	1.84
95	$2.16^{+0.82}_{-0.69}$	1.73
100	$2.33^{+0.94}_{-0.67}$	2.18
105	$2.57^{+1.02}_{-0.73}$	2.15
110	$2.71^{+1.18}_{-0.80}$	2.67
115	$2.74^{+1.16}_{-0.80}$	2.70
120	$3.11^{+1.27}_{-0.86}$	3.90
125	$3.62^{+1.44}_{-1.02}$	6.75
130	$4.56^{+1.88}_{-1.37}$	6.71
135	$5.99^{+2.46}_{-1.77}$	8.93
140	$7.98^{+3.38}_{-2.28}$	12.72
145	$11.80^{+4.80}_{-3.42}$	19.17
150	$18.43^{+7.65}_{-5.23}$	27.19

TABLE VIII: The predicted and observed cross-section limits of the  $ZH$  and  $WH$  processes combined when  $H \rightarrow b\bar{b}$  divided by the SM cross-section

### C. Results

Observing no significant excess in the data, we place 95% confidence level upper limits on the Higgs boson production cross section. Considering the systematic uncertainties listed above, we computed the expected limit for the Higgs boson cross-section when the Higgs boson is produced with a  $Z/W$  boson and decays to two  $b$  quarks where  $Z$  decays to neutrinos and  $W$  to leptons. We use a Bayesian method for deriving the limits [18]. Table VIII shows the final results. All the cross-sections times branching fraction are ratios with respect to the standard model cross-section.

### D. Diboson interpretation

The production of  $WZ$  and  $ZZ$  boson pairs where  $Z \rightarrow b\bar{b}$  is an important test of the electroweak sector of the SM. The production rate for these processes is higher than that of  $WH$  and  $ZH$  associated production, and thus the measurement of  $\sigma(WZ + ZZ)$  using the tools designed for the Higgs boson search can provide a powerful validation of the  $VH \rightarrow \cancel{E}_T b\bar{b}$  results. We perform the measurement using the same event selection and tools as described in this note.

Figure 6 shows three discriminants used to measure  $\sigma(WZ + ZZ)$  in the  $\cancel{E}_T+b\text{-jets}$  signature; the outcome of the measurements are shown in figure 7. All three measurements are compatible with the SM expectation, the most significant being  $\sigma(WZ + ZZ)/\text{SM} = 0.7 + 0.5 - 0.4$ ;

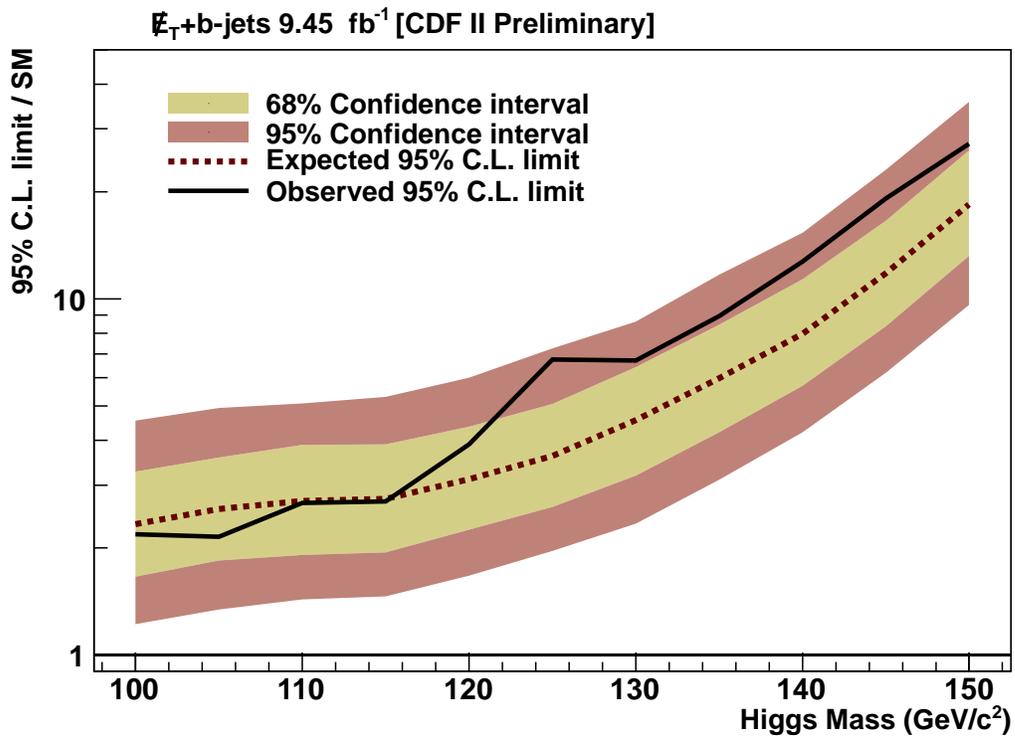


FIG. 4: 95% C.L. exclusion limits in the  $VH \rightarrow \cancel{E}_T b\bar{b}$  channel divided by the SM cross-section

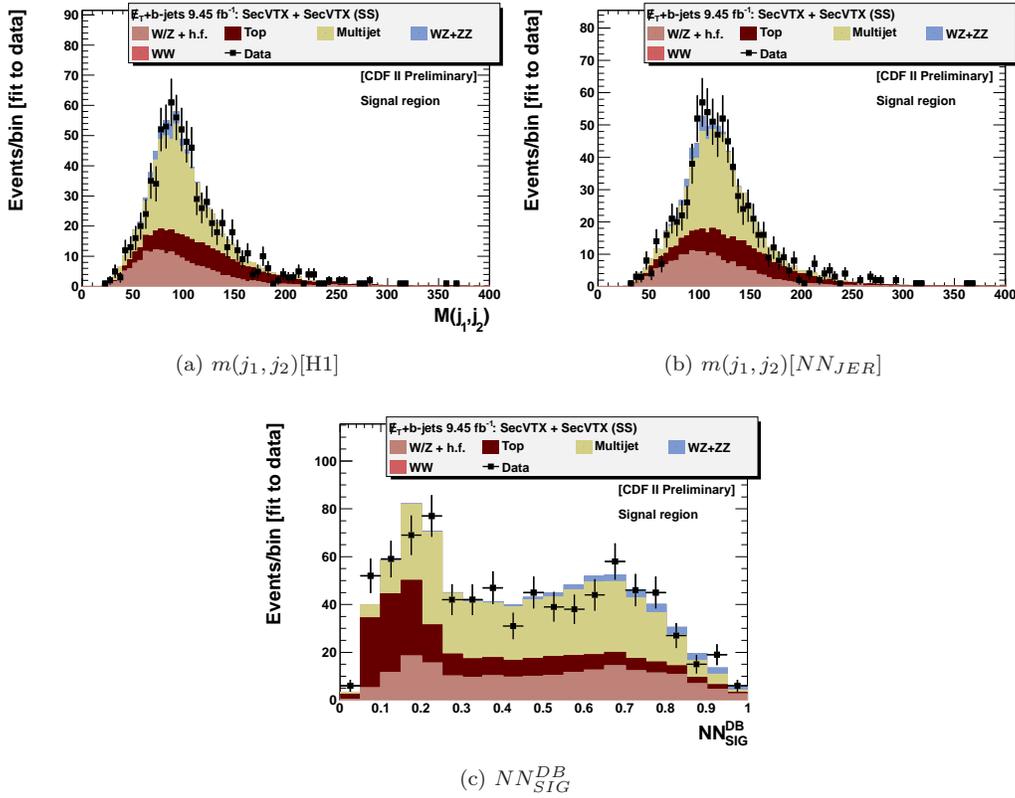


FIG. 5: Discriminant functions for measuring  $\sigma(WZ + ZZ)$ , ordered by sensitivity to the  $WZ + ZZ$  signal.

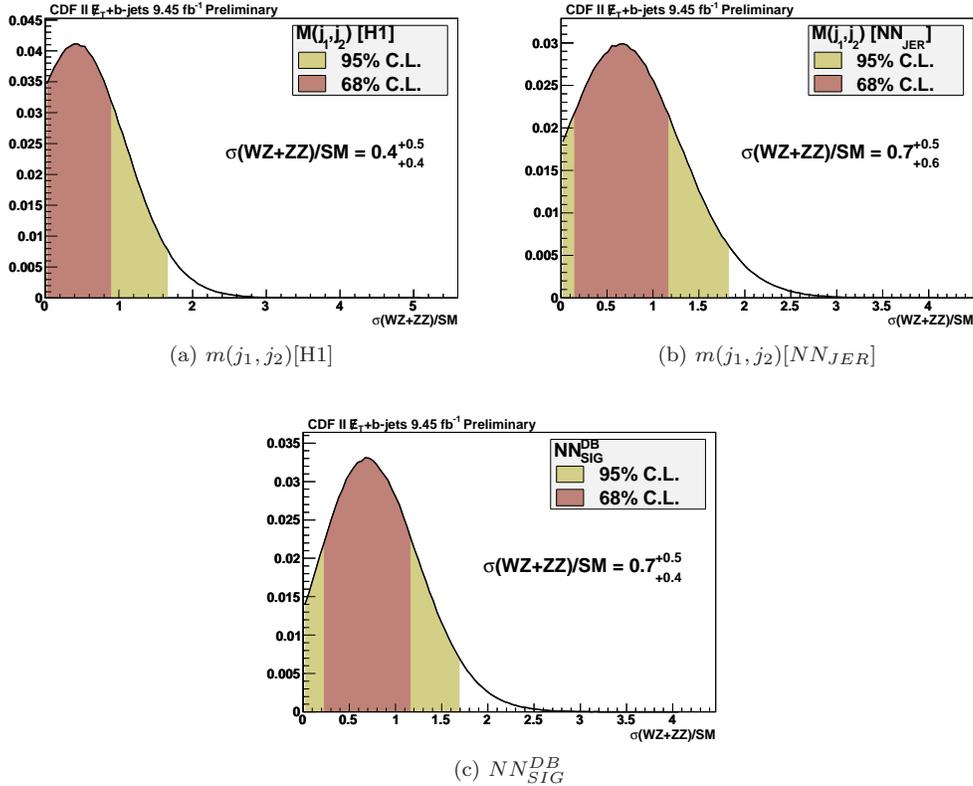


FIG. 6: Posterior probability distributions for  $\sigma(WZ + ZZ)$  for each discriminant function.  $NN_{SIG}^{DB}$  is the most sensitive;  $m(j_1, j_2)[NN_{JER}]$  significantly improves over  $m(j_1, j_2)[H1]$ .

## VI. SUMMARY

We have presented an updated search of the Standard Model Higgs boson in events  $VH \rightarrow \cancel{E}_T + b\bar{b}$  using  $9.45 \text{ fb}^{-1}$  of CDF data. We correct the jet energies using a  $NN_{JER}$  and use a  $NN$  to suppress the dominant QCD background. An additional  $NN$  is used to discriminate the signal from the surviving backgrounds. We expect to set a limit on the SM Higgs boson cross section times the branching ratio of 3.6 in the hypothesis of  $m_H = 125 \text{ GeV}$ . In absence of a significant signal excess, we observe a limit of 6.7 times the SM prediction. This result is one of the most sensitive at the Tevatron.

## Acknowledgments

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 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 fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

- 
- [1] The Tevatron Electroweak Working Group, "Combination of CDF and D0 Results on the Mass of the Top Quark", FERMILAB-TM-2466-E, TEVEWWG-TOP-2010-07, CDF-NOTE-10210, D0-NOTE-6090, 2010 [arXiv:1007.3178]
  - [2] LEP Electroweak Working Group, <http://lepewwg.web.cern.ch/LEPEWWG/>
  - [3] The TEVNPWWG Working Group, "Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to  $10 \text{ fb}^{-1}$  of Data", FERMILAB-CONF-12-065-E, CDF-NOTE-10806, D0-NOTE-6303, 2012
  - [4] The CDF Collaboration, CDF-NOTE-10583
  - [5] Aaltonen, T. *et al.*, "Search for single top quark production in pbar p collisions at  $\sqrt{s}=1.96 \text{ TeV}$  in the missing transverse energy plus jets topology", Phys. Rev. D **81**, 072003 (2010) [arXiv:hep-ex/1001.4577]
  - [6] T. Aaltonen *et al.*, The CDF Collaboration, "Observation of Electroweak Single Top Quark Production", FERMILAB-PUB-09-059-E, arXiv:0903.0885
  - [7] F. Abe, *et al.*, Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988); D. Amidei, *et al.*, Nucl. Instrum. Methods Phys. Res. A **350**, 73 (1994); F. Abe, *et al.*, Phys. Rev. D **52**, 4784 (1995); P. Azzi, *et al.*, Nucl. Instrum. Methods Phys. Res. A **360**, 137 (1995); The CDFII Detector Technical Design Report, Fermilab-Pub-96/390-E
  - [8] CDF Collaboration, "Searches for direct pair production of supersymmetric top and supersymmetric bottom quarks in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96 \text{ TeV}$ ", Phys. Rev. D **76**, 072010 (2007)
  - [9] CDF Collaboration, "A Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and  $b$ -quark Jets in  $p\bar{p}$  Collisions at  $\sqrt{s} = 1.96 \text{ TeV}$ ", Phys. Rev. Lett. **104**, 141801 (2010), FERMILAB-PUB-09-586-E, 2010 [arXiv:0911.3935]
  - [10] H1 Collaboration, C. Adloff *et al.*, Z. Phys. C **74** (1997) 221.
  - [11] A. Acosta *et al.*, Phys. Rev. D **71**, 052003 (2005)
  - [12] A. Abulencia *et al.*, Phys. Rev. D **74**, 072006 (2006)
  - [13] A. Hocker *et al.*, "TMVA: Toolkit for multivariate data analysis," arXiv:physics/0703039.

- [14] T. Sjostrand et al., High-Energy-Physics Event Generation with PYTHIA 6.1, *Comput. Phys. Commun.* **135**, 238 (2001).
- [15] See: [http://www-d0.fnal.gov/~msanders/Higgs/higgsXsec\\_WadeFisher.txt](http://www-d0.fnal.gov/~msanders/Higgs/higgsXsec_WadeFisher.txt)
- [16] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, “ALPGEN, a generator for hard multiparton processes in hadronic collisions”, *JHEP*, 07:001, (2003.)
- [17] CDF Collaboration: “*Neural Network Search for Standard Model Higgs Boson in Met Plus Jets Channel with 1.7 fb<sup>-1</sup>*”, CDF/PUB/EXOTIC/PUBLIC/9166, 2008
- [18] T. Junk, Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties, CDF/DOC/STATISTICS/PUBLIC/8128
- [19] Not only are *b*-quarks more often produced in pairs, but our selections to isolate a large sample of QCD events free from contamination biases the pairs.
- [20] Gluon-jets have a central energy scale that is  $2\sigma$  lower than that of quark-jets.