

## Combination of CDF's Higgs Boson Searches With Up to $10 \text{ fb}^{-1}$ of Data

---

**Azeddine Kasmi**<sup>\*†</sup>

*Baylor University*

*E-mail:* [kasmi@fnal.gov](mailto:kasmi@fnal.gov)

A combination of the results of searches for the standard model Higgs boson is presented. The searches use up to  $10 \text{ fb}^{-1}$  of Tevatron collider Run II data. We present upper bounds at the 95% confidence level on the production rate of a standard-model-like Higgs boson in the mass range 90 – 200 GeV. We further present measurements of the cross section times branching ratio for Higgs bosons decaying to  $b\bar{b}$ ,  $\tau^+\tau^-$ ,  $W^+W^-$ , and  $\gamma\gamma$ . We evaluate the significance of observed excesses over the standard model background predictions.

*36th International Conference on High Energy Physics,  
July 4-11, 2012  
Melbourne, Australia*

---

<sup>\*</sup>Speaker.

<sup>†</sup>On behalf of the CDF Collaboration.

## 1. Introduction

The Higgs boson ( $H$ ) is the last particle of the standard model (SM) of particle physics that remains to be observed in nature. In electroweak theory, the Higgs boson serves both to give the  $W$  and  $Z$  bosons their masses and to give the fermions mass. In the SM, the couplings of the Higgs boson to other particles is completely specified and is a function of the mass of the Higgs boson ( $m_H$ ), which itself is unspecified. Direct searches at LEP give a lower limit on  $m_H$  of  $114.4\text{ GeV}/c^2$  [1]. With the latest  $W$  boson and top quark mass measurements from the Tevatron [2], the precision electroweak data is consistent with a Higgs boson mass less than  $152\text{ GeV}/c^2$  at the 95% confidence level (C.L.), assuming no physics beyond the SM. A new particle has been observed by the LHC collaborations ATLAS [3] and CMS [4]. The properties of this particle measured so far are consistent with those expected for the SM Higgs boson with a mass near  $125\text{ GeV}/c^2$ . The decay channels contributing the most to the significance of these observations are  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ \rightarrow 4l$ . The center of mass energy of the Tevatron is lower than that of the LHC, so the Tevatron experiments are sensitive to a different mixture of production and decay modes than the LHC experiments. In this note, we summarize the CDF published analyses with the full data set.

## 2. The CDF II Detector

The CDF II detector is described in detail in [5]. With the exception of the  $H \rightarrow \tau^+\tau^-$  search, all other results combined in this note make use of the full CDF Run II data sample. To properly model the data, different data quality criteria are applied, which is reflected in different luminosities.

## 3. SM Higgs Boson Predictions

The kinematic distributions of Higgs boson signal events are modeled via the PYTHIA [6] Monte Carlo (MC) generator, with CTEQ5L and CTEQ6L1 [7] leading-order (LO) parton distribution functions. We scale these MC predictions to the most recent higher-order calculations of cross sections to model, for example, the Higgs boson  $p_T$  spectrum and the number of associated jets. The Higgs boson decay branching ratio predictions used for these results are those of Ref. [8].

## 4. Search Channels

The event selections typically consist of a preselection based on event topology and kinematics, and a final subsequent selection using multivariate analysis techniques (MVA) that combine several discriminating variables into a single final discriminant to separate signal from background. Each channel is divided into exclusive subchannels according to various lepton, jet multiplicity, and  $b$ -tagging characterization criteria. Events with similar signal-to-background ratio are grouped to optimize the overall sensitivity. Such subdivision allows the inclusion of less sensitive topologies while ensuring that the sensitivity of the analysis as a whole will not be degraded. In addition, sensitivity is further improved by exploiting the different dominant signal and background components

in each subchannel during the training of the multivariate discriminants. The final multivariate discriminant is optimized to separate signal from backgrounds for each subchannel and at each Higgs boson mass hypothesis within the test range.

#### 4.1 $H \rightarrow b\bar{b}$ Final States

At the Tevatron center of mass energy, the most relevant SM Higgs production mechanisms are gluon-gluon fusion, associated production (VH), and vector boson fusion (VBF). The study of a Higgs produced by gluon-gluon fusion is very cumbersome if it subsequently decays via  $H \rightarrow b\bar{b}$  because the multijet production largely overwhelms the signal. However, associate production of a Higgs with a  $W$  or  $Z$  boson provides a good trigger if the  $W$  decays leptonically or the  $Z$  decays into an  $e$  or  $\mu$  pair. These channels have been the center of focus of Higgs searches at Tevatron at low masses [9]. These analyses benefit from the neural network  $b$  tagger [10].

##### 4.1.1 $WH \rightarrow lvb\bar{b}$ Channels

In this analysis [11], events are analyzed in two-and-three-jet sub-channels separately, and in each of these samples the events are grouped into various lepton and  $b$ -tag categories. Events are grouped into separate analysis categories based on the quality of the identified lepton. Separate categories are used for events with each lepton category, such as a high quality central  $\mu$  or central  $e$  candidate, etc. Within the lepton categories there are five  $b$ -tagging categories considered for two-jet events: two tight  $b$  tags, one tight  $b$  tag and one loose  $b$  tag, a single tight  $b$  tag, two loose  $b$  tags, and a single loose  $b$ -tag. For three-jet categories only the double-tag categories are considered.

##### 4.1.2 $ZH \rightarrow llb\bar{b}$ Channels

This analysis [12] requires two isolated leptons and at least two jets. Neural networks are used to select di-electron and di-muon candidates. The lack of missing energy from neutrinos allows for a significantly better dijet mass resolution in this channel than in the WH analysis. The jet energies are corrected for  $\cancel{E}_T$  using a neural network approach. The analysis uses four of the WH tagging categories except the single loose  $b$  tag. Events with two or three jets are separated into independent analysis sub-channels.

##### 4.1.3 $WH, ZH \rightarrow \cancel{E}_T b\bar{b}$ Channels

In this analysis [13], the event selection is similar to the WH selection except all events with isolated leptons are rejected and stronger multijet background suppression techniques are applied. The analysis uses a track-based missing transverse momentum calculation as a discriminant against false  $\cancel{E}_T$ . A NN is trained to reject multijet background before  $b$  tagging. Events not passing a requirement on this NN are used to calibrate the multijet rate. A NN network is constructed to separate the Higgs boson signal from the remaining backgrounds, such as  $W+$  heavy flavor and  $t\bar{t}$  production. The analysis uses three non-overlapping categories of  $b$ -tagged events.

##### 4.1.4 $H \rightarrow W^+W^-$ Channels

The signal events are characterized by large  $\cancel{E}_T$  and two oppositely signed, isolated leptons. The signal events are categorized in five nonoverlapping samples, split into high  $s/b$  and low  $s/b$

categories defined by lepton types and the number of reconstructed jets: 0, 1, or 2+ jets. The inputs include likelihoods constructed from calculated matrix-element probabilities as additional inputs for the 0-jet bin. Another channel is the low di-lepton mass  $m_{l+l^-}$  channel, which accepts events with  $m_{l+l^-} < 16 \text{ GeV}/c^2$ . Moreover, opposite-sign channels in which one of the two lepton candidates is a hadronically decaying tau lepton are also included. The final discriminants are obtained from boosted decision trees, which incorporate both hadronic tau identification and kinematic event variables as inputs.

#### 4.1.5 All Other Channels

Although the sensitivity to a low-mass SM Higgs boson at the Fermilab Tevatron is highest for channels involving  $H \rightarrow b\bar{b}$  decay, other channels still contribute significantly to the combined Higgs search. Among the channel included in the current combination are the diphoton [14] final states, ditau final states [15], all-hadronic final states [16], four-lepton final states [17], and a Higgs boson produced in association with two top quarks [18].

### 5. Systematic Uncertainties

The Higgs boson signal is expected to be small compared with the large backgrounds. Systematic uncertainties on the background predictions for the signal selection criteria are usually larger than the expected signal rates. The MVA techniques purify subsets of the expected signals in bins of the final discriminants and help constrain the background rates. Extrapolation of background rate information from one part of a histogram to another is affected by uncertainties in the shapes of the predicted distributions. Systematic uncertainties are also correlated. The nature of the fits that are performed requires careful evaluation of the common and independent sources of systematic uncertainty.

### 6. Statistical Methods

The upper limits on Higgs boson production and the cross section are set based on a Bayesian calculation taking a uniform prior on the signal strength. The p-values are computed with a Frequentist method, although the handling of the systematic uncertainties is Bayesian. The likelihood function is a product over all channels of the Poisson probability of observing the data given the predictions, which depends on the values of the nuisance parameters that parameterize the systematic uncertainties:

$$L(R, \vec{s}, \vec{b}, \vec{n}, \vec{v}) \times \pi(\vec{v}) = \prod_{i=1}^{N_c} \prod_{j=1}^{N_{bins}} (R s_{ij} + b_{ij})^{n_{ij}} \frac{e^{-(R s_{ij} + b_{ij})}}{n_{ij}!} \times \prod_{k=1}^{n_{sys}} e^{-v_k^2/2} \quad (6.1)$$

where the first product is over the number of channels ( $N_c$ ), and the second product is over histogram bins containing  $n_{ij}$  events. The observed number of events in bin  $j$  of channel  $i$  is  $n_{ij}$ . The SM signal prediction in bin  $ij$  is  $s_{ij}$ , summed over all production and decay modes to which channel  $i$  is sensitive, and  $b_{ij}$  is the corresponding background prediction in that bin. The predictions  $s_{ij}$  and  $b_{ij}$  are functions of the nuisance parameters  $\vec{v}$ . The prior distributions for the nuisance parameters are independent Gaussians, and the units in which the nuisance parameters are expressed are in

standard deviations with respect to their nominal values. The factor  $R$  is a simultaneous scaling of all signal components. Thus, each combination presented here assumes that the relative ratios of the contributing Higgs boson production and decay modes are as predicted by the model being tested, within their theoretical uncertainties.

To calculate the best-fit value of  $R$  given all the data we assume a uniform prior density  $\pi(R)$  for positive values of  $R$  and zero for negative values of  $R$ , and integrate the likelihood function  $L$  multiplied by the priors for the nuisance parameters over the values of the nuisance parameters.

$$L'(R, \vec{s}, \vec{b}, \vec{n}) = \int L(R, \vec{s}, \vec{b}, \vec{n}, \vec{v}) \pi(\vec{v}) d\vec{v}. \quad (6.2)$$

The best-fit value of  $R$ ,  $R_{fit}$ , is the value that maximizes the posterior density  $L'(R)\pi(R)$ . The 95% confidence upper limit on  $R$ ,  $R_{95}$ , is defined as follows:

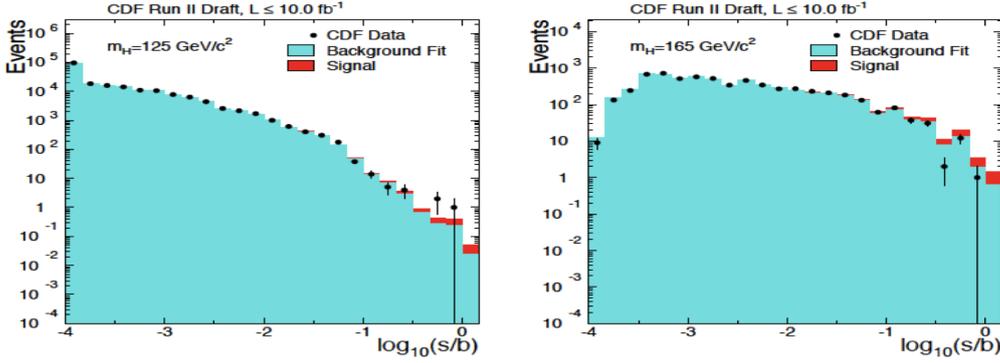
$$0.95 = \int_0^{R_{95}} L'(R)\pi(R) dR. \quad (6.3)$$

To evaluate the significance of excess data events compared with the background prediction, we compute a p-value, which is the probability to observe data as more signal-like than the data assuming that no signal is truly present. A data outcome with a p-value less than  $1.35 \times 10^{-3}$  is customarily identified as having a three standard deviations excess, where the correspondence between the p-value and the number of standard deviations is computed using the integral of one tail of a Gaussian distribution.

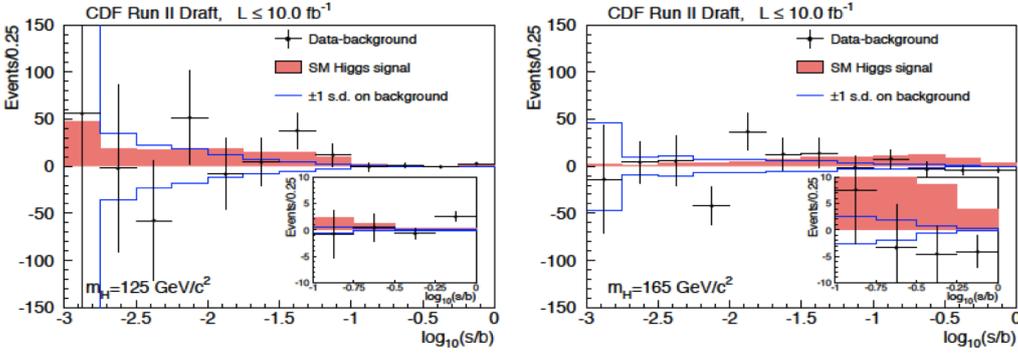
## 7. Results

The data are categorized in a total of 81 sub-channels at  $m_H = 125 \text{ GeV}/c^2$ . In order to visualize the results and identify which data events are causing the observed limits and p-values to be high or low, we perform a joint fit of all channels' background predictions to the corresponding data floating all nuisance parameters, and then collect together all histogram bins, adding together bins with similar s/b. The results are shown in Fig. 1, for the searches at  $m_H = 125 \text{ GeV}/c^2$  and  $165 \text{ GeV}/c^2$ . The signal is shown at its nominal prediction, summed over the same bins as the background and data. These data are also shown in Fig. 2, grouped with wider bins in the s/b ratio, but this time with the best-fit background subtracted for the same two  $m_H$  hypotheses. An excess of data events can be seen in the rightmost bins, those with the highest s/b, in the  $m_H = 125 \text{ GeV}/c^2$  distributions, and a deficit of data is seen in the highest s/b bins for the  $m_H = 165 \text{ GeV}/c^2$  search.

We set upper limits for the SM Higgs boson production cross section as shown in Fig. 3. The limits are expressed as a multiple of the SM prediction for test masses from 90 to 200  $\text{GeV}/c^2$  with a 5  $\text{GeV}/c^2$  step. The points are joined by straight lines for better readability. The bands indicate the 68% and 95% probability regions where the limits can fluctuate, in the absence of signal. The dot-dashed line represents the assumption that the SM Higgs boson is present with a mass of  $125 \text{ GeV}/c^2$ .



**Figure 1:** Collected discriminant histograms, summed for bins with similar signal-to-background ratio ( $s/b$ ), for the search for the SM Higgs boson at  $m_H = 125 \text{ GeV}/c^2$  (left), and at  $m_H = 165 \text{ GeV}/c^2$  (right). The background is fit to the data. The signal model, scaled to the SM expectation, is shown by a filled histogram.



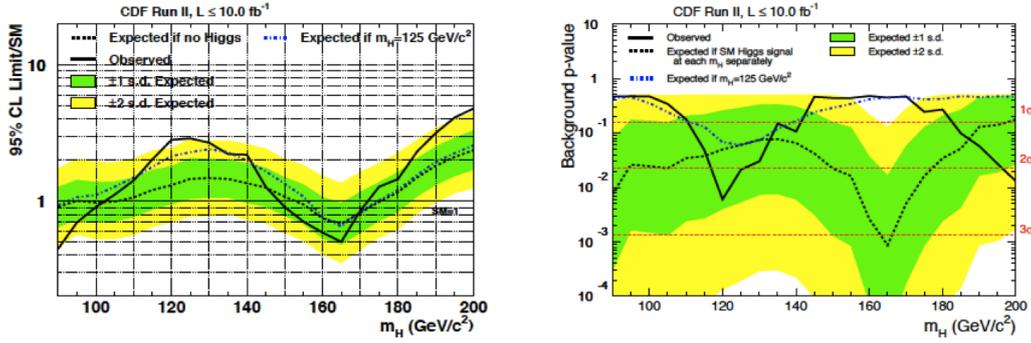
**Figure 2:** Background-subtracted distribution for the discriminant histograms, summed for bins with similar signal-to-background ratio ( $s/b$ ), for the search for the SM Higgs boson at  $m_H = 125 \text{ GeV}/c^2$  (left), and at  $m_H = 165 \text{ GeV}/c^2$  (right). The background is fit to the data, and the uncertainty on the background, shown with the unfilled histogram, is after the fit. The signal model, scaled to the SM expectation, is shown with filled histograms. Uncertainties on the data points correspond to the square root of the sum of expected signal and background yields in each bin.

## 8. Conclusion

The combination of CDF's final searches for Higgs bosons in the context of the SM has been presented. It excludes the mass ranges  $90 < m_H < 102 \text{ GeV}/c^2$  and  $149 < m_H < 172 \text{ GeV}/c^2$  at the 95% C.L. An excess of data with respect to the background prediction is present, which has a local significance of 2.0 standard deviations at  $m_H = 125 \text{ GeV}/c^2$ .

## References

- [1] ALEPH Collaboration and DELPHI Collaboration and L3 Collaboration and OPAL Collaboration and The LEP Working Group for Higgs Boson Searches, Physics Letters B, 565 (17): 61–75, 2003.



**Figure 3:** (Left) Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits for the SM Higgs boson search as functions of the Higgs boson mass. (Right) The background p-value for Higgs boson production for all channels combined.

- [2] The CDF and D0 Collaborations and the Tevatron Electroweak Working Group, arXiv:1204.0042 (2012); T. Aaltonen et al. (CDF Collaboration, D0 Collaboration) (2012), arXiv:1207.1069.
- [3] G. Aad et al. (ATLAS Collaboration), Phys. Lett. B 716, 1 (2012).
- [4] S. Chatrchyan et al. (CMS Collaboration), Phys.Lett. B 716, 30 (2012).
- [5] D. Acosta et al. (CDF Collaboration), Phys. Rev. D 71, 052003 (2005); D. Acosta, et al. (CDF Collaboration), Phys. Rev. D 71, 032001 (2005); A. Abulencia, et al. (CDF Collaboration), J. Phys. G Nucl. Part. Phys. 34, 2457 (2007).
- [6] T. Sjostrand, S. Mrenna, and P. Skands, J. High Energy Phys. 05 (2006) 026. We use PYTHIA version 6.216 to generate the Higgs boson signals.
- [7] H. L. Lai et al., Eur. Phys. J. C 12, 375 (2000); J. Pumplin et al., J. High Energy Phys. 0207, 012 (2002).
- [8] S. Dittmaier et al. (LHC Higgs Cross Section Working Group), arXiv:1201.3084 (2012); S. Dittmaier et al. (LHC Higgs Cross Section Working Group) (2011), arXiv:1101.0593.
- [9] T. Aaltonen et al., The CDF Collaboration, Phys. Rev. Lett. 109, 111802 (2012) [arXiv:1207.1707].
- [10] J. Freeman, T. Junk, M. Kirby, Y. Oksuzian, T.J. Phillips, F.D. Snider, M. Trovato, J. Vizan, and W.M. Yao, Nucl. Instrum. Methods A 697, 64 (2012).
- [11] T. Aaltonen et al., The CDF Collaboration, Phys. Rev. Lett. 109, 111804 (2012) [arXiv:1207.1703].
- [12] T. Aaltonen et al., The CDF Collaboration, Phys. Rev. Lett. 109, 111803 (2012) [arXiv:1207.170].
- [13] T. Aaltonen et al., The CDF Collaboration, Phys. Rev. Lett. 109, 111805 (2012) [arXiv:1207.1711].
- [14] T. Aaltonen et al., The CDF Collaboration, Physics Letters B 717 (2012) 173-181.
- [15] T. Aaltonen et al., The CDF Collaboration, arXiv:1201.4880.
- [16] T. Aaltonen et al., The CDF Collaboration, arXiv:1208.6445 (submitted to JHEP).
- [17] T. Aaltonen et al., The CDF Collaboration, Phys. Rev. D 86, 072012 (2012).
- [18] T. Aaltonen et al., The CDF Collaboration, Phys.Rev.Lett. 109 (2012) 181802 [arXiv:1208.2662].