

Searches for the Higgs Boson at the Tevatron

Alexei Safonov for the CDF and DØ Collaborations

Department of Physics and Astronomy,
Texas A&M University,
College Station, TX 77843, USA

DOI: <http://dx.doi.org/10.5689/UA-PROC-2010-09/53>

Before the LHC experiments have accumulated significant amounts of data, the Tevatron experiments remain the frontier of searches for Higgs, both in the context of the Standard Model as well as scenarios of new physics. With both the CDF and DØ experiments in a smooth operating regime and well understood detectors and data, the focus of both experiments is now on improving the sensitivity of searches with the existing data by utilizing advanced analysis methods. This contribution reports on the status and perspectives of the searches for Higgs boson at Fermilab Tevatron.

1 Introduction

Understanding the mechanism of electroweak symmetry breaking has been a major goal of particle physics for many years. With the Higgs boson, an integral part of such mechanism in the Standard Model (SM), still not seen experimentally, Higgs searches are a central part of the Fermilab Tevatron physics program. Search strategies for the Higgs boson at the Tevatron are determined by the phenomenology of production and decay for particular mass of Higgs m_h as well as the dynamics of the backgrounds. The dominant production mechanism for Higgs is gluon fusion $gg \rightarrow h$, as shown in Fig. 1(a), followed by the associated production of Higgs with a W or Z boson. The next is the vector boson fusion (VBF) Higgs production $qq \rightarrow qqh$, which becomes more significant than the associated production for higher Higgs masses. Higgs decay branching fractions strongly depend on mass of the Higgs boson, as shown in Fig. 1(b). For lower masses, $m_h \leq 135 \text{ GeV}/c^2$, the dominant decay mode is $h \rightarrow bb$ followed by $h \rightarrow \tau\tau$ and the rest of the decay modes being essentially insignificant. The only other decay worth mentioning is $h \rightarrow \gamma\gamma$, which is interesting because at the LHC this is the main search mode for the low mass Higgs and relies on very good resolution of electromagnetic calorimeters. At higher masses, the dominant Higgs decay modes are $h \rightarrow WW$ and $h \rightarrow ZZ$.

Because SM backgrounds in Higgs searches are large, search strategies are largely modified compared to simple estimations based on cross-section and branching fraction considerations. Particularly for low m_h , the search channel using dominant production and decay modes $gg \rightarrow h \rightarrow bb$ is completely dominated by the QCD bb backgrounds. Because of that, search strategies at low mass rely mainly on the associated production $pp \rightarrow VH$ (V can be Z or W boson) where a substantial background rejection can be obtained by requiring the presence of a lepton from Z or W and/or significant missing transverse energy \cancel{E}_T from W or $Z \rightarrow \nu\nu$. The targeted Higgs decay mode in these searches is usually $h \rightarrow b\bar{b}$, but one can also look for e.g. the $h \rightarrow \tau\tau$

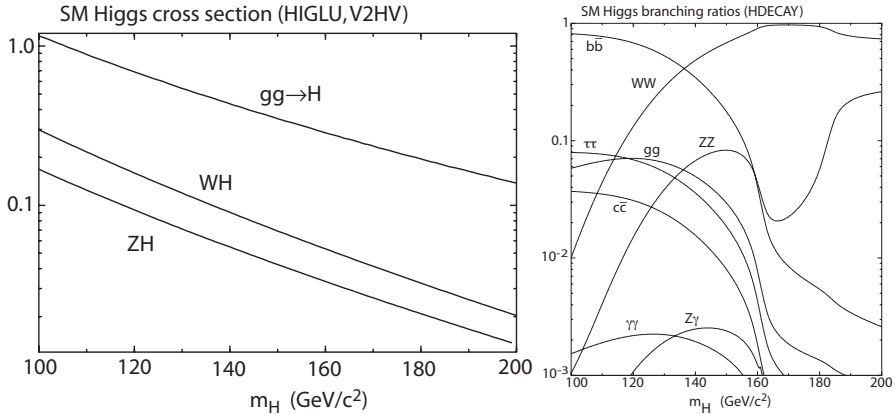


Figure 1: (a) SM Higgs boson production cross section for the Tevatron [1]. (b) Branching fractions as a function of Higgs mass [2].

decay mode. Searches in the $h \rightarrow \tau\tau$ channel can also rely on the direct production $gg \rightarrow h$ process, but the background from the irreducible $Z \rightarrow \tau\tau$ is fairly significant reducing significance of the search in this mode. For $m_h > 135 \text{ GeV}/c^2$, lower backgrounds and cleaner final states of $h \rightarrow WW/ZZ$ with multiple leptons from W and Z bosons allow relying on the dominant $gg \rightarrow h$ production process.

With the CDF and $D\bar{O}$ detectors and systematics being well understood, the focus of both experiments is on maximizing the sensitivity of searches with the existing data while continuing accumulation of the new data statistics. This is achieved by increasing the acceptance using less strict selections, employing advanced multi-variate methods and analysis techniques as well as combining results from multiple searches and between the two experiments. The following describes several selected Higgs search analyses and summarizes the current combined sensitivity of the Tevatron experiments.

2 Searches for Low Mass SM Higgs

As discussed earlier, searches for low mass Higgs mainly rely on $p\bar{p} \rightarrow WH$ and $p\bar{p} \rightarrow ZH$ production modes. While having smaller cross-section compared to the gluon fusion, these processes allow for much cleaner final states. In addition to leptonic decay modes, another important search channel is the $\cancel{E}_T b\bar{b}$ final state, where either Z decays to neutrinos or a lepton from a W escapes detection.

In the $D\bar{O} WH \rightarrow l\nu b\bar{b}$ analyses [3], the data are split by lepton type and jet multiplicity (2- or 3-jet sub-channels). Each category is further sub-divided into non-overlapping sub-categories based on the presence of a single “tight” b -tagged jet or two “loosely” b -tagged jets. The tight and loose b -tagging criteria are defined with respect to the mis-identification rate for light quark or gluon jets, which is typically $\sim 0.5 - 1.5\%$. Each sub-channel is analyzed separately. The outputs of random forests, trained separately for each sample and for each Higgs mass, are used as the final discriminating variables in the limit setting procedure.

The CDF $WH \rightarrow l\nu b\bar{b}$ analysis [4] uses events categorized in two and three jet sub-channels, and in each of these samples the events are grouped into various lepton and b -tag categories.

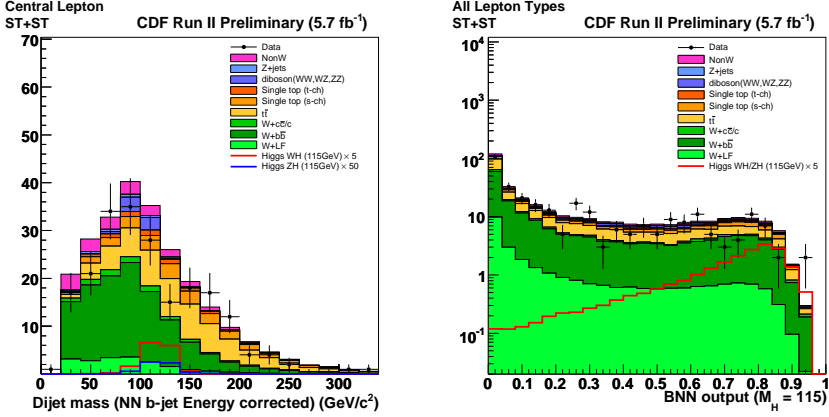


Figure 2: (a) The dijet invariant mass after NN b-jet energy correction in the CDF search in $WH \rightarrow l\nu bb$ channel for the sample with two tight b -tagged jets and a central lepton. (b) The BNN output for all event categories contributing to the signal region.

To increase acceptance, events with an isolated track failing lepton selection requirements in the two jet sample, or an identified loose muon in the extended muon coverage in the three jet sample, are added as separate categories in addition to the samples with a well defined lepton. Within each lepton category, there are four b -tagging categories considered in the two jet sample: two tight b -tags, one tight and one loose b -tag, one tight and one looser b -tag, and a single tight b -tag. The same b -tagging categories are used for the three jets channel, except the category with one tight and one looser b -tags and those events are added to one tight b -tag category. A Bayesian neural network (BNN) discriminant is trained at each m_h in the test range for the two jet sample, separately for each lepton and b -tagging category, while for the three jet sample a matrix element (ME) discriminant is used. An illustration showing the invariant mass of the dijet system for the two tight b -tagged sample is shown in Fig. 2(a) and the output of the BNN is shown in Fig. 2(b).

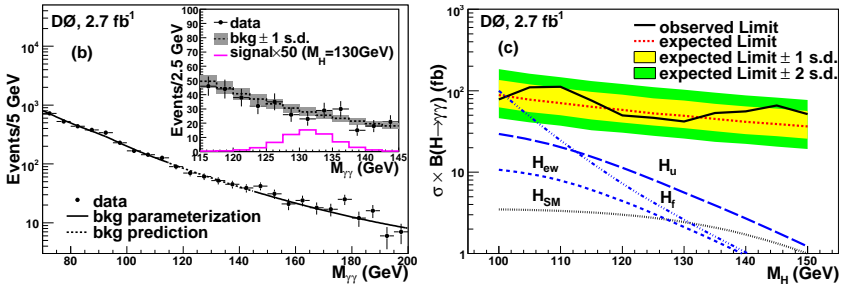


Figure 3: (a): $m_{\gamma\gamma}$ spectrum in data compared to the background parameterization. The inset figure compares the data to the background prediction inside the search region including its one standard deviation uncertainty band and the expected SM Higgs signal scaled by a factor of 50 for $m_h = 130 \text{ GeV}/c^2$. (b): Observed and expected 95% C.L. upper limits on $\sigma \times B$ as a function of m_h compared to the SM prediction and several non-SM EWSB scenarios.

For the $ZH \rightarrow \nu\nu bb$ analyses [5], selections are similar to the WH analyses, except all events with isolated leptons are vetoed and stronger multijet background suppression is applied. Both CDF and $D\mathcal{O}$ analyses use a track-based missing transverse momentum calculation as a discriminant against false \cancel{E}_T . In addition, both CDF and $D\mathcal{O}$ utilize multi-variate techniques, a boosted decision tree (BDT) at $D\mathcal{O}$ and a neural network (NN) at CDF, to further discriminate against the multi-jet background before b -tagging. There is a sizable fraction of the $WH \rightarrow l\nu bb$ signal in which the lepton is undetected that is selected in the $ZH \rightarrow l\nu bb$ samples, so these analyses are also referred to as $VH \rightarrow \cancel{E}_T bb$. The CDF analysis uses three non-overlapping categories of b -tagged events (two tight b -tags, one loose and one tight, and a single tight b -tag as in the three jet $WH \rightarrow \cancel{E}_T bb$ channels). $D\mathcal{O}$ uses orthogonal single tag and tight-loose double-tag channels. The CDF analysis relies on NN outputs for the final discriminating variables, while $D\mathcal{O}$ uses the BDT outputs.

The $ZH \rightarrow llbb$ analyses [6] require two isolated leptons and at least two jets. $D\mathcal{O}$'s $ZH \rightarrow llbb$ analyses separate events into non-overlapping samples of events with one tight b -tag and two loose b -tags. CDF separates events into single tag, double tag and loose double tag samples. To increase signal acceptance, $D\mathcal{O}$ has loosened the selection criteria for one of the leptons to include an isolated track not reconstructed in the muon detector or an electron from the intercryostat region of the $D\mathcal{O}$ detector. Combined with the dielectron and dimuon analyses, these provide four orthogonal samples, and each uses 4.2 fb^{-1} of data. Most recently the ee and $\mu\mu$ channels have been updated to include 6.2 fb^{-1} of data. CDF has added for this combination additional sub-channels for candidate events with two loose muon candidates selected using a NN discriminant. Events in the new category are analyzed separately depending on the trigger path (muon or \cancel{E}_T -based), and the jet energies include the NN-based corrections for \cancel{E}_T . For the $D\mathcal{O}$ analysis random forests of decision trees provide the final variables for setting limits, while CDF utilizes outputs of two-dimensional NN incorporating likelihoods based on event probabilities obtained from ME calculations as additional inputs.

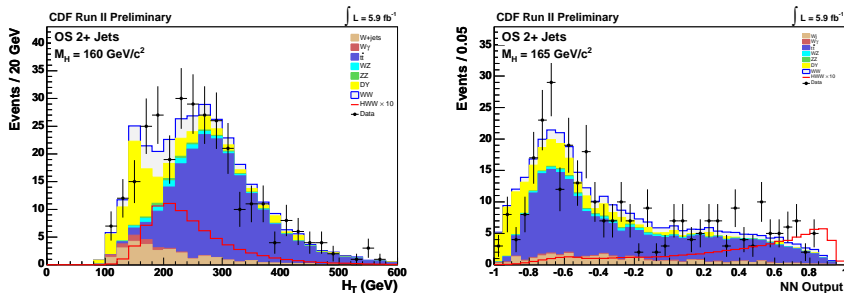


Figure 4: (a): The distribution of H_T defined as a scalar sum of lepton p_T , jet E_T and \cancel{E}_T for the CDF $H \rightarrow WW$ search used as one of the input variables in the NN discriminant. (b) NN output templates for the channel with two leptons and 2+ jets and $m_h = 165 \text{ GeV}/c^2$.

The background subtracted data is shown in Fig. 5(a) compared to the uncertainties from the background estimation and the expected Higgs signal for $m_h = 115 \text{ GeV}/c^2$. The combined limit using CDF and $D\mathcal{O}$ results [7] for the WH and ZH channels can be seen in Fig. 5(b). To give a feel of the sensitivities of each of these channels, Table 1 shows expected production cross section limit divided by the expected SM Higgs production cross-section for each of the search channels assuming $m_h = 115 \text{ GeV}/c^2$.

Channel	$\sigma_{exp}^{95\%}/\sigma^{SM}$
$WH \rightarrow l\nu bb$	3.5
$VH \rightarrow \cancel{E}_T + bb$	4.0
$ZH \rightarrow llbb$	5.5
$VH \rightarrow qqbb$	18
$h \rightarrow \gamma\gamma$	21
$h \rightarrow \tau\tau jj$	23
Combined	1.9

Table 1: Illustration of relative sensitivities of different channels used in the low mass Higgs searches at CDF for $m_h = 115 \text{ GeV}/c^2$.

GeV/c^2 with the background estimated from a sideband fit to the mass spectrum in the 70 to 200 GeV/c^2 range excluding the search region. The invariant mass $m_{\gamma\gamma}$ and exclusion limits for this analysis are shown in Figs. 3(a) and (b) for both the SM case as well as the non-SM scenarios with reduced Higgs coupling to fermions.

Both CDF and $D\emptyset$ have recently completed analyses [8] targeting the $h \rightarrow \gamma\gamma$ decay mode. While the expected sensitivity of this analysis is not expected to compete with the channels relying on associated production, it is particularly interesting as this very channel will be used as the primary discovery mode at the LHC if the Higgs boson is light. In the $D\emptyset$ search, events are selected by requiring at least two photon candidates with $p_T > 25 \text{ GeV}/c$ in the central part of the detector. To suppress multi-jet and γ +jet backgrounds where jets are misidentified as photons, a NN is deployed trained using a set of variables sensitive to the tracker activity and energy deposits in the calorimeter and preshower detectors. The dominant backgrounds include γ +jet and dijet events and are directly estimated from data. The $Z/\gamma^* \rightarrow ee$ background with both electrons misreconstructed as photons is estimated using the MC simulation. The signal is searched in windows of $\pm 15 \text{ GeV}/c^2$ (five times the m_h resolution) from 100 to 150

3 Searches for High Mass SM Higgs

For higher Higgs masses (above $m_h = 135 \text{ GeV}/c^2$), the large branching ratio and relatively clean final states make $h \rightarrow WW$ the most sensitive channel. $h \rightarrow ZZ$, while very clean has lower branching and becomes significant above the ZZ threshold.

For the $H \rightarrow WW$ analyses, signal events are characterized by large \cancel{E}_T and two opposite-signed, isolated leptons. The presence of neutrinos in the final state prevents the accurate reconstruction of the candidate Higgs boson mass. $D\emptyset$ selects events containing large \cancel{E}_T and electrons and/or muons, dividing the data sample into three final states: ee , $e\mu$, $\mu\mu$. Final states involving leptonic tau decays and mis-identified hadronic tau decays are included. The ee and $\mu\mu$ analyses are as in ref [21]. The $e\mu$ channel has been updated and now uses 6.7 fb^{-1} of data as well as subdividing the dataset according to the number of jets in the event: 0, 1, or 2+ jets. CDF separates the $H \rightarrow WW$ events into five separate non-overlapping categories split into both “high s/b” and “low s/b” categories based on lepton types and the number of reconstructed jets: 0, 1, or 2+ jets. The sample with two or more jets is not split into low s/b and high s/b lepton categories due to low statistics. A sixth CDF channel is the low dilepton mass m_{ll} channel, which accepts events with $m_{ll} < 16 \text{ GeV}$. This channel increases the sensitivity of the $h \rightarrow WW$ analyses at low m_h by adding $\sim 10\%$ additional acceptance. Because this analysis is also sensitive to $WH + ZH \rightarrow jjWW$, and vector-boson fusion production modes, the division of events into jet categories allows the analysis discriminants to separate three different categories of signals from the backgrounds more effectively and increase the overall sensitivity of the analysis. The $D\emptyset ee$ and $\mu\mu$ channels use NN discriminants, including the number of jets as an input, as the final discriminant while the $e\mu$ channel relies on BDT outputs with additional input variables now included for the 1 and 2+ jet sub-channels. CDF uses NN outputs, including

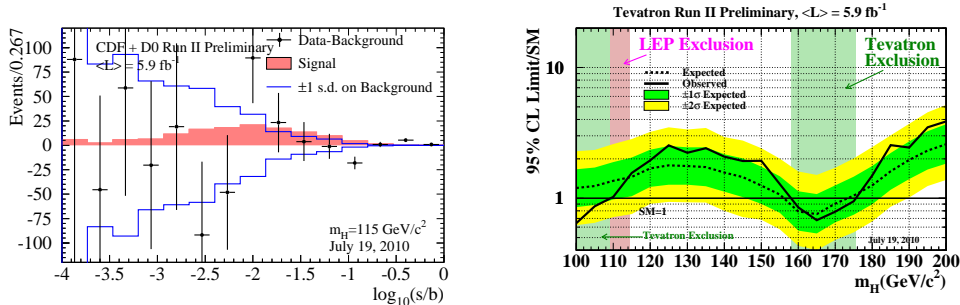


Figure 5: (a): Background-subtracted data distributions for all channels in bins of s/b , for $m_h = 115 \text{ GeV}/c^2$, the error bars are the square roots of the predicted background in each bin. The unshaded histogram shows the systematic uncertainty on the background model, and the shaded histogram shows the expected SM Higgs signal. (b): Observed and expected 95% C.L. upper limits on the Higgs production cross section (as a multiple of the SM prediction) versus m_h for the combined CDF and D0 analyses.

likelihoods constructed from calculated ME probabilities as additional inputs for the 0-jet bin.

4 Summary

Searches for Higgs at the Tevatron continue to provide an important insight on the mechanism of electroweak symmetry breaking. Both experiments continue accumulating data and improving analysis techniques. In addition to excluding a range of allowed masses 158-175 GeV/c^2 , the combined sensitivity of the D0 and CDF experiments for 95% C.L. exclusion in the low mass region approaches the SM expectation.

References

- [1] M. Spira, Nucl. Instrum. Meth. **A389** (1997) 357 .
- [2] M. Spira, Comput. Phys. Commun. 108 (1998) 56.
- [3] V. M. Abazov et al. (D0 Collaboration), arXiv:1012.0874.
- [4] CDF Collaboration, CDF Conference Note 10223.
- [5] V. M. Abazov et al. (D0 Collaboration), Phys. Rev. Lett. 104:071801 (2010); (CDF Collaboration), CDF Conf. Note 10235.
- [6] V. M. Abazov et al. (D0 Collaboration), arXiv.org:1008.3564, CDF Collaboration, CDF Conf. Note 10212.
- [7] D0 and CDF Collaborations, arXiv:1007.4587.
- [8] V. M. Abazov et al. (D0 Collaboration), Phys. Rev. Lett. 102:231801 (2009); (CDF Collaboration), CDF Conf. Note 10065.
- [9] D0 and CDF collaborations, Phys. Rev. D 82:011102 (2010); V. M. Abazov et al. (D0 Collaboration), Phys. Rev. Lett. 104:061804 (2010); CDF Collaboration, CDF Conf. Note 10232.