



CDF/PUB/EXOTIC/PUBLIC/10792

Version 1.0

March 6, 2012

A search for the Higgs Boson in the All Hadronic Channel using 9.45 fb^{-1} of CDF data

Ankush Mitra, Yen-Chu Chen, Song-Ming Wang ¹
Academia Sinica, Taiwan

Francesco Devoto ²
Helsinki University, Finland

Abstract

We report the results of a search for the standard model Higgs boson, in which either the Higgs boson is produced together with a W or Z boson (associated production, $p\bar{p} \rightarrow W/ZH$), or produced via the fusion of two W or Z bosons (vector boson fusion, $p\bar{p} \rightarrow qqH$). The decay channels considered in this analysis are that the W and Z bosons in the associated production decay hadronically, and the Higgs boson decays into a $b\bar{b}$ pair. The final state signature consists of four or more jets, with at least two b -jets. The search is performed on 9.45 fb^{-1} of data recorded at CDF. The data are in agreement with the background model and we set 95% confidence limits on the Higgs production as function of the Higgs mass.

The CDF Collaboration
URL <http://www-cdf.fnal.gov>

Preliminary Results for Winter 2012 Conferences

¹mitra@fnal.gov, chenyc@fnal.gov, smwang@fnal.gov

²francesco.devoto@helsinki.fi

Contents

1	Introduction	3
2	Data Sample, Event Selections and Backgrounds	4
2.1	Initial PreTag Selection	4
2.2	Signal Regions for the Analysis	5
2.3	Background and Signal	5
2.4	Expected Signal and Backgrounds	5
3	bjet energy Neural Network correction	7
4	Jet Width	9
5	QCD Multi-jet Background Prediction	11
5.1	Tuning the Modeling of the Mass $M(qq)$	12
6	Untagged Jets Neural Network	14
7	Neural Network Training	16
7.1	Neural Network Output Tuning	23
8	Systematics	24
9	Unblinded Signal Region	26
10	Results	28
11	Conclusions	32
12	Appendix	33
12.1	Z+jets Generator Level Filter	33

1 Introduction

The Higgs boson plays a central role in the standard model (SM) as it endows particles with mass. Most Higgs searches at CDF focus on final state combinations of leptons, jets and missing energy which help to reduce the background. However the Higgs signal yield is also small. This analysis considers the all-hadronic final state which has the largest signal yield. But the challenge is to reduce and accurately model the overwhelming QCD multijet background.

We search for a Higgs boson decaying to a pair of bottom-quark jets ($b\bar{b}$) accompanied by two additional quark jets (qq') for Higgs masses $100 \leq m_H \leq 150 \text{ GeV}/c^2$. The search is most sensitive to a Higgs boson with mass $< 135 \text{ GeV}/c^2$ where the Higgs boson decay to $b\bar{b}$ is dominant [1]. The two production channels studied in this search are the associated production and vector boson fusion (VBF). The associated production, $p\bar{p} \rightarrow VH \rightarrow qq' + b\bar{b}$, where V is a W/Z vector boson which decays to a pair of quarks. The VBF channel, $p\bar{p} \rightarrow qq'H \rightarrow qq' + b\bar{b}$, where incoming partons each radiate a V and fuse to form a Higgs boson. Both production processes are illustrated in Fig. 1.

This note describes the third iteration of the all-hadronic Higgs search using 9.45 fb^{-1} of $p\bar{p}$ collision data at $\sqrt{s} = 1.96 \text{ TeV}$ recorded by the CDF detector. Articles on the previous 2 fb^{-1} and 4 fb^{-1} search can be found at [2] [3] and references therein. The improvements, with respect to the previous version are: all CDF Run-II data is analyzed (9.45 fb^{-1}), the b -jet resolution is improved, the acceptance for Higgs bosons has increased, the QCD multijet background modelling has improved and an improved neural network (NN) to separate QCD from Higgs.

CDF is a general-purpose detector that is described in detail in reference [4]. The components relevant to this analysis are briefly described here. The charged-particle tracking system is closest to the beam pipe, and consists of a multi-layer silicon detectors (SVX) [5] and a large open-cell drift chamber covering the pseudorapidity region $|\eta| < 1$ [6]. The silicon detectors allow a precise measurement of a track's impact parameter with respect to the primary vertex in the plane transverse to the beam direction. The CDF central tracking region covers the pseudorapidity region $|\eta| \leq 1$ and is used to reconstruct charged particle momenta and the collision vertex. The tracking system is enclosed in a superconducting solenoid, which in turn is surrounded by a calorimeter. The CDF calorimeter system is organized into electromagnetic and hadronic sections segmented in a projective tower geometry, and covers the region $|\eta| < 3.6$. The electromagnetic calorimeters utilize a lead-scintillator sampling technology, whereas the hadron calorimeters use iron-scintillator technology. The central muon-detection system is located outside of the calorimeter and covers the range $|\eta| < 1$.

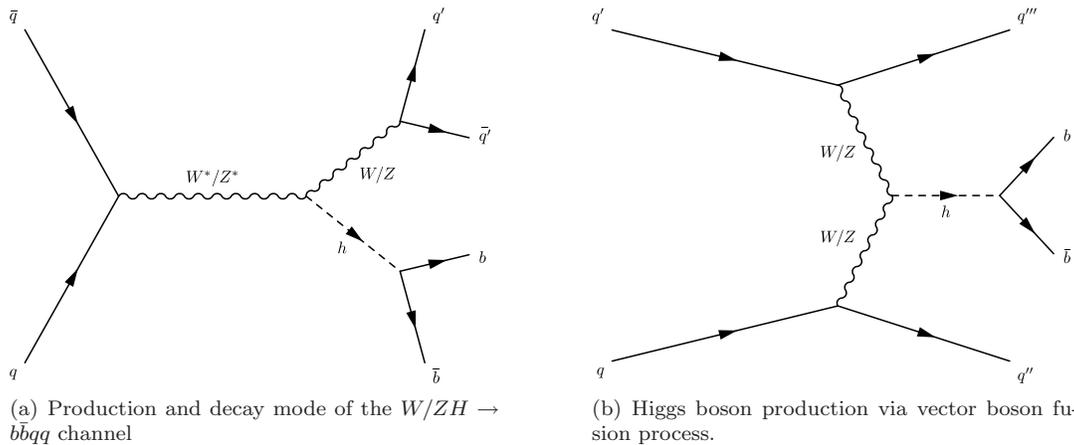


Figure 1: Feynman diagrams for the two Higgs production channels studied in this analysis: Associated Vector Boson Production & Vector Boson Fusion.

2 Data Sample, Event Selections and Backgrounds

The data for this analysis were collected by CDF's multi-jet triggers. They were designed to select events with multiple high E_T jets with large Sum-Et, which is the characteristic of an all-hadronic Higgs event. The first 3.0 fb^{-1} of CDF data were collected by a multi-jet trigger that requires at least four jets with $E_T > 15 \text{ GeV}$ and Sum-Et $> 175 \text{ GeV}$. The remaining 6.45 fb^{-1} data was collected by another trigger that requires at least three jets with $E_T > 20 \text{ GeV}$ and Sum-Et $> 130 \text{ GeV}$. The latter trigger increase the acceptance for low mass Higgs boson events.

2.1 Initial PreTag Selection

The events selected for the analysis have to pass the selection criteria defined below.

- The event must pass the multijet trigger
- The detector components essential to the analysis are operating and producing good data during the data taking period.
- The event's reconstructed primary vertex to be in the luminous region ($|V_Z| < 60 \text{ cm}$, where V_Z is the Z position of the reconstructed primary vertex along the beam axis.
- The missing transverse energy significance ³ must be less than 6.
- The event should have at least four or five jets with $E_T > 15 \text{ GeV}$ and within the pseudorapidity region $|\eta| < 2.4$. Jets are reconstructed from the calorimeter towers using a cone algorithm with fixed radius $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ in $\eta-\phi$ space [7]. The jet E_T measurements are corrected for detector effects [8].
- The jets passing the event selection are ordered by descending b -jet corrected ⁴ E_T and any fifth jet is no longer used.
- The scalar E_T sum of the four leading jets (SumEt) $> 220 \text{ GeV}/c^2$

We search for b jets in events passing the initial PreTag selection. A b jet is identified by its displaced vertex as defined by the SECVTX algorithm [9], or by using the probability that the tracks within the jet are inconsistent with originating from the primary $p\bar{p}$ collision as defined by the JETPROB algorithm [10]. Jets which are tagged by both algorithms, SECVTX takes precedence as it has a lower mistag rate.

The signal/background ratio is enhanced by dividing the data into two non-overlapping b -tagging categories:

- SS: Exactly two jets are tagged by SECVTX
- SJ: Exactly one jet is tagged by SECVTX and exactly one jet is tagged by JETPROB.

The jets from the events which satisfy either b -tagging category are labelled as b_1, b_2, q_1, q_2 where $b(q)$ is the tagged(untagged) jet and $E_T^{b_1/q_1} > E_T^{b_2/q_2}$.

Events with exactly one SECVTX tagged jet are used in the QCD background prediction model:

- events with exactly 1 SECVTX tagged jet are used to predict QCD background for SS events
- events with exactly 1 SECVTX tagged jet and zero JETPROB tagged jets are used to predict the QCD background for SJ events.

Further details on the multijet QCD background are available in section 5.

Events with zero b -tagged tagged or at least 3 b -tagged jets are not used in the analysis and are discarded.

³MET significance = MET/ $\sqrt{\text{Total Transverse Energy}}$

⁴ b -jet energy correction is described in section 3

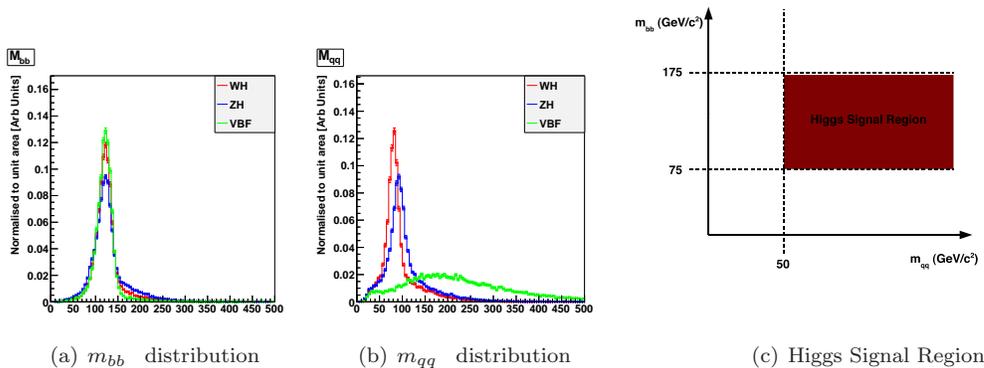


Figure 2: m_{bb} and m_{qq} distributions for $m_H = 120 \text{ GeV}/c^2$. These distributions are used to define the Higgs signal region.

2.2 Signal Regions for the Analysis

The signal region is defined by the mass of the two b -tagged jets (m_{bb}) and two untagged jets (m_{qq}). The range for m_{bb} is defined by the mass of the Higgs bosons we are searching for. The m_{qq} range has a resonance from the W/Z vector decay from the associated production mode. The VBF mode has no m_{qq} resonance. However the q -jets from VBF tend to be produced with large separation in η which gives an effective large broad m_{qq} mass (Fig. 2). The Higgs signal range is defined as:

- $75 \text{ GeV}/c^2 < m_{bb} < 175 \text{ GeV}/c^2$
- $m_{qq} > 50 \text{ GeV}/c^2$

2.3 Background and Signal

The backgrounds that have a similar final state signature as the all hadronic Higgs signal are QCD multi-jet production, top quark pair production, single top quark production, $W \rightarrow qq$ plus $b\bar{b}/c\bar{c}$ jets production, $Z \rightarrow b\bar{b}/c\bar{c}$ plus jets production, and diboson productions (WW , WZ , ZZ). 98% of the total background are from QCD multi-jet production, making the dominant background source. Its contribution is estimated from a data driven based technique, which is described in section 5. The non-QCD backgrounds are estimated from Monte-Carlo (MC) simulation. The W plus heavy-flavor jets production is simulated with the ALPGEN [11] generator to simulate the W boson plus parton production, with PYTHIA [12] used to model parton showers. The other non-QCD backgrounds and the SM Higgs signal samples were simulated with the PYTHIA generator. All the non-QCD background samples are generated using the CTEQ5L PDFs [13].

2.4 Expected Signal and Backgrounds

The number of signal and background events which are used in the analysis, after passing the trigger and event selection, are given in tables 1 and 2.

Table 1: *CDF Run II Preliminary* $9.45fb^{-1}$ Expected number of signal events which pass the trigger, event selection have two b-tags (SecVtx-SecVtx (SS) or SecVtx-JetProb (SJ)).

Signal Region	WH		ZH		VBF		Total	
Higgs Mass (GeV/c^2)	SS	SJ	SS	SJ	SS	SJ	SS	SJ
100	17.8	6.4	13.8	4.4	10.2	3.4	41.8	14.2
105	16.7	6.1	12.9	4.1	9.9	3.5	39.5	13.7
110	15.5	5.6	12.5	4.1	10.0	3.4	38.0	13.1
115	14.3	5.2	11.1	3.6	9.3	3.3	34.7	12.1
120	13.0	4.6	9.9	3.2	8.8	3.0	31.7	10.8
125	10.9	3.8	8.3	2.7	7.8	2.6	27.0	9.1
130	9.0	3.1	6.8	2.2	6.6	2.3	22.4	7.6
135	7.0	2.5	5.3	1.7	5.5	1.8	17.8	6.0
140	5.1	1.8	3.9	1.3	4.2	1.5	13.2	4.6
145	3.5	1.2	2.6	0.9	3.0	1.0	9.1	3.1
150	2.2	0.8	1.7	0.6	2.0	0.7	5.9	2.1

Table 2: *CDF Run II Preliminary*, $9.45fb^{-1}$ Expected number of background and signal (MH120) events which pass the trigger, event selection have two b-tags (SecVtx-SecVtx (SS) or SecVtx-JetProb (SJ)). The number QCD events is defined to be Data - total non-QCD

Backgrounds	Signal Region	
	SS	SJ
$t\bar{t}$	1032.2 ± 155.9	383.9 ± 56.8
Single Top S channel	110.5 ± 18.5	37.8 ± 6.2
Single Top T channel	44.0 ± 7.3	25.5 ± 4.2
$W+b\bar{b}$	77.0 ± 40.0	28.5 ± 14.8
$W+c\bar{c}$	8.3 ± 4.3	7.4 ± 3.6
$Z(\rightarrow b\bar{b}/c\bar{c})+jets$	872.6 ± 452.0	337.6 ± 174.5
WW	5.6 ± 0.8	5.6 ± 0.8
WZ	19.7 ± 2.9	7.9 ± 1.1
ZZ	21.4 ± 3.1	7.9 ± 1.1
total non-QCD	2191.3 ± 480.3	842.1 ± 184.3
Data	87272	46818
QCD*	85080.7	45975.9
Higgs signal (120GeV)	31.7 ± 4.7	10.8 ± 1.6

3 bjet energy Neural Network correction

The di-jet invariant mass has a straight correlation to jet energy measurement and to optimize the di-jet mass we need to improve the resolution of jet energy measurement. These improvements will help to improve other parts of the analysis, such as the QCD modeling and jet shaping; both depend on the jet- E_T . A Neural Network (NN) is trained to correct the measured b -jet energy to the b -parton energy.

The choice of the variables to train the NN carry information of the jet energy. Two different NN were trained using SecVtxTag and JetProb events and two sets of variables was identified for them. All energies are corrected for detector effects [8]. We had 9 variables for SecVtxTag events:

- the Jet E_t ,
- the Jet P_t ,
- the Jet Raw E_t ,
- the transverse mass⁵,
- the decay length of the jet in 2-dim and its uncertainty,
- the P_t of secondary vertex,
- the maximum P_t of the track inside the jet cone,
- the sum of all tracks inside the jet cone,

and 6 variables for JetProb events:

- the Jet E_t ,
- the Jet P_t ,
- the Jet Raw E_t ,
- the transverse mass,
- the decay length of the jet in 2-dim and its uncertainty.

The NN was trained with b tagged jets matched to b -partons. The match criteria is defined as ΔR between the b -jet and the b -parton is ≤ 0.4 .

The NN correction function is trained to estimate the ratio between the energy of b -parton and the b -jet. We started to train, for each sample (VBF, WH and ZH), a dedicated NN. After the training we compared the resolution for each mass before and after the NN correction.

For each mass point, the resolution is define as the ratio between the RMS and the mean value of its own M_{bb} distribution; the values of these two parameter are calculated with a gaussian fit in a range of 2σ around the mean value. This operation is iterated ten times.

To identify the best NN performance, we applied to each sample the others NN. We obtained the best performance with the VBF NN, so we decided to use that one for the analysis (fig. 3).

Figure 4 shows the comparison of mass distribution before and after NN correction; we have an improvement of value of M_{bb} , the mean value shift to $\sim 124 \text{ GeV}/c^2$, the RMS value reduces by 5 – 6% and the resolution increase by $\sim 12 - 13\%$.

⁵The jet transverse mass is define as $P_t/P * M$.

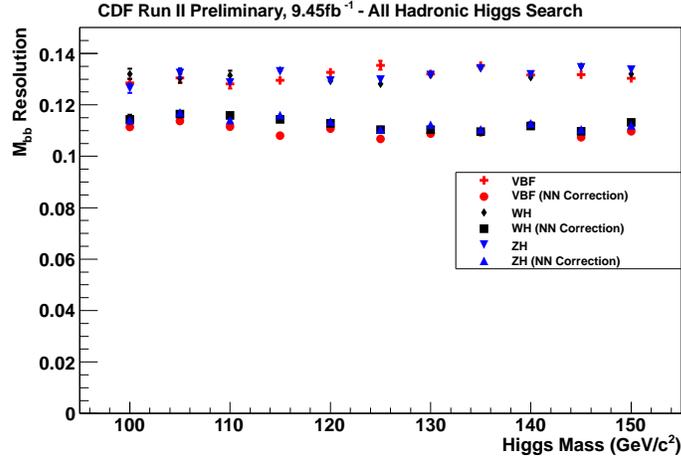


Figure 3: The resolution for each mass point before and after NN correction samples with VBF NN train.

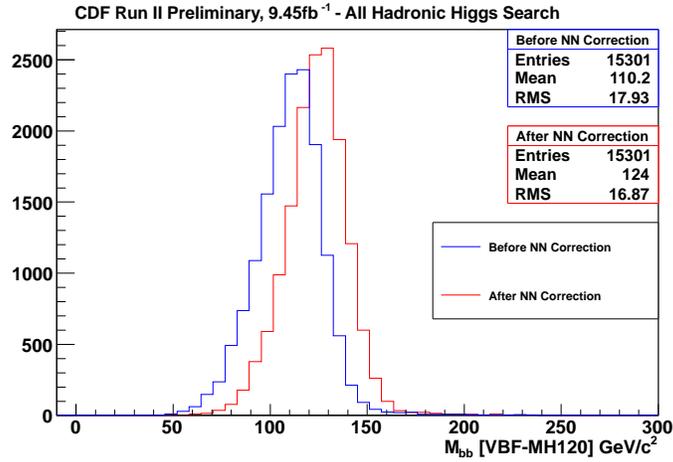


Figure 4: Comparison of M_{bb} distribution before and after NN Correction for VBF sample with $M_H = 120\text{GeV}/c^2$.

4 Jet Width

The untagged q -jets from QCD are a mixture of quark and gluon jets whereas the q jets from the Higgs signal are quark jets. On average, gluon jets tend to be broader than quark jets. Thus the width of a jet can be used to discriminate quarks from gluons and so improve the separation of the QCD multijet background from the Higgs signal. The definitions of the jet width measured by the calorimeter ($\langle R \rangle_{CAL}$) and tracker ($\langle R \rangle_{TRK}$) are given in equation 1.

$$\langle R \rangle_{CAL} = \sqrt{\sum_{\text{towers}} \left[\frac{E_t^{\text{tower}}}{E_t^{\text{jet}}} \left(\Delta R(\text{tower}, \text{jet}) \right)^2 \right]} \quad (1a)$$

$$\langle R \rangle_{TRK} = \sqrt{\sum_{\text{tracks}} \left[\frac{P_t^{\text{track}}}{P_t^{\text{jet}}} \left(\Delta R(\text{track}, \text{jet}) \right)^2 \right]} \quad (1b)$$

where

$$\Delta R(\text{tower}, \text{jet}) \equiv \sqrt{(\text{tower } \eta - \text{jet } \eta)^2 + \Delta\phi(\text{tower } \phi, \text{jet } \phi)^2} \quad (2a)$$

$$\Delta R(\text{track}, \text{jet}) \equiv \sqrt{(\text{track } \eta - \text{jet } \eta)^2 + \Delta\phi(\text{track } \phi, \text{jet } \phi)^2} \quad (2b)$$

and $\Delta\phi(\text{tower}/\text{track } \phi, \text{jet } \phi)$ is smallest angular difference between tower/track ϕ and jet ϕ . All calorimeter towers inside the jet cone of $\Delta R < 0.4$ are used in the $\langle R \rangle_{CAL}$ calculation. All tracks within the jet cone of $\Delta R < 0.4$ are selected for $\langle R \rangle_{TRK}$

The jet-width depends not only on the parton initiating the jet but also varies with jet kinematics and detector effects. A neural network is trained to parameterise the jet width as a function of these jet kinematic and detector variables. The calorimeter jet-width is parameterised by two neural network functions: one for the central calorimeter ($|\eta| \leq 1.1$) and another for the forward calorimeter ($|\eta| > 1.0$). Each calorimeter neural network function parameterises the jet width as a function of jet- E_T , jet- η and number of reconstructed vertices (NVtx). The tracker measured jet width is parameterised by one neural network as a function of jet- E_T . A separate neural network is trained for data and MC.

The neural network is trained on a sample of light-flavored (untagged) jets from MC and data using jets from the hadronic W boson decay from $t\bar{t} \rightarrow b\bar{b}l\nu qq$. These events are used as one can extract a pure sample of $t\bar{t} \rightarrow b\bar{b}l\nu qq$ from data and the leading untagged jet pair whose mass is $M_W \pm 30 \text{ GeV}/c^2$ are assumed to be jets from a W boson decay and so are light flavored.

The neural network parameterisations of the jet-width are used to remove the jet-kinematic and detector dependencies by rescaling the measured jet-width to a common reference of jet- $E_T = 50 \text{ GeV}/c^2$, jet- $\eta=0$, and NVtx=1. Equation 3 demonstrates how the jet-widths are rescaled.

$$\langle R \rangle_{CAL}^{Data'} = \langle R \rangle_{CAL}^{Data} \times \frac{\langle R \rangle_{CAL}^{Ref}}{f_{CAL}^{Data}(\text{jet- } E_T, \text{jet-}\eta, \text{NVtx})} \quad (3a)$$

$$\langle R \rangle_{CAL}^{MC'} = \langle R \rangle_{CAL}^{MC} \times \frac{\langle R \rangle_{CAL}^{Ref}}{f_{CAL}^{MC}(\text{jet- } E_T, \text{jet-}\eta, \text{NVtx})} \quad (3b)$$

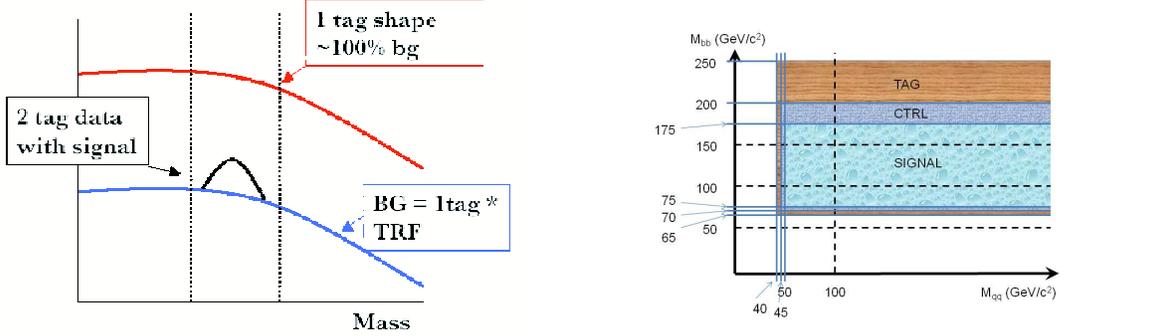
$$\langle R \rangle_{TRK}^{Data'} = \langle R \rangle_{TRK}^{Data} \times \frac{\langle R \rangle_{TRK}^{Ref}}{f_{TRK}^{Data}(\text{jet- } E_T)} \quad (3c)$$

$$\langle R \rangle_{TRK}^{MC'} = \langle R \rangle_{TRK}^{MC} \times \frac{\langle R \rangle_{TRK}^{Ref}}{f_{TRK}^{MC}(\text{jet- } E_T)} \quad (3d)$$

$\langle R \rangle_{CAL}^{Ref}$ and $\langle R \rangle_{TRK}^{Ref}$ are the data common reference jet-width for the calorimeter and tracker measured jet-width, respectively. The MC uses the same reference values as this ensures the MC measures the same jet-width as data. The function $f_{CAL}^{Data/MC}(\text{jet- } E_T, \text{jet-}\eta, \text{NVtx})$ is the data/MC jet width

parameterization and $f_{TRK}^{MC}(\text{jet- } E_T)$ is the NN data/MC jet width parameterization. After rescaling the measured jet-widths, any difference in the jet-width can now be assumed to be due to the type of parton (quark or gluon) initiating the jet.

After these described corrections are applied, the MC has a systematic uncertainty of $\pm 2.6\%$ (5.5%) for the calorimeter(tracker) measured jet-width.



(a) TRF Principal: The red-line corresponds to the 1-tag distribution which is $\approx 100\%$ background. The 2-tag background is assumed to be a scaled version of the 1-tag (background) distribution (blue line). The TRF is derived from the regions outside the signal peak.

(b) M_{bb} - M_{qq} plane to define the TAG (brown) and CTRL (blue) regions.

Figure 5: The Tag-Rate function and the $m_{bb} - m_{qq}$ plane

5 QCD Multi-jet Background Prediction

The critical component to this analysis is an accurate prediction of the QCD background as it is the dominant background. In this analysis a data driven model was devised to predict the two-tagged background from the background-rich one-tagged data. The assumption is that the two-tagged background distribution is a scaled replica of the one-tagged distribution (figure 5). The scale factor which deduced from the one-tag data is a multi-dimensional function, called the Tag Rate Function (TRF). The TRF is the probability of a jet being b-tagged in the event that already has one other jet tagged as a b-jet. The probability is measured in a kinematic region that has very little contribution from the Higgs signal, the TAG region. This is applied to the one tagged events in the signal region to predict the double b-tagged QCD background. As a systematic, another TRF is derived from the control (CTRL) region. This is also applied to the 1-Tag data in the signal region to give an alternative background prediction. The difference of these two background predictions is applied as a systematic error.

The key issue of this method is to make sure that the technique can correctly predict the shapes of the kinematic distributions of the double b-tagged QCD multi-jet events which will be used later in the NN training to separate the Higgs signal from the QCD background. This TRF method does not necessary predict the right normalization of the double b-tagged QCD background.

The TRF is parameterised as a function of three parameters which are:

- ΔR of Tag-Probe jet pair
- E_T of the Probe jet
- pseudorapidity of the Probe jet ($|\eta|$)

$$TRF(\Delta\eta_{bb}, E_t, |\eta|) = \frac{\text{Number of 2-Tag events}(\Delta\eta_{bb}, E_t, |\eta|)}{\text{Number of events with } \geq 1 \text{ tight SecVtx tagged jet}(\Delta\eta_{bb}, E_t, |\eta|)} \quad (4)$$

The TRFs are measured separately for SS and SJ double b-tagged categories. For the SJ category, events with only one tight SecVtx tagged jets are considered in the measurement of the TRF(SJ).

The data used to derive the TRFs come from examining data outside the signal region (figure 5) in the $M(bb)$ - $M(qq)$ plane. The signal region is defined by $75 < M(bb) < 175 \text{ GeV}/c^2$ and $M(qq) > 50 \text{ GeV}/c^2$ mass window. Two regions outside the signal region are defined:

- CTRL region: an open-surrounding area around the signal region; $70 < M(bb) < 200 \text{ GeV}/c^2$ and $M(qq) > 45 \text{ GeV}/c^2$, except the signal region.
- TAG region: an open-surrounding area around the CTRL region; $65 < M(bb) < 250 \text{ GeV}/c^2$ and $M(qq) > 40 \text{ GeV}/c^2$, except the CTRL region and signal region.

The default TRF uses data from the TAG region. The CTRL region is used to derive systematic errors (figure 5). In the TRF deriving the contribution of $t\bar{t}$ and Z plus jets are subtracted.

If a variable shows power in discriminating the background but is not well modeled by the TRF, it is not used in the analysis. Variable like $M(qq)$, although not very well described by the raw TRF, with fine tuning on the TRF based on correction function obtained in the TAG region it can be used eventually. This is discussed in the next section.

5.1 Tuning the Modeling of the Mass $M(qq)$

The TRF generated via this method does predict well the shapes of various kinematic variables except a few. Two variables are not very well described but important to this analysis are $M(bb)$ and $M(qq)$. Possible reasons for the miss matching are the following. The ratio of 2 b tagged over 1 b tagged events is assumed to be flat. This assumption is true in large scale but not quite so in local area in the $M(bb)$, $M(qq)$ phase space. The other reason is that while developing the TRF both 1 b tagged and 2 b tagged events contribute to the denominator. While in predicting 2 b events in the signal region only 1 b tagged events are used. This creates slightly inconsistency in the composition of the denominator. These are limitations of the method. We can only choose the variables that well described to proceed with the analysis so we need to correct for $M(qq)$ and hopefully the $M(bb)$ shape will follow.

To correct for $M(qq)$, we measure the correction function by applying the TRF to the one tagged events in the TAG region and measure the ratio of the predicted double b-tagged events to the observed b-tagged events as a function of the mass $M(qq)$. The correction function is then applied in the signal region when we are predicting the double b-tagged events in the signal region. As a systematic check, we also measured another correction function for $M(qq)$ in the CONTROL region and apply the correction function in predicting the 2 b-tagged events, in the signal region, while using the TRF from the CONTROL region. The difference in the predictions between using the TRF's (with their corresponding correction function) measured from the TAG region and from the CONTROL region is part of the source of systematic uncertainty of the modeling.

After the fine tuning based on $M(qq)$ the kinematic distributions of the predicted double b-tagged events in the signal region are compared to the observed double b-tagged events. Figure 6 shows the distributions of $M(bb)$, $M(qq)$ and $\cos\theta_3$ as examples from SS channel. The corresponding distributions from SJ channel are shown in Figure 7. Not all variables can be well described as shown.

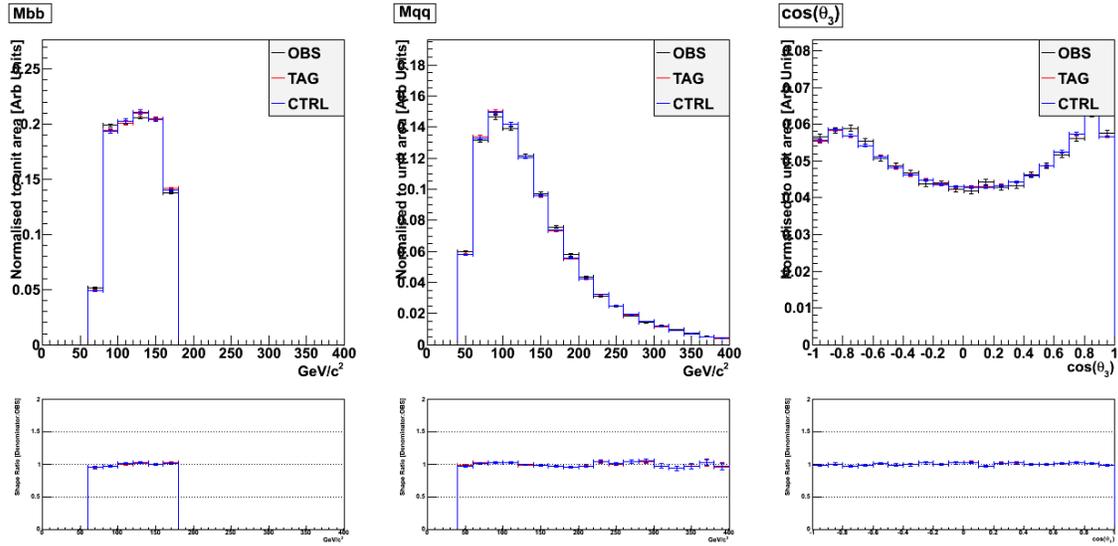


Figure 6: The kinematic distributions of the predicted double b-tagged events in the signal region are compared to the observed double b-tagged events for the SS category. The left side plot shows the invariant mass of the two b tagged jets system, $M(bb)$. The middle plot shows the invariant mass of the two not b tagged jets system, $M(qq)$. The right side plot shows the $\cos\theta_3$. The red hashed histograms are the predicted double b-tagged events, and the black points are the observed double b-tagged events.

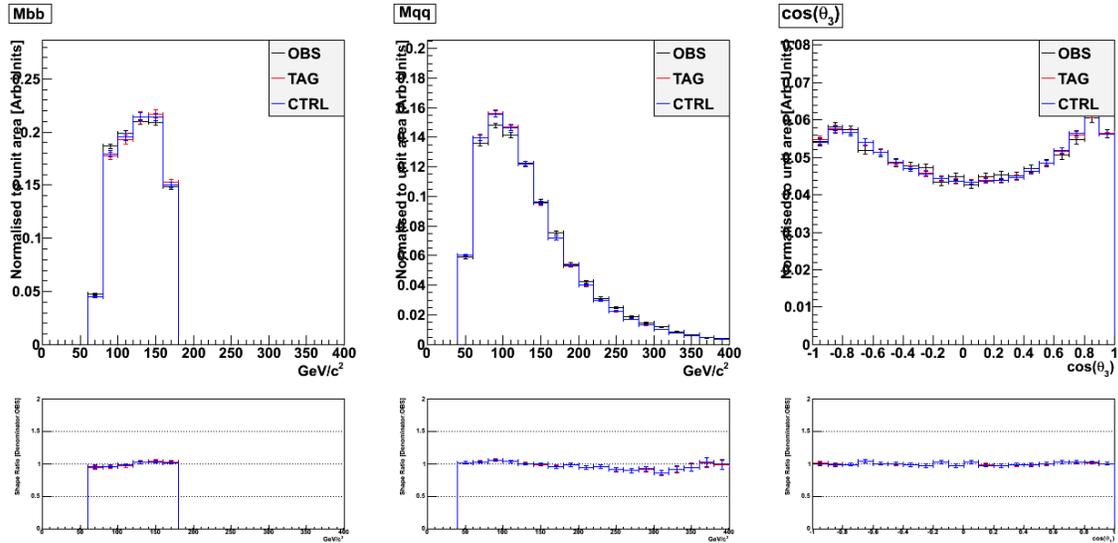
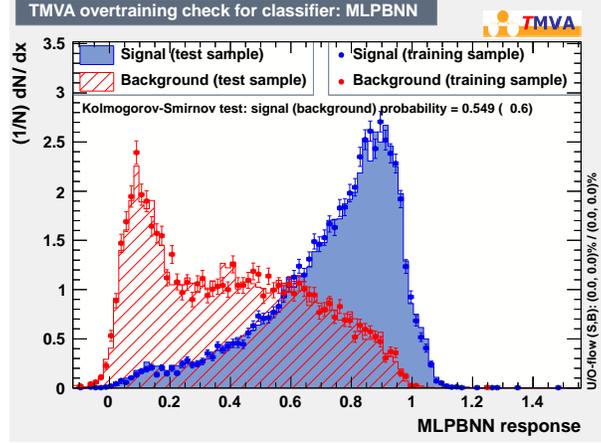
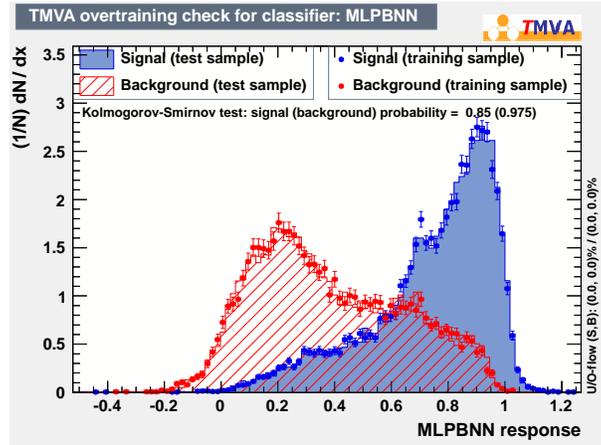


Figure 7: The kinematic distributions of the predicted double b-tagged events in the signal region are compared to the observed double b-tagged events for the SJ category. The left side plot shows the invariant mass of the two b tagged jets system, $M(bb)$. The middle plot shows the invariant mass of the two not b tagged jets system, $M(qq)$. The right side plot shows the $\cos\theta_3$. The red hashed histograms are the predicted double b-tagged events, and the black points are the observed double b-tagged events.



(a) ZHQCD



(b) WHQCD

Figure 8: The performance of YC.ZH (shown in the plot (a)) and YC.WH (shown in the plot (b)).

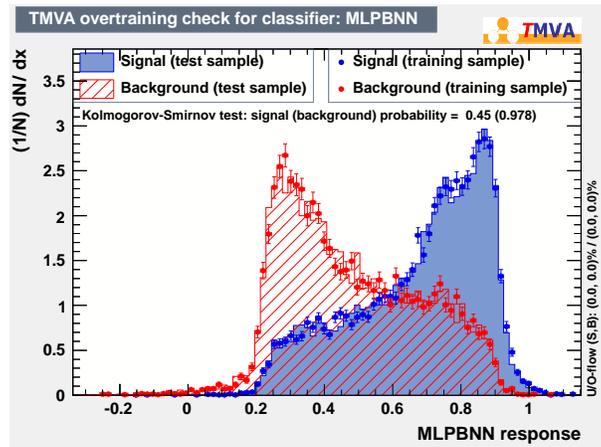
6 Untagged Jets Neural Network

In the association production of Higgs particle there are light quark pairs from Z, W decay or through vector boson fusion, VBF. These light quark pairs have special features that can be used to help discriminate the QCD background thus help the search of Higgs particle. So we train special Neural Networks to identify light quark pairs in the association production of Higgs particle event. Below we describe the features of the light quark pairs and how the Neural Networks are trained.

The variables used are $M(qq)$, $d\phi$ and $d\eta$, $dR(qq)$, Ptq , where $M(qq)$ is the invariant mass of the two light quarks; $d\phi$ and $d\eta$ are the differences of the two light quarks in ϕ and η ; $dR(qq)$ is defined as $\sqrt{d\phi^2 + d\eta^2}$; finally Ptq is the transverse momentum of one q jet in the qq system.

To train the Neural Networks to identify qq in ZH, WH and VBF+H, special attention is paid to reject events that could have three or more b quarks. In case of ZH, this particularly important because when both Z and H decay into b quark pair and when only two jets are b tagged, it is hard to say the two untagged jets are from truly from Z.

The performance of the three Neural Networks are shown in Figure 8 and in Figure 9.



(a) VBFQCD

Figure 9: The performance of YC_VBF.

7 Neural Network Training

For the Higgs analysis, a multivariate discriminant has the ability to combine the information from several variables. This improves the ability to separate a Higgs signal from background events far greater than a standard cuts analysis. The TMVA package [14] allows one to evaluate several multivariate classifiers. For this analysis, we considered an Artificial Neural Network ⁶.

A dedicated Neural Net was trained for each process, WH, ZH and VBF. Because the processes WH and ZH are similar we continue to identify them as VH, although the Neural Net was trained separately. The output of this first Neural Net was used as the input of a second one (SuperDiscriminant - SD). We used the output of the last Neural Net for the analysis.

As the background is dominated by QCD, the 1-Tag background, weighted by the TRF, is used as the background sample for the NN.

The Neural Net was trained at three target Higgs masses; 100 GeV, 120 GeV and 140 GeV. These three trained neural nets were used to search for a Higgs boson between 100 GeV to 150 GeV. For each mass point, the closest trained neural net was used as follows:

- 100 GeV Higgs used Neural Net trained on 100 GeV Higgs sample
- 105,110,115,120,125,130 GeV Higgs used Neural Net trained on 120 GeV Higgs sample
- 135,140,145,150 GeV Higgs used Neural Net trained on 140 GeV Higgs sample

For all Neural Nets the same window cut was applied:

$$\begin{aligned} \text{Signal Window} & : 75 < M(b\bar{b}) < 175\text{GeV} \\ & M(q\bar{q}) > 50\text{GeV} \end{aligned} \tag{5}$$

After experimenting with combinations of variables for the training, we decided on the following set of variables for the VH training:

- Mass of the two b-tagged jets ($M(bb)$)
- Mass of the two non b-tagged jets ($M(qq)$)
- cosine of the leading-jet scattering angle in the 4-jet rest-frame ($\cos(\theta_3)$)
- χ variable [3]
- Jet Width Tower of leading non b-jet
- Jet Width Track of leading non b-jet
- Jet Width Tower of second leading non b-jet
- Jet Width Track of second leading non b-jet
- Aplanarity, Sphericity and Centrality
- ΔR of the two b-tagged jets
- ΔR of the two non b-tagged jets
- Difference between the two ϕ angles of the two b-tagged jets
- Difference between the two ϕ angles of the two non b-tagged jets

⁶We followed TMVAs recommendation of the Multi-layer Perceptron algorithm for the artificial neural network

- Neural Network distribution for the identification of WH non b-jet
- Neural Network distribution for the identification of ZH non b-jet

and the following set of variables for VBF training:

- Mass of the two b-tagged jets ($M(bb)$)
- Mass of the two non b-tagged jets ($M(qq)$)
- χ variable [3]
- Jet Width Tower of leading non b-jet
- Jet Width Track of leading non b-jet
- Jet Width Tower of second leading non b-jet
- Jet Width Track of second leading non b-jet
- η of leading non b-jet
- η of second leading non b-jet
- Difference between the two η angles of the two b-tagged jets
- Invariant mass of four jets system
- Sum of momentum Z component of the four jets system
- Sphericity and Centrality
- ΔR of the two b-tagged jets
- ΔR of the two non b-tagged jets
- Difference between the two ϕ angles of the two b-tagged jets
- Neural Network distribution for the identification of VBF non b-jet

The definition for $\cos(\theta_3)$ can be found in [15]

The χ variable is defined as [3]:

$$\chi = \text{Min}(\chi_W, \chi_Z) \quad (6)$$

$$\chi_{W/Z} = \sqrt{(M_{W/Z} - M_{qq})^2 + (M_H - M_{bb})^2} \quad (7)$$

For the SuperDiscriminant (SD) neural net training the list of training variables is:

- Neural Net distribution for WH
- Neural Net distribution for ZH
- Neural Net distribution for VBF

Figures 10-17 show the signal & background plots for the selected variables for the VH and VBF channels and the figure 18

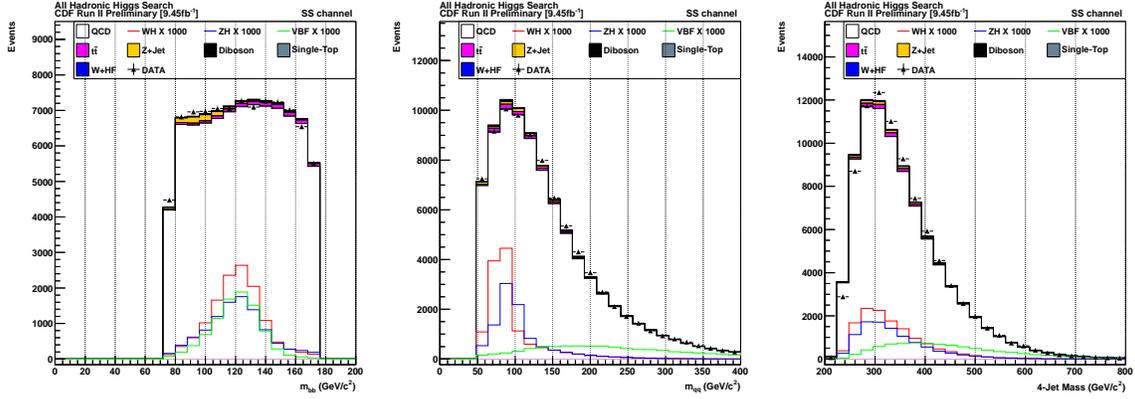


Figure 10: The SS TRF prediction of the training variables. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

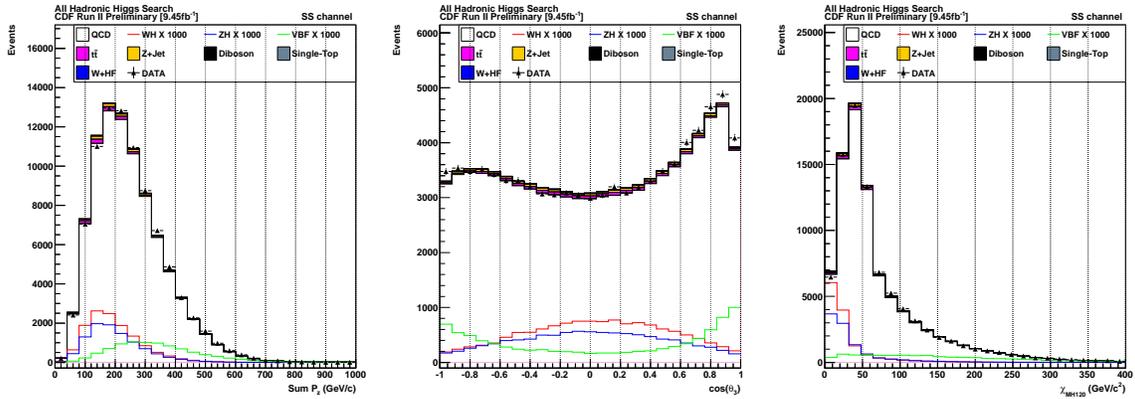


Figure 11: The SS TRF prediction of the training variables. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

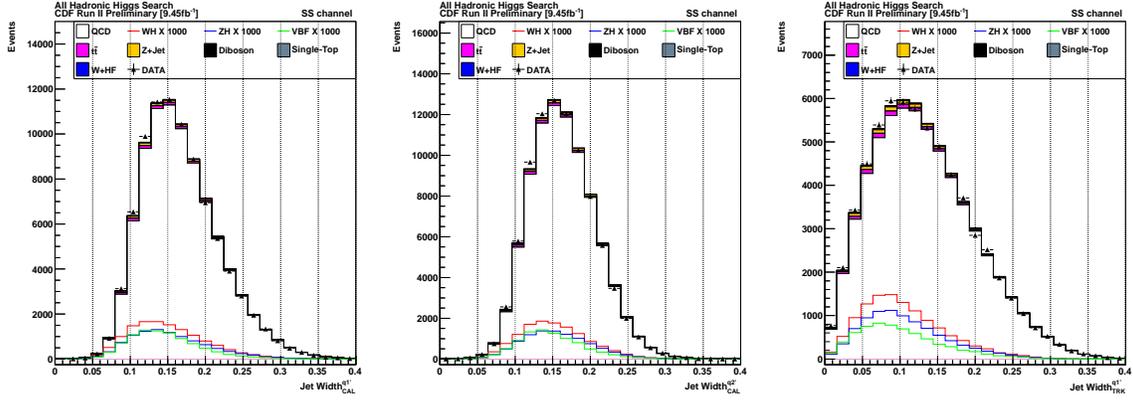


Figure 12: The SS TRF prediction of the training variables. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

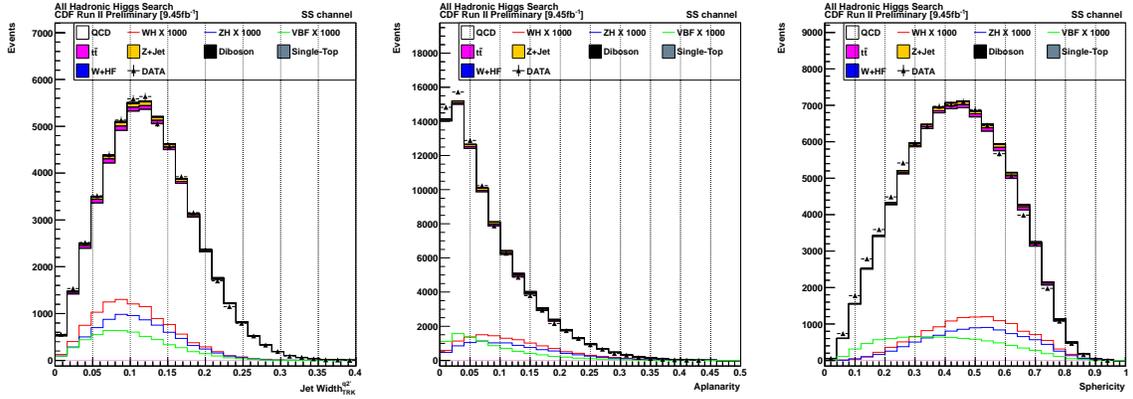


Figure 13: The SS TRF prediction of the training variables. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

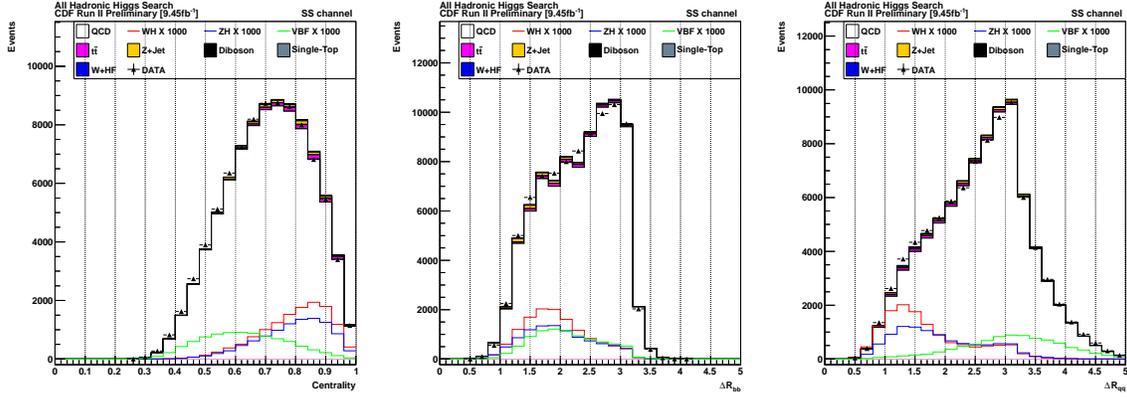


Figure 14: The SS TRF prediction of the training variables. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

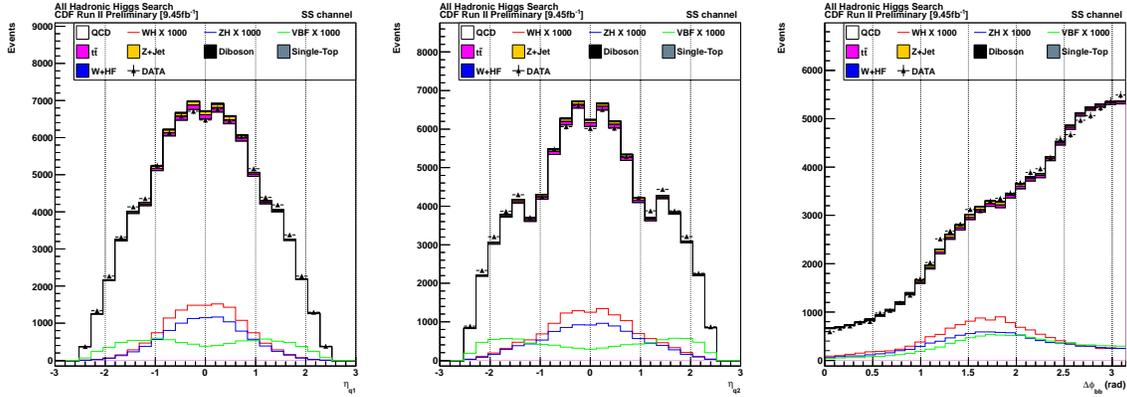


Figure 15: The SS TRF prediction of the training variables. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

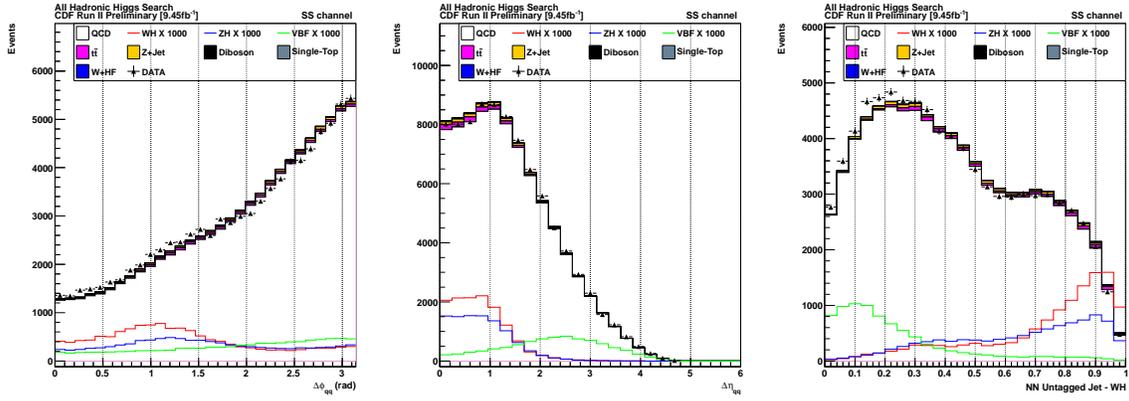


Figure 16: The SS TRF prediction of the training variables. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

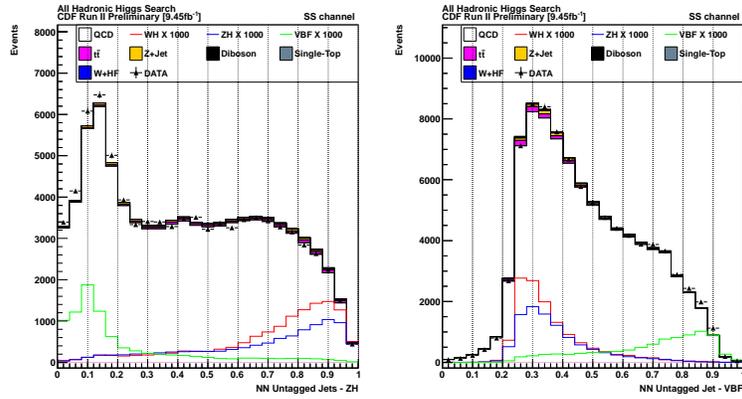


Figure 17: The SS TRF prediction of the training variables. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

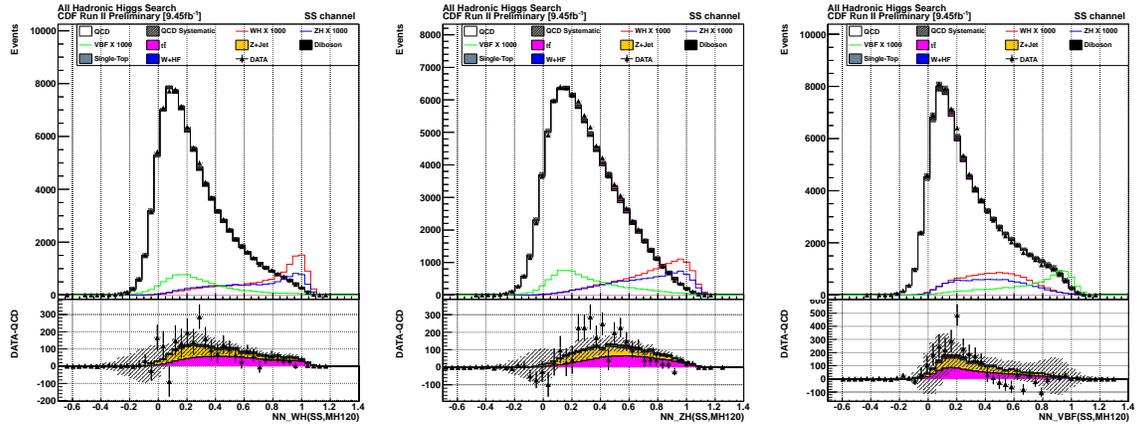


Figure 18: The SS TRF prediction of the training variables for SuperDiscriminat. The black histograms are the TRF prediction and the black triangles are the data for the Higgs signal region.

7.1 Neural Network Output Tuning

The Super Discriminant Neural Network output presents a disagreement between Data and TRF prediction in the Higgs Signal Region. We also see a similar mismodelling for the VBF Neural Network output.

We decided to correct the VBF Neural Network Output (NN_VBF) so to improve also the SD Neural Network Output (NN_SD). The idea is to apply a correction function measured in a control region to the Signal Region.

The two control region we considered are:

- Tag Region
- NJet6 Region

In the two control region is present the same mis-modelling we see in the Signal Region, so we assume that they have the same source. The correction function is the fit of the ratio between Data and TRF Prediction in a control region and we use that fit values to reweight the TRF events, assuming that the same correction works also in Signal Region, we apply it in the Signal Region, if the NN_VBF present an improvement we also should see an improvement in NN_SD.

We calculated the TRF correction function for all three mass point used for training: 100, 120 and 140GeV and for the two bjet categories: SS and SJ in the two control regions. To choose which correction to use, we applied each correction function fit to the TRF predict events in the Control Region (CTRL) and we measured the χ^2/NDF and the Kolmogorov-Smirnov value of corrected TRF prediction to Data. The function with best values is selected and we used the other one to estimate the systematic uncertainty.

After examining all results, we decided to use the Tag Correction Function to correct the Neural Network Output for all samples and the NJet6 one for the systematics uncertainties.

8 Systematics

We consider systematics which affect the shapes and normalisation of the Neural Net distributions for the signal and background. The systematics which are considered are:

- QCD Modelling, which is divided into three categories:
 - TRF interpolation uncertainty
 - m_{qq} tuning Uncertainty
 - VBF-NN tuning uncertainty
- Jet Width
- Jet Energy Scale
- Initial and final state radiation (ISR/FSR).
- Luminosity
- b-tag scale factor
- Parton distribution uncertainties
- cross-section uncertainties

The nominal background prediction uses the TRF derived from the TAG region which is interpolated into the signal region. The systematic error for this interpolation is taken from the background prediction using the TRF derived from the CTRL region. The difference between these two background shapes is used as the interpolation shape systematic.

Section 5.1 discussed the tuning required for TRF prediction of m_{qq} . The tuning used functions derived from the TAG region and CTRL region. The nominal correction used the TAG derived tuning function. As a measure of the systematic error, the CTRL region tuning function is used. The difference in the neural net output shape from using the two tuning functions is used as the m_{qq} tuning uncertainty shape systematic.

The VBF-NN requires additional tuning to correct the TRF prediction, akin to the m_{qq} tuning. The VBF-NN tuning functions are measured in the TAG and the SIXJET control region. The TAG region tuning is used for the nominal QCD background prediction. The SIXJET correction function is used to measure the systematic uncertainty. The difference of the Higgs-NN shape from using the two VBF-NN tuning functions is used as the VBF-NN tuning uncertainty shape systematic.

The Jet Energy Scale affects the shape of the Jet E_t related quantities. The training variables which are affected most are M_{bb} and M_{qq} . Although the Jet Energy Scale systematics affects the M_{bb} and M_{qq} distributions, the neural network shape is less affected. The Jet Shape Systematic also affect the acceptance on account of the $\text{SumEt} > 220.0 \text{ GeV}$ event selection cut, which gives rise to a $\pm 9\%$ rate systematic error in addition to the shape errors.

The jet-width uncertainty is applied by shifting the jet-width value by $\pm 1\sigma$ and propagating this change through the NN which results in a change in shape of the Neural Network.

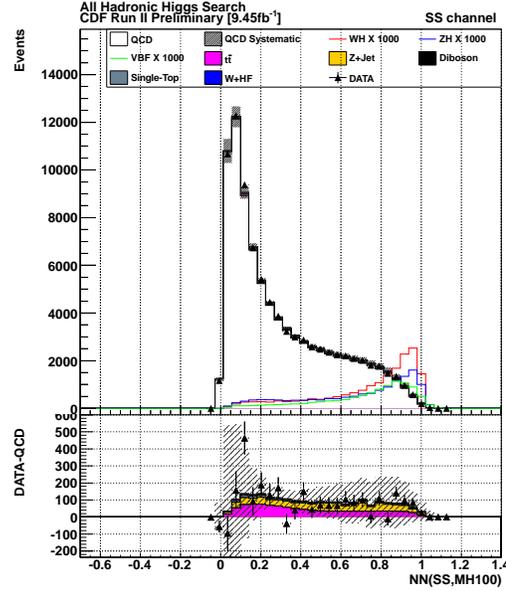
The initial and final state radiation systematics also include a shape systematic as well as a $\pm 3\%$ rate systematic.

Uncertainties on the cross-section for all MC samples are applied: 7% for $t\bar{t}$ and single-top, 6% for diboson, 50% for W +HF and Z +Jets, 5% for WH/ZH and 10% for VBF. Other rate systematics include a $\pm 2\%$ PDF rate systematic and the 6% luminosity uncertainty. For B-Tagging, all MC SecVtx-SecVtx events have to apply a scale factor of 0.902 with a 6.7% rate systematic. For the SecVtx-JetProb events, the MC scale factor is 0.655 with a 6.4% rate systematic. Finally there is a 3.55% rate systematic for all signal samples according to our trigger study.

Table 3 summarises all the rate uncertainties & shape systematics which are applied to calculate the limit.

Table 3: Summary of all Systematic Uncertainties used to calculate the limit

TRF (QCD) Uncertainties	
TRF Interpolation	Shape
TRF m_{qq} Tuning	Shape
TRF VBF-NN tuning	Shape
Higgs and Non-QCD Uncertainties	
Luminosity	$\pm 6\%$ Rate
Trigger	$\pm 3.55\%$ Rate
SecVtx+SecVtx	$\pm 7.1\%$ Rate
SecVtx+JetProb	$\pm 6.4\%$ Rate
Jet Energy Correction	$\pm 9\%$ Rate
	Shape
Jet Width	Shape
Higgs and Non-QCD cross-section uncertainties	
$t\bar{t}$ & single-top	$\pm 7\%$ Rate
Diboson (WW/WZ/ZZ)	$\pm 6\%$ Rate
W+HF & Z+Jets	$\pm 50\%$ Rate
WH/ZH	$\pm 5\%$ Rate
VBF	$\pm 10\%$ Rate
Higgs Uncertainties	
PDF	$\pm 2\%$ Rate
ISR/FSR	$\pm 3\%$
	Shape



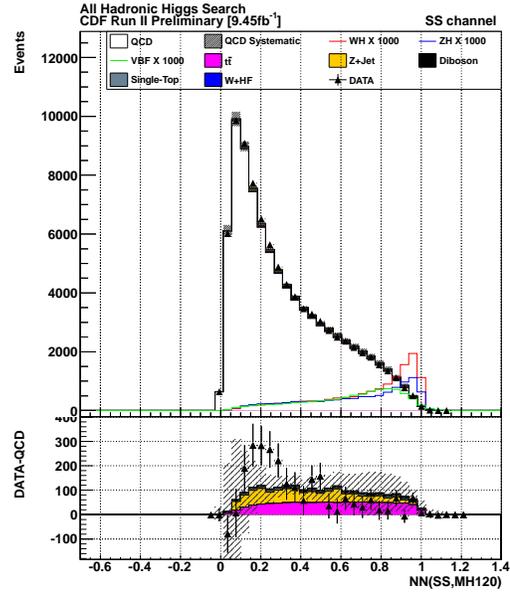
(a) MH100-SS

Figure 19: Higgs-NN distribution for $m_H = 100 \text{ GeV}/c^2$. The plot is from the SS b-tag category. All backgrounds are stacked and the superimposed Higgs signal is scaled by x1000. As the QCD background is large, plots of the DATA-QCD are plotted with stacked plot of non-QCD background and QCD systematic.

9 Unblinded Signal Region

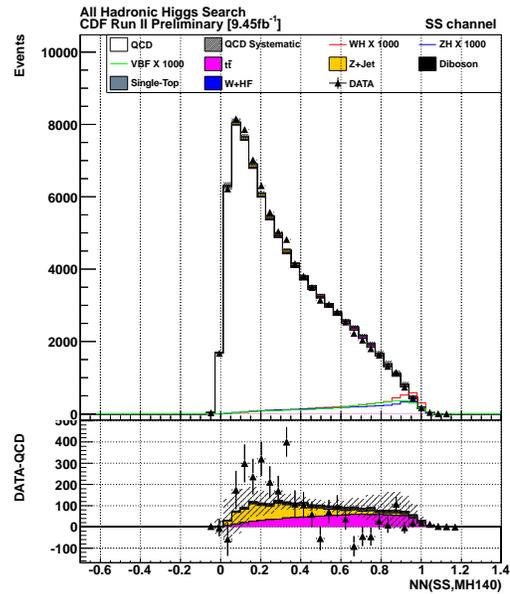
Figures 19-21 show the Higgs-NN output for Higgs boson masses of 100, 120 and 140 GeV/c^2 for the SS b-tag category ⁷. The histograms show the data, a stacked plot of background and the Higgs signal scaled by x1000. Below each histogram is QCD(TRF) subtracted data and background. Higgs bosons of 105, 110, 115, 125, 130 GeV/c^2 use the same NN as Fig. 20 and share the same data and background prediction; only the signal template differs. Similarly, Higgs bosons of 135, 145, 150 GeV/c^2 use the same NN as Fig. 21 and only differ by using a different Higgs signal template. All three NN show no evidence of a Higgs boson signal and no disagreement between the background and observed data

⁷The equivalent plots for the SJ b-tag category can be found at the public web page: http://www-cdf.fnal.gov/physics/new/hdg/Results.files/results/vhqqbb_120307



(a) MH120-SS

Figure 20: Higgs-NN distribution for $m_H = 120 \text{ GeV}/c^2$. The plot is from the SS b-tag category. All backgrounds are stacked and the superimposed Higgs signal is scaled by x1000. As the QCD background is large, plots of the DATA-QCD are plotted with stacked plot of non-QCD background and QCD systematic.



(a) MH140-SS

Figure 21: Higgs-NN distribution for $m_H = 140 \text{ GeV}/c^2$. The plot is from the SS b-tag category. All backgrounds are stacked and the superimposed Higgs signal is scaled by x1000. As the QCD background is large, plots of the DATA-QCD are plotted with stacked plot of non-QCD background and QCD systematic.

Table 4: Expected and observed 95% CL upper limits for the combined SS and SJ channels using 9.45 fb^{-1} of $p\bar{p}$ data collected by the CDF detector. The limits are normalised to the expected Higgs cross-section.

Higgs mass	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Observed
100	1.4	3.6	7.7	14.5	24.4	10.9
105	1.8	3.8	7.5	13.6	22.3	7.5
110	2.0	4.0	7.6	13.2	21.7	7.0
115	2.3	4.4	8.3	14.5	23.4	7.2
120	2.4	4.6	8.9	15.6	25.3	8.4
125	2.8	5.7	11.0	19.5	31.6	9.0
130	3.4	7.1	13.8	24.3	39.5	13.2
135	5.3	10.8	19.5	32.2	49.6	21.2
140	7.3	14.3	25.8	42.7	66.1	26.2
145	10.2	20.4	36.7	60.5	93.4	35.1
150	17.1	32.5	58.7	98.2	152.0	64.6

Table 5: Expected and observed 95% CL upper limits for the SS channel using 9.45 fb^{-1} of $p\bar{p}$ data collected by the CDF detector. The limits are normalised to the expected Higgs cross-section.

Higgs mass	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Observed
100	3.6	6.7	10.9	16.6	23.7	15.9
105	4.3	6.7	10.1	14.9	21.2	8.2
110	4.6	6.7	9.8	14.2	20.1	8.3
115	4.8	7.3	10.8	15.5	21.7	8.9
120	5.3	7.8	11.6	16.8	23.6	10.2
125	6.4	9.5	14.0	20.4	29.0	11.6
130	8.2	12.2	18.3	26.8	38.1	14.6
135	12.4	18.4	27.3	39.5	55.6	25.6
140	17.0	24.8	36.5	52.7	74.1	34.8
145	23.9	35.1	51.6	74.1	103.9	50.9
150	39.6	57.7	84.3	121.0	167.6	84.1

10 Results

Figures 19-21 show no evidence of a Higgs boson. We calculate upper limits on the excluded Higgs boson cross-section at the 95% confidence level (CL). Table 4 has the limits from the combination of the SS and SJ b -tagging category. The limits for the SS and SJ are in tables 5 and 6, respectively. All the limits in the tables are normalised to the expected Higgs signal cross-section.

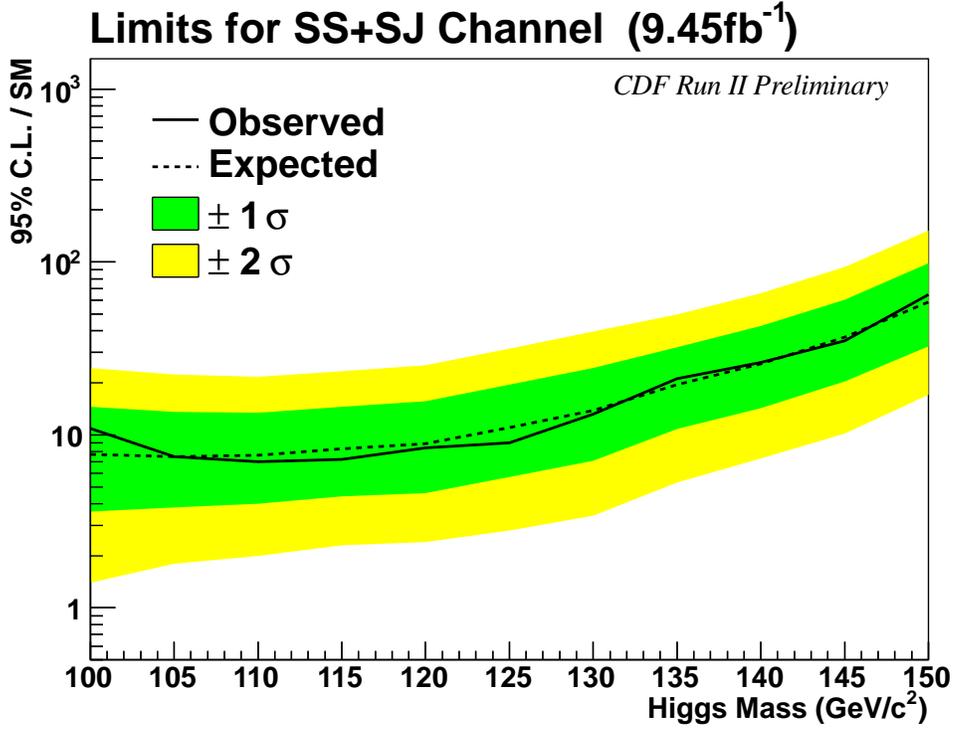


Figure 22: Limits for combined SS & SJ channels: The expected & observed limits are plotted as a function of the Higgs mass. The limits are normalised to the expected Higgs cross-section.

Table 6: Expected and observed 95% CL upper limits for the SJ channel using 9.45 fb^{-1} of $p\bar{p}$ data collected by the CDF detector. The limits are normalised to the expected Higgs cross-section.

Higgs mass	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Observed
100	7.8	10.9	15.5	22.0	30.6	12.2
105	8.7	12.2	17.3	24.5	34.2	15.0
110	8.4	11.7	16.7	23.5	32.5	13.8
115	9.3	12.7	17.9	25.2	35.1	14.7
120	10.3	14.3	20.3	28.9	40.3	15.4
125	12.7	17.6	25.1	35.5	49.3	19.4
130	16.2	22.4	32.2	46.1	65.1	24.7
135	18.6	25.6	36.2	50.8	70.2	26.1
140	24.1	32.9	46.3	65.3	90.7	33.9
145	34.8	47.6	67.0	95.2	130.6	48.0
150	56.0	76.1	104.8	142.9	185.6	78.2

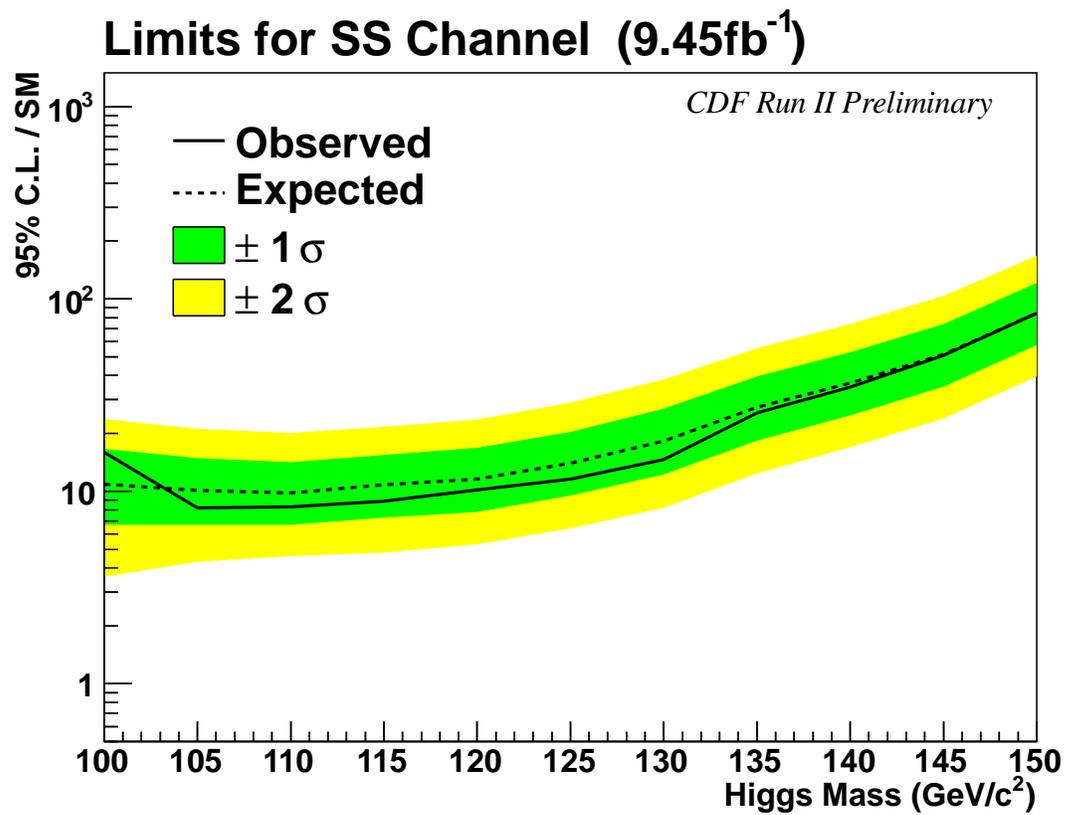


Figure 23: Limits for the SS channel: The expected & observed limits are plotted as a function of the Higgs mass. The limits are normalised to the expected Higgs cross-section.

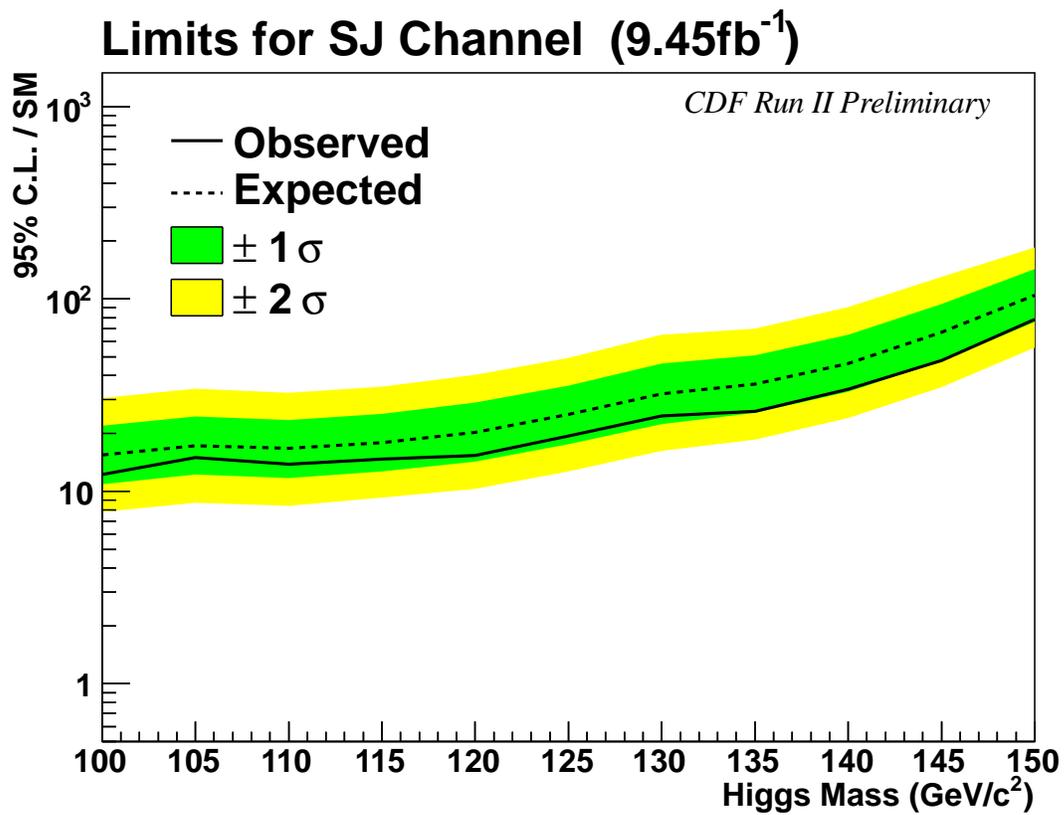


Figure 24: Limits for the SJ channel: The expected & observed limits are plotted as a function of the Higgs mass. The limits are normalised to the expected Higgs cross-section.

11 Conclusions

We have presented the search for the Standard Model Higgs boson in the all hadronic mode using 9.45 fb^{-1} of data collected by the CDF detector. A Neural Network was used to separate the background events from the signal. As the presence of a Higgs signal was not observed, 95% confidence limits were calculated. The median expected limit for Higgs mass 120 GeV is 8.9 while the observed limit is 8.4.

The improvements to the analysis with respect to the previous 4 fb^{-1} version are:

- increase of analyzed data from 4 fb^{-1} to 9.45 fb^{-1}
- improvements to the b-jet energy resolution
- new jet-width measurement
- improvements to QCD modeling (TRF).
- unification of VH and VBF signal regions
- more variables used in NN: result of improved TRF modeling
- improved NN training
- tuning of TRF NN output using information from control regions.

12 Appendix

12.1 Z+jets Generator Level Filter

The large cross-section for Z+jets would produce an extremely large number of events; of which many would be rejected by the trigger. A filter was devised to select events which were likely to pass the trigger.

- At generator level, select events with ≥ 1 b or c parton.
- ≥ 3 jets with $E_T > 5$ GeV. The jets were defined by cone sizes of 0.4, 0.7 and 1.0.
- The Sum- E_T for the 0.4, 0.7 and 1.0 jets are calculated. Accept the event if any of these sums ≥ 60 GeV

References

- [1] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998. 3
- [2] Ankush Mitra, Shang-Yuu Tsai, and Song-Ming Wang. A search for the higgs boson in the all hadronic channel with data sample of 4fb^{-1} . CDF/ANAL/EXOTIC/CDFR 9955, Academia Sinica, December 2009. 3
- [3] T. Aaltonen. Search for the higgs boson in the all-hadronic final state using the cdf ii detector”. *Phys. Rev. D; Physical Review D*, 84(5), 2011. 3, 16, 17
- [4] CDF Collaboration. *FERMILAB-PUB-96/390-E*. 3
- [5] A.Sill et al. *Nucl. Instrum. and Methods A*, 447:1, 2000. 3
- [6] CDF uses a cylindrical coordinate system in which θ is the polar angle to the proton beam, ϕ is the azimuthal angle about the beam axis, and pseudorapidity is defined as $\eta = -\ln \tan(\theta/2)$. The transverse energy and transverse momentum are defined as $E_T = E\sin\theta$ and $p_T = p\sin\theta$, where E is energy measured in the calorimeter and p is momentum measured by the tracking system. 3
- [7] F. Abe et al. Topology of three-jet events in $p\bar{p}$ collisions at $s = 1.8\text{ tev}$. *Phys. Rev. D*, 45(5):1448–1458, Mar 1992. 4
- [8] A. Bhatti et al. Determination of the jet energy scale at the Collider Detector at Fermilab. *Nucl. Instrum. Meth.*, A566:375–412, 2006. 4, 7
- [9] D. Acosta et al. 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. *Physical Review D (Particles and Fields)*, 71(5):052003, 2005. 4
- [10] A. Abulencia et al. 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 jet probability b-tagging. *Physical Review D (Particles and Fields)*, 74(7):072006, 2006. 4
- [11] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003. 5
- [12] Torbjorn Sjostrand, Leif Lonnblad, and Stephen Mrenna. PYTHIA 6.2: Physics and manual. 2001. hep-ph/0108264. 5
- [13] H. L. Lai et al. Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions. *Eur. Phys. J.*, C12:375–392, 2000. 5
- [14] A. Hocker et al. TMVA: Toolkit for multivariate data analysis. 2007. 16
- [15] S. Geer and T. Asakawa. Analysis of multijet events produced at high energy hadron colliders. *Phys. Rev. D*, 53(9):4793–4805, May 1996. 17