CDF/ANAL/EXOTIC/GROUP/9596

Combined $WH \to \ell\nu b\bar{b}$ search with 2.7 fb^{-1} of CDF data

The CDF Collaboration

URL http://www-cdf.fnal.gov

(Dated: November 19, 2008)

Abstract

Within the CDF experiment, there are two efforts to search for a Higgs bosons produced in association with a W boson. The analyses use the same 2.7 fb^{-1} of CDF data and very similar $W + 2$ jets event selections, but differ in terms of multivariate techniques. One uses a neural network based on kinematic information, while the other uses a boosted decision tree with matrix element, kinematic variable, and flavor separating neural network information as inputs. In order to improve sensitivity, these two analyses are combined into a single analysis by using them as inputs to another neural network to produce a super-discriminant. No evidence of a Higgs boson is observed and limits on its production rate are obtained for Higgs masses between 100 GeV/c^2 through 150 GeV/c^2. The combined result has an expected sensitivity of 4.8 times the Standard Model Higgs cross section for a Higgs mass of 115 GeV/c^2. The observed limit at that mass is 5.6 times the SM cross section.

Preliminary Results Winter 2008/2009

I. INTRODUCTION

The standard model of particle physics has proven to be an extremely successful theory through its accurate predictions of many experimental results over the last few decades. Although the addition of the Higgs mechanism [1] completed the standard model in the late 1960's [2], its eponymous particle, the Higgs boson, has yet to be discovered. Theoretically the Higgs boson is expected to be about the same mass as the W and Z bosons, while direct limits from the LEP experiments exclude masses below 114 GeV/c^2 [3]. In addition, electroweak precision measurements place an indirect upper limit on the mass of a Standard Model Higgs boson of 144 GeV/c^2 [4].

The search for a light Higgs boson ($115 < m_h < 150\ GeV/c^2$) is well motivated, but also quite challenging at the Tevatron. In this mass range, the $b\bar{b}$ decay mode is important, in fact it is the dominant decay mode for $m_h < 135\ GeV/c^2$. The largest production mode at the Tevatron is gluon fusion [5]. Since the dijet background production rate is many orders of magnitude larger, the $gg \to h \to b\bar{b}$ channel has negligible sensitivity. Instead, associate production with a W or Z boson where the W or Z decays leptonically provide the most sensitive search channels. Here, we consider a search for the Higgs boson in the $WH \to \ell\nu b\bar{b}$ channel.

II. ANALYSIS TECHNIQUE

At CDF the search for $WH \to \ell\nu b\bar{b}$ consists of two different techniques both using the 2.7 fb^{-1} dataset. Both techniques select essentially the same events: they use a sample of events containing a W boson candidate and exactly two jets. The events are triggered either by a high-p_T lepton or a $\not\!\!E_T+$ jets signature. The events are required to have at least one jet tagged by the `SECVTX` b-tagging algorithm [6], and the events are separated into single- and double-tagged categories. Both use multivariate discriminants to enhance the separation of signal from background and the resulting sensitivity is competitive between techniques.

The main differences between these two searches is in the multivariate techniques used. One analysis, which we will refer to as the `NN` analysis, uses quantities calculated from the lepton, $\not\!\!E_T$, and jets (with both tight and loose definitions) kinematics in an artificial neural network to discriminate signal from background [7, 8]. The second analysis, referred to here as the MEBDT analysis, uses a boosted decision tree (BDT) [9] whose inputs include matrix element information calculated from the lepton and tightly defined jet kinematics, a neural network sensitive to the flavor of tagged jets [10], and other event kinematic information [11]. Because these two analyses achieve comparable results using different kinematic information, it is interesting to pursue an analysis technique which combines these two approaches. This interest is further justified by noting that the NN and MEBDT outputs for signal and different backgrounds are not fully correlated.

To achieve a combination of the NN and MEBDT analyses, we employ a super-discriminant technique first developed to combine analyses in the CDF single-top search [12, 13]. The basic idea of a super-discriminant is to take the discriminant outputs from two or more multivariate analyses and use them as the inputs to a new discriminant which will combine the information from the input analyses, hopefully to obtain greater sensitivity. For this combination, we will use as our super-discriminant a neural network, optimized using genetic algorithms [14]. The super-discriminant will have two inputs for each event, the NN and MEBDT output values from the individual analyses. The output of the super-discriminant will provide a new distribution, which will then be analyzed using the same statistical techniques used for the NN and MEBDT analyses.

It should be noted that along the path of doing this combination, we will also incorporate small changes in the event selection so that the NN and the MEBDT analyses use exactly the same selection. In particular, the NN will use the full dataset for the PHX leptons. (Only the first 1.9 fb^{-1} was used for the previous results due to minor technical difficulties.) In addition, the MEBDT analysis will use the isolated track event selection rather than the extended muon coverage selection because the extended muons are largely a subset of the isolated tracks, but the isolated tracks offer additional signal acceptance. Finally, the MEBDT analysis will adopt the same b-tagging categories as were used by the NN analysis, including the double-tag category in which one jet is tagged by the silicon vertex SECVTX algorithm while the other jet is tagged with the jet probability (JetProb) algorithm [15]. These changes result in improvements in the individual analyses, and the combination benefits from these improvements as well.

For the remainder of the note, we will outline the steps necessary to achieve the combination. First we will briefly look at the correlations between the `NN` and `MEBDT` discriminants. Next, we will describe the procedure for constructing the super-discriminant used in this combination. Finally, we will explain the statistical techniques used to analyze the super-discriminant distribution and obtain the combination results.

III. CORRELATIONS

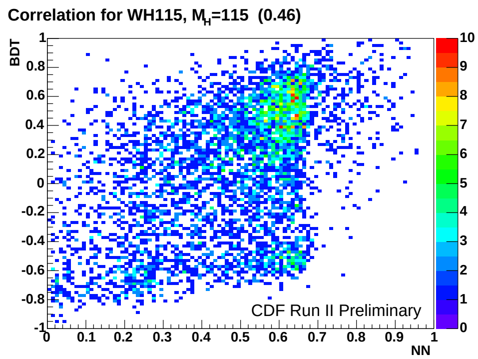
Since they use some of the same input variables, the `NN` and `MEBDT` discriminant outputs are certainly correlated; however, they are not fully correlated. For example, the `MEBDT` analysis uses a neural network b-tagger [10] and this information is not used by the `NN` analysis. Likewise, the `NN` analysis considers quantities calculated from jets with looser P_T and η definitions, which are not used by the `MEBDT` analysis.

It is interesting to look at the correlations among the two analyses for different MC samples. In the most sensitive channel (W + 2 jets, 2 or more b-tag), for a given MC sample we fill from every event a 2D histogram containing: `NN`, `MEBDT`. Fig. 1 shows the 2D correlations for a $WH \rightarrow \ell\nu b\bar{b}$ signal with $M_H = 115$ GeV/c^2 and several background samples. It is clear from the Figs. that the discriminants are highly correlated; however, it is also clear that this correlation is not 100 % so one can hope that combining the information in these two discriminants will provide a more sensitive result. Furthermore, the correlations between signal and background samples appears different, which is a feature that a neural network may be able to exploit.

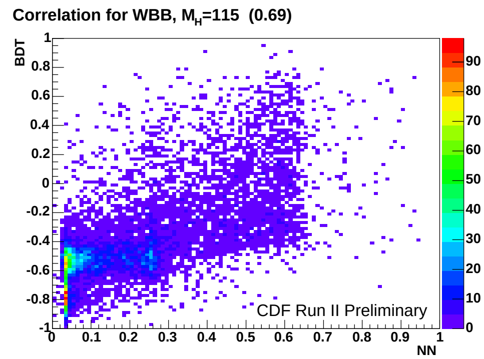
IV. CONSTRUCTING THE SUPER-DISCRIMINANT THROUGH NEURO-EVOLUTION

A. Introduction

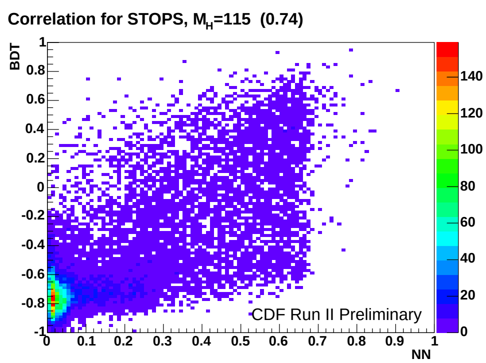
Although there are several multivariate options for combining these two techniques, we choose to form the super-discriminant using an artificial neural network (ANN) with two inputs and a single output, similar to the approach used in the single top combination [13]. In addition, we make use of a novel NN training technique, known as neuro-evolution, in which the neural network architecture and weights are optimized through the use of



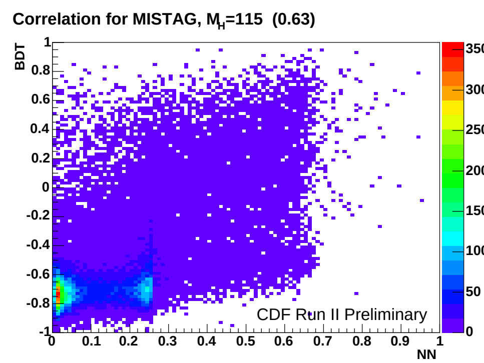
(a) WH signal for Higgs mass of 115 GeV/c^2



(b) Wbb



(c) Single Top s-Channel



(d) Mistags (W + light flavor)

FIG. 1: 2D plots of the correlations between the discriminants used in the `NN` and `MEBDT` for a Higgs boson signal and some of the important backgrounds in the two tag channel. The number in parentheses in the title shows the correlation coefficient.

genetic algorithms. A similar technique was previously employed to improve the event selection in a dilepton top mass measurement [16].

B. Neuro-Evolution

A typical approach to ANN training involves using a gradient descent method, such as backpropagation, to minimize the classification error, defined by $\sum (o_i - t_i)^2$, where o_i is the output of the neural network and t_i is the desired output, usually zero for background and one for signal. Although backpropagation is a powerful and fast technique for training neural networks, it is not necessarily true that an ANN that minimizes the classification error will also provide the greatest sensitivity in a search. Therefore, it is interesting to explore additional training methods capable of optimizing quantities directly related to the problem of interest. One such technique known as neuro-evolution involves using genetic algorithms, rather than backpropagation, to search the space of possible ANN weights. Because this search proceeds stochastically, it can be used to optimize an arbitrary figure of merit, in contrast to gradient descent methods which require a figure of merit that is well-behaved with calculable derivatives. The neuro-evolution package used for this analysis is Neuro-Evolution of Augmenting Topologies (NEAT) [14, 17]. NEAT has the advantage that in addition to optimizing the neural network weights, it also varies the network topology, adding complexity as needed to improve performance.

Neuro-evolution with NEAT begins from a population of neural networks generated from a seed network by randomly varying the network weights. Evolution then proceeds in generations. In each generation, the following steps are completed:

1. The fitness of each neural network is evaluated by calculating the networks performance using a figure of merit.

2. Networks with poor fitness are removed from the population.

3. The remaining networks are allowed to replenish the population through mutation and breeding. Possible mutations include randomly changing one or more of the ANN weights, randomly adding a link between nodes, and randomly adding new nodes. Breeding involves blending randomly selected features from two networks.

The population of networks remaining at the end of this process for one generation becomes the initial population for the next generation.

1. *Fitness Calculation*

We evaluate the fitness of a given neural network as follows: First, the network is used to calculate an output value for each event in a training sample. In this case, the training sample is formed by selecting one-half of the signal and background Monte Carlo samples. The other half of the Monte Carlo is reserved to check the final network for evidence of overtraining. All background processes are included in the training except non-W because it is a relatively minor background and suffers from extremely low statistics. After calculating the ANN output, the values are stored in histograms which can then be used to generate pseudodata and serve as fit templates in the figure of merit calculation. This process is repeated every generation for each network in the population.

The key to obtaining good results in NEAT is to define a figure of merit that is closely related to the quantity to be optimized, but which can be calculated quickly enough to be used to evaluate ANN fitness repeatedly for large populations of neural networks. In this case, the quantity we would like to optimize is the expected limit on the cross section times branching fraction for $WH \to \ell\nu b\bar{b}$. However, calculating the full expected limit is computationally expensive. As a faster alternative figure of merit, we calculate the quadrature sum of expected signal divided by the square root of the expected background ($S/\sqrt{B}$) in each bin of the ANN templates. This figure of merit can be calculated very quickly since there is no integration over nuisance parameters.

2. *Automatic Binning*

Because the neural networks generated during NEAT training are not required to have any specific output range, the binning of the neural network output histograms cannot be determined in advance. Therefore, it is necessary to determine a binning automatically. However, the choice of binning can also strongly affect the performance of a given discriminant. Rather than attempting to develop an algorithm to determine automatically an optimal binning, we choose instead to employ a fixed binning scheme and to ensure the NEAT has sufficient freedom to evolve networks that make optimal use of this binning, as described in the next section. Our fixed binning scheme uses one-hundred bins equally distributed between zero and one. Output values less than zero are placed in the first histogram bin and outputs greater than one are put in the last bin. Furthermore, measures are taken to prevent insufficient Monte Carlo statistics from biasing the expected limit calculation.

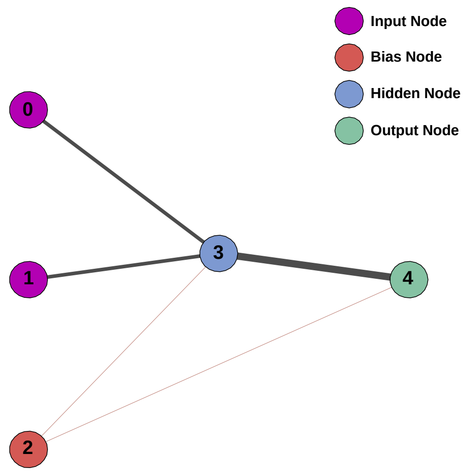


FIG. 2: Starting topology used for NEAT evolution.

3. Network Topology

We configure NEAT to evolve feed-forward neural networks with a single output node. Recursion, or feedback, is disabled. A sigmoid activiation function is used for nodes in hidden layers while the output node uses a linear activation. A basic premise of NEAT is that the initial network topology should be as simple as possible, allowing NEAT to evolve complexity as needed. In general, this means that the seed network for a run of NEAT evolution should be one in which each input is connected directly to the output node with no hidden layers. However, in this case, because we are allowing NEAT not only to search for the optimal ANN shape, but also the optimal binning, it is advantageous to begin with the slightly more complicated initial topology as shown in figure 2. The inputs are connected to a hidden node. The weight between the single hidden node and the output node, therefore, becomes a scaling factor to adjust the range of the ANN output. The connection between the bias node and the output sets the offset for the ANN output. With these two additional degrees of freedom beyond the minimal configuration, NEAT has the ability to adjust the ANN output to take best use of the predefined binning.

Evolution begins from a seed network. For each channel, we use one of the following starting configuration for each of the multiple parallel runs (see below):

- All inputs given equal weights.

- All weights set to zero.
- The weight for one input set to one, and the others set to zero.

4. Training and Selection of the Final Network

Because neuro-evolution performs a stochastic search for the optimal network, using multiple, parallel runs with different random seeds can improve the search speed. For each channel, we use five different parallel runs seeded from the four starting configurations described above. Each run evolves a population of 150 neural networks through up to 200 generations. For each generation, the structure of the best performing network (champion) is saved, yielding up to 200 network configurations for each run. At the conclusion of the evolution, we select a subset of the champion networks with the highest figure of merit and calculate their actual expected limits. The final selected network is chosen to be the one with the best expected limit.

The training procedure outlined above is repeated for each of the three tag channels and for each Higgs mass between 100 GeV/c^2 and 150 GeV/c^2, in 5 GeV/c^2 steps. As an example, the topology the networks selected for the 115 GeV/c^2 Higgs mass is shown in figure 3.

C. Validation

The first step in validating our NEAT neural network is to check for signs of overtraining. Overtraining occurs when an ANN begins to learn the specific features of the training sample. Because we reserve half of the MC sample for testing, we can check for overtraining by looking for shape differences between the ANN output calculated on the training sample and on the testing sample. There are no signs of overtraining revealed by these studies.

V. STATISTICAL TREATMENT

The combination follows exactly the same statistical treatment as the individual analyses. Specifically, a framework is used to calculate limits using a Bayesian approach in

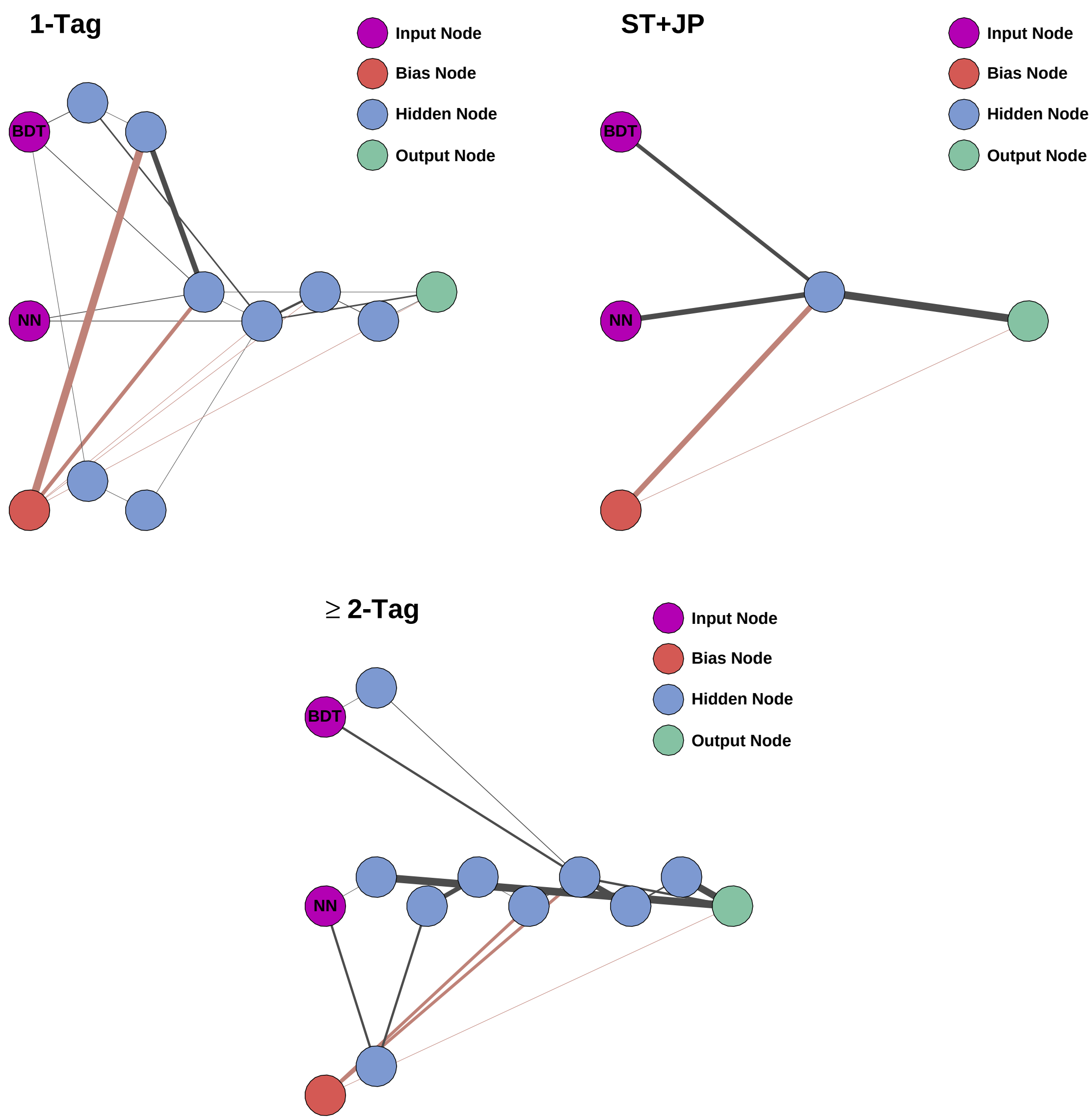


FIG. 3: Topology of the selected networks.

which nuisance parameters are marginalized [18]. The systematic uncertainties are the same as those used in the individual `NN` and `MEBDT` analyses.

A. Channels

The data used for this analysis comes from either the lepton triggered events for the $\not\!\!E_T+$ jets triggered events. The triggered leptons include electrons and muon candidates, while in the $\not\!\!E_T$+ jets sample, we use isolated tracks to identify the lepton. Within each class of events, the data can be further subdivided into three independent tag categories:

- Events having only one `SECVTX` tag and no `JetProb` tags are called "1-Tag" events.
- Events having one `SECVTX`-tagged jet and one `JetProb`-tagged jet are called "ST+JP" events.
- Events with two or more `SECVTX`-tagged jets are called "$\geq$ 2-Tag" events.

Combining the two lepton categories and three tag categories yields six independent channels. These six separate channels are fit simultaneously to obtain the final results.

VI. RESULTS

Figs. 4 - 6 show the NEAT NN output distributions for a Higgs mass of 115 GeV/c^2 observed in data compared to the background prediction. Table I shows the expected and observed limits calculated for all Higgs masses. The limits are displayed graphically in Fig. VI. The combined result has an expected sensitivity of 4.8 times the Standard Model Higgs cross section for a Higgs mass of 115 GeV/c^2. This represents a 15 % gain over previous $WH \rightarrow \ell\nu b\bar{b}$ search results from CDF.

Mass	100	105	110	115	120	125	130	135	140	145	150
Expected	3.54	3.80	4.14	4.81	5.91	7.18	8.72	12.2	17.5	25.6	40.5
Observed	3.27	3.56	4.87	5.59	5.93	7.96	8.89	13.2	26.5	42.2	75.5

TABLE I: The expected and observed limits for the NEAT combination

VII. 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

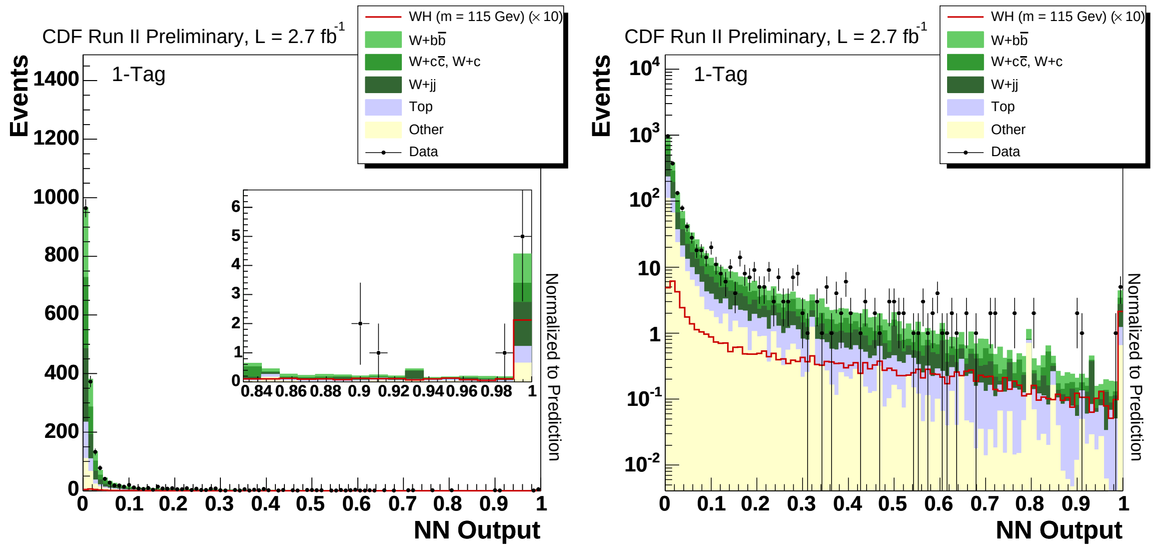


FIG. 4: The NEAT output distributions for signal and background in the one tag bin.

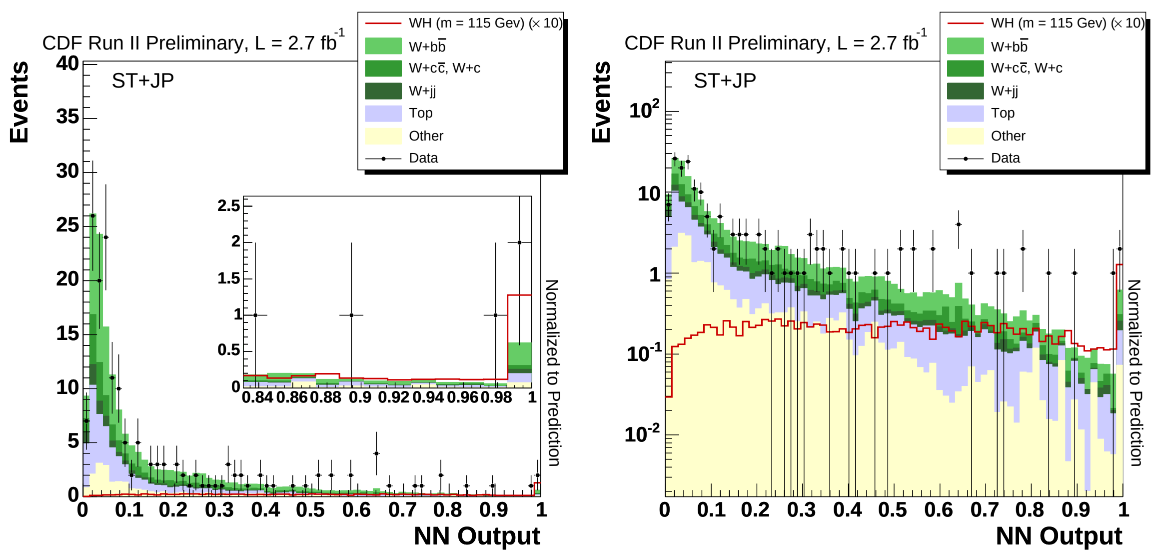


FIG. 5: The NEAT output distributions for signal and background in the ST + JP bin.

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 für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique

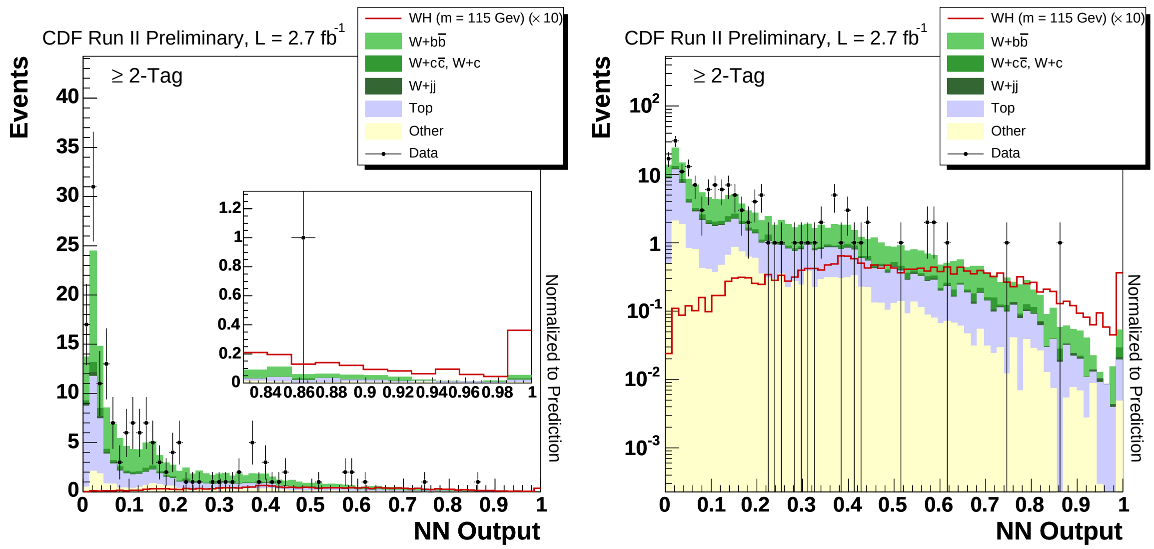


FIG. 6: The NEAT output distributions for signal and background in the ≥ 2 tag bin.

Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

---

[1] P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964).

[2] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[3] R. Barate et al. (LEP Working Group for Higgs boson searches), Phys. Lett. **B565**, 61 (2003), hep-ex/0306033.

[4] J. Alcaraz et al. (LEP) (2007), 0712.0929.

[5] T. Han and S. Willenbrock, Phys. Lett. **B273**, 167 (1991).

[6] D. E. Acosta et al. (CDF Collaboration), Phys. Rev. **D71**, 052003 (2005), hep-ex/0410041.

[7] A. Abulencia et al. (CDF Collaboration), Phys. Rev. Lett. **97**, 081802 (2006), hep-ex/0605124.

[8] J. Slaunwhite et al. (CDF Collaboration), CDF public note 9468 (2008).

[9] A. Hocker et al., PoS **ACAT**, 040 (2007), physics/0703039.

[10] S. Richter (2007), FERMILAB-THESIS-2007-35.

[11] F. Canelli et al. (CDF Collaboration), CDF public note 9463 (2008).

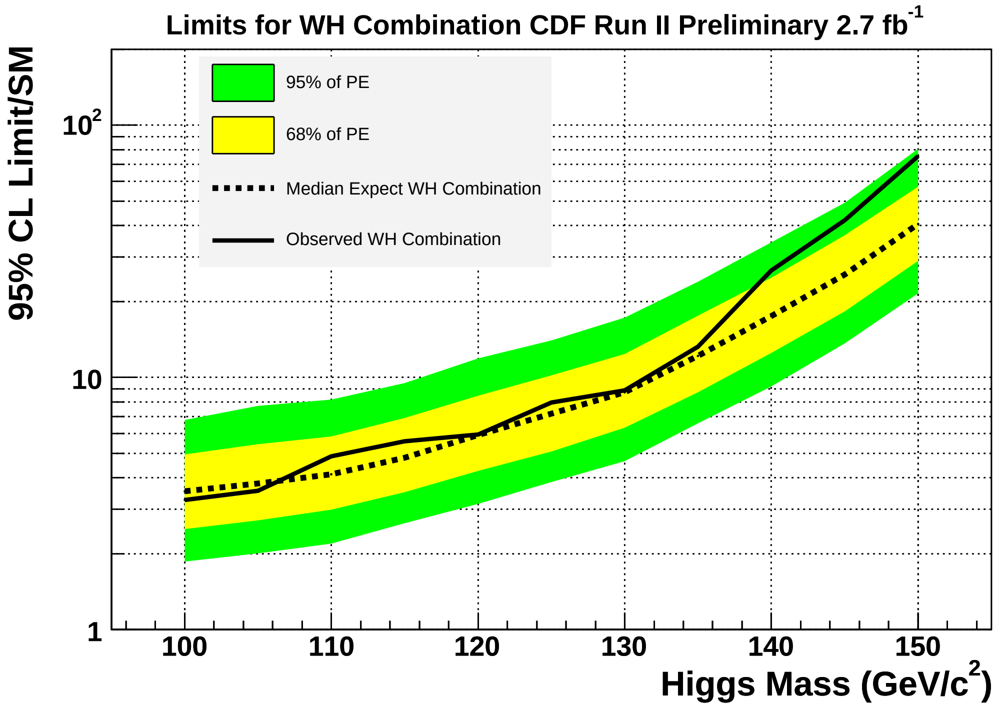


FIG. 7: The expected and observed limits for the combined result using NEAT

[12] T. Aaltonen et al. (CDF Collaboration) (2008), 0809.2581.

[13] K. Lannon et al. (CDF Collaboration), CDF public note 9251 (2008).

[14] K. Stanley and R. Miikkulainen, Evolutionary Compuation **10** (2002).

[15] A. Abulencia et al. (CDF Collaboration), Phys. Rev. **D74**, 072006 (2006), hep-ex/0607035.

[16] T. Aaltonen et al. (CDF Collaboration) (2008), 0807.4652.

[17] S. Whiteson and D. Whiteson (2006), hep-ex/0607012.

[18] T. Junk (CDF Collaboration), CDF public note 8128 (2008).