Improving the search for a Standard Model Higgs boson at the DØ Detector

William Johnson
University of Washington, Seattle
(Dated: 31 July 2013)

One of the available channels for a Standard Model Higgs search at the Tevatron is the production of a Higgs boson in association with a W boson, where the Higgs decays to two b-quarks and the W decays to a lepton and a neutrino. In this paper we provide an introduction to the Higgs boson and how it fits within the Standard Model, and the DØ detector that was used to collect data for this search. Next we describe the analysis process and the development of a pair of optimization tools that were designed to improve the Higgs search capabilities at DØ. Finally, the results of these improvements are presented with prior results as comparison.

I. INTRODUCTION

A. The Higgs Boson and the Standard Model

Originally predicted in the 1960s, the Higgs boson has recently, since the announcement of its tentative discovery in July 2012, become a symbol of the predictive power of particle physics theory. It occupies a very special place in the Standard Model (SM), the framework in which particle physics theory resides. This paper will begin by describing the SM to shed some more light on where the Higgs boson fits into it, followed by summarizing the DØ Experiment's role in the search for the new particle.

1. The Standard Model

The Standard Model describes, in essence, the interactions between matter. In particular, it describes matter of being made up of *quarks* (building blocks of things such as protons and neutrons, which in turn make up atomic nuclei) and *leptons* (a class of particles which includes the electron, responsible for chemical reactions between atoms), which interact with each other by means of three known fundamental forces.[1]

- ***Electromagnetism***—the most ubiquitous in daily life of the three forces, electromagnetism (EM) describes such things as fridge magnets or current traveling through wires, and microwaves, radio, and visible light. It is carried by the *photon*, which makes up what we see as light as well as the "force fields" that stick magnets to refrigerators.
- ***The weak nuclear force***—responsible for radioactive decay, the weak force is carried by three particles: $W^{\pm}$ and Z. These particles are unique in the world of force-carriers in the sense that they are massive (in fact, they are *quite* massive, weighing around 80 and 90 times as much as a proton, respectively). The Higgs boson is intimately related to this observation, as we shall see.
- ***The strong nuclear force***—responsible for holding quarks together in protons, neutrons, and multitudinous other "hadrons" (the name for a thing which is made out of quarks). Its strength is characterized in one of physicists' favorite factoids: the energy that one has to put into a hadron to pull the quarks apart is so great that *new quarks* are produced out of the vacuum to bind with the quarks you were trying to pull apart!

---

[1] Not including gravity, which the Standard Model has yet to describe correctly.

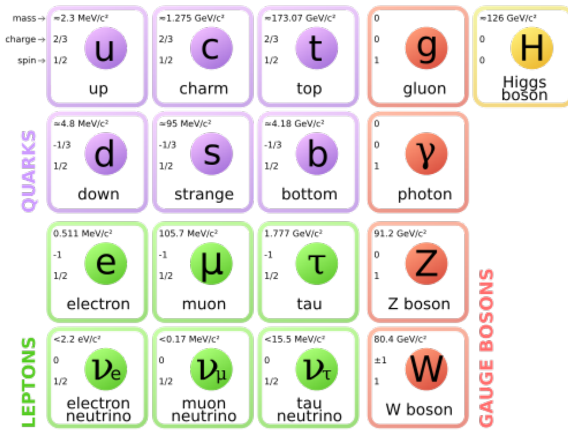

FIG. 1: The particles of the Standard Model.

These pieces are collected in the Figure 1, with the matter particles in green and purple, and the force carriers in red. The important thing distinguishing the modern Standard Model from early quantum mechanics is that physics is framed in terms of *fields*, which is a mathematical construct that takes a value at every point in space. Lots of commonplace concepts are best described as fields—temperature is one example, or wind speed. Particles themselves are perturbations on these fields, like a wave traveling along an otherwise flat lake.

In the late 1960s, Glashow, Salam, and Weinberg discovered that electromagnetism and the weak nuclear force could be described by one theory, which describes the *electroweak* interactions. In fact, at the high temperatures where electromagnetism and the weak force are unified, the W and Z bosons are massless! This is because of *gauge symmetry*, which states that there are certain mathematical changes that can be made without changing the actual physics of a situation. In the case of force-carriers like the W and Z, their masses *must* be zero or else the gauge symmetry is not preserved—we say it is *broken*. Somehow, as the temperature of the universe decreased after the Big Bang, these two forces must have diverged to become the distinct phenomena we observe today, and the weak force carriers must have acquired masses. The Higgs mechanism is the process by which this occurs.

2. The Higgs Mechanism

In order to give masses to the W and Z bosons, and to break the symmetry between electromagnetism and the weak nuclear force, Higgs (and, simultaneously, Anderson, Brout, Englert, Guralnik, Hagen, Kibble, and 't Hooft) postulated a new field to add to the theory. This field, which has come to be known simply as the Higgs field, has four components, three of which are given up (when the temperature lowers beyond a certain critical value) to give masses to the W and Z bosons, thus breaking electroweak symmetry. One component is left over: it shows up as a new particle, the Higgs boson. In addition, there are ways to give masses to the matter particles using the Higgs mechanism (though we must resort to ad-hoc methods to do so—quark and lepton masses do not fall straight out of the theory like the W and Z masses do). Although several of its properties could be predicted by the theory, its mass was a free parameter, requiring us to actually find it experimentally and measure its mass.

B. Higgs Searches at DØ

The DØ detector is, like most modern detectors, made up of several layers surrounding the interaction point. These are, from innermost to outermost: [3]

1. ***Tracking detectors***—the silicon microstrip tracker (SMT) and central scintillating fiber detector (CFT) track the position of charged particles, measuring how their paths bend inside a 2T magnetic field to measure their momentum. The magnetic field is provided by a superconducting solenoid, which separates the tracking layers from the calorimetry layers.

2. ***Preshower detectors***—three thin layers of scintillating strips are used between the trackers and the calorimeters in order to improve energy resolution, which is degraded when the particles

move through the solenoidal magnet, perhaps depositing some energy into the magnet itself rather than the calorimeters where we can measure it. This serves to bolster the usefulness of the real-time event selection (the "trigger" which selects interesting events *as they happen*) as well as later data analysis stages such as tagging jets which come from b quarks [6].

3. ***Calorimeters***—the calorimetry system is made up of a cylindrical portion surrounding the interaction point and two end caps. Each portion is made up of three sections: one finely segmented electromagnetic layer, for catching leptons and photons; and one fine plus one coarse hadronic calorimeter, for protons, pions, and the like. Calorimeters are used to measure the energy of all particles (save neutrinos, which are inferred from the lack of a signal where conservation of energy would otherwise require one).

4. ***Muon system***—The muon system is made of three layers, with drift tubes and scintillation counters making up each layer, and a 1.8 T toroidal magnet between the first and second layers. Because muons interact less frequently than other particles, but still live long enough to escape the detector, this outer system is needed to catch them.

A cross-sectional diagram of the detector is in Figure 2.

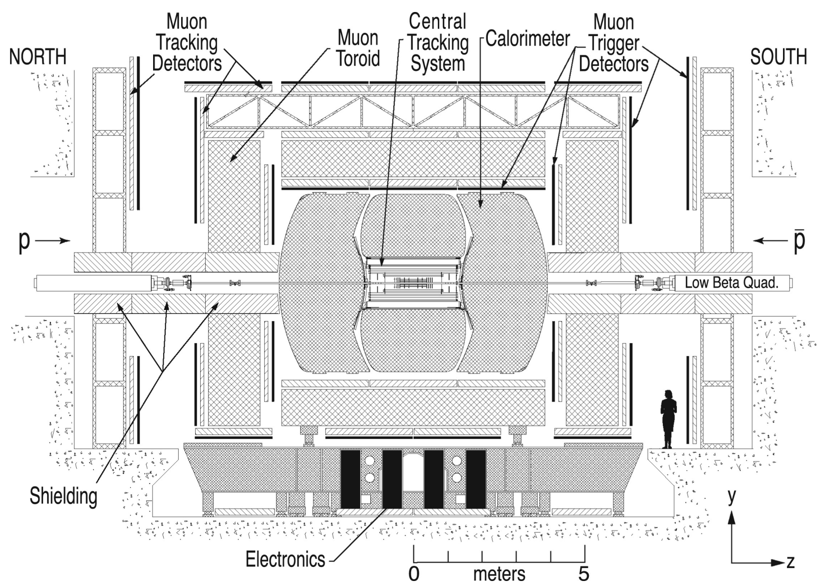

FIG. 2: A cross section of the DØ detector. *Source*: [2]

The detector observes the results of colliding a proton beam and an antiproton beam accelerated by the Tevatron to a center of mass energy of 1.96 TeV. At these energies, the dominant production mode for a Higgs boson is by gluon fusion, followed by a Higgs being produced in association with a vector boson (W or Z). After a Higgs boson is produced, the dominant decay mode for a Higgs mass less than about 135 GeV is into a pair of b quarks; being unstable, each b quark will decay within a few millimeters of where they were produced into a spray of hadrons known as a jet.

Unfortunately, there are a tremendous amount of physics processes that look just like a gluon-fusion Higgs decaying into two b quarks (the so-called *background* events). Hence we eschew this process in favor of the less-common but more analysis-friendly associated production channel. We will be looking in particular for events where a W boson radiates off a Higgs then decays into a lepton (e or μ) and its associated neutrino (ν_e or ν_μ, respectively). This produces a much more distinctive signature than the gluon-fusion process: one lepton signal with two jets tagged as having originated from b quarks (which in turn originated from the same parent particle), and a significant amount of missing energy that was carried away by the neutrino (which we cannot detect at DØ).

Now comes the problem of data analysis. There are many kinematic variables at play: the momentum of the lepton, or the angle between the b quarks, or how "b-like" the quarks looked according to the tagging software. The list is very long, and it is not obvious which ones will provide the best picture of whether an event contains a Higgs boson or not. It may even be that several variables are correlated! A process known as multivariate analysis (MVA) is used to sort this out, which will be described in much more detail in Section II.

MVA has a weakness though, and that is the age-old problem that if you don't give it good data, it can't give you good results! One of the most important variables for our Higgs analysis is the "b-ID value," a number between 0 and 1, which characterizes how much a jet looks like it came from a b quark. The software that analyzes the data to assign these numbers to each jet is always being improved, and depending on which version of the software you are using, a number like 0.5 could mean either a very good or a very poor tag! Hence it is necessary to optimize the analysis for each version of the b-tagger, to determine which choices of b-ID values provide the best discrimination between Higgs events and everything else. This process is described in Section III. Finally, we conclude with an overview of how the improvements described in this picture fit with the other improvements that were designed by other students this summer, with a discussion of the resulting improvement in the DØ Collaboration's Higgs search sensitivity.

II. MVA VARIABLES OPTIMIZATION

A. Multivariate Analysis

Multivariate analysis (MVA) is a powerful class of techniques to use for data analysis when there are a large number of variables that characterize an event [1]. There are several different ways of doing an MVA, and at DØ two variations on the *decision tree* method, which is good at handling correlated variables, are used [5]. A decision tree is pictured in Figure 3, and shows the end result: a given event is sorted into a bin R_i based on the values of the variables x_i. During the first phase of the analysis, the "MVA training," the MVA software uses simulated events which it knows are signal (i.e., a Higgs event) or background and attempts to create a tree which sorts signal events and background events into different regions. In the second phase, when the newly trained tree is used on real data, one would expect that any Higgs events will show up in the appropriate regions.

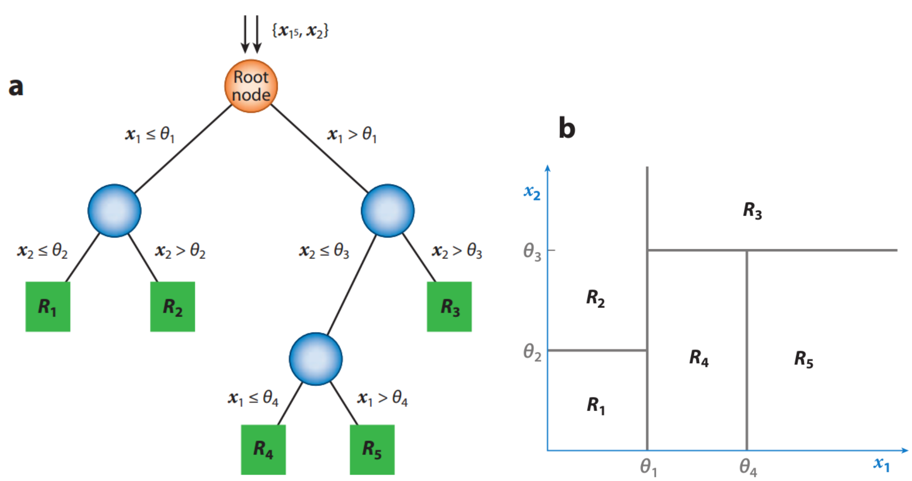

FIG. 3: A generic decision tree, with two variables x_i. (b) shows how the decision tree splits the x_1x_2 plane into distinct regions. *Source: [1], Fig. 5*

- ***Boosted decision tree***—in this process, the training is repeated several times, each time paying special attention to events which are misclassified in the previous iteration.
- ***Random forest***—here many trees are created, each one trained using only a random sample of all the available variables.

B. Choosing Variables

Of course, one can imagine an MVA is complicated enough with just a handful of variables to choose from—but there are dozens available, and we would like to restrict the analysis to only those variables which have the best chance of discriminating between signal and background events. With this in mind, Ben Rabe (another summer student) and I developed code which examines each variable (in particular, it examines the distribution of the variable over all events for signal and background) to see which ones are the most different between background and signal samples. There are several tests which provide a measure of how alike two histograms are:

- ***Kolmogorov-Smirnov test***—examines the shape of each histogram and returns a "p-value" between 0 and 1 which is supposed to represent the probability that the two plots came from the same parent distribution.
- ***Max KS test***—the KS test is based on a cumulative integral of both histograms; inspired by this, the "max KS test" is simply the maximum difference between the two cumulative integrals that is obtained in a given histogram.
- ***Reduced chi-squared test***—contrary to the KS test, which examines overall shape, the chi-squared test examines each bin of the histograms independently and produces a value based on the difference between each bin.
- ***Significance tests***—there are four different methods of measuring signal significance that we use. If S is the number of signal events in a given bin, and B is the number of background events, these are:

  $$S/B \qquad S/\sqrt{B} \qquad S/\sqrt{S+B} \qquad S\log\left(1+S/B\right).$$

  We examine the maximum value that these tests obtain over all bins in the histogram.

The first three mentioned measure how well our simulated data models the real data for that variable; we do not want to use variables for which the simulation looks nothing like what actually happened! The significance tests then give us an idea of whether or not that variable is good at picking Higgs processes out of backgrounds. We do some initial sorting to discard the especially badly modeled variables and the ones that do not discriminate well against background events, then the MVA is run with variables hand-picked from the remainder.

III. b-TAGGER OPTIMIZATION

A. Finding the Best Values

The selection process for b-jets in our analysis takes the form of a simple cut on the highest jet b-ID values. Hence to optimize the selections, the b-ID values were plotted against a specialized significance value:

$$\Sigma_b = \frac{S}{\sqrt{S+B+\sigma_{\text{hf}}^2+\sigma_{\text{lf}}^2}}$$

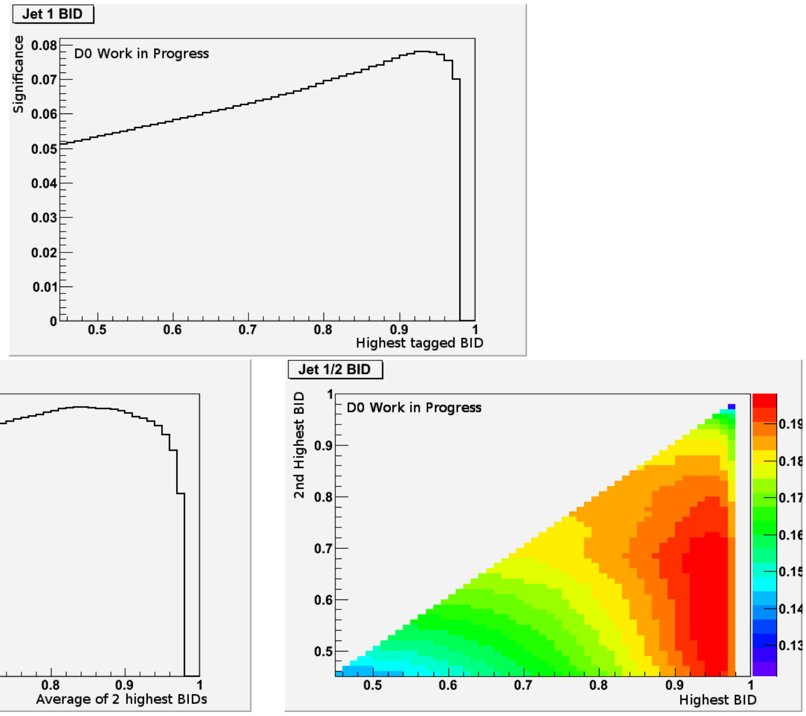

FIG. 4: Three plots of the b-ID value versus the significance Σ_b defined above. In the lower-right plot, the highest b-ID for the event is plotted on the x-axis, while the second-highest is on the y-axis. The colors represent Σ_b, with lowest values in blue and highest values in red.

where $\sigma_{\rm hf}$ and $\sigma_{\rm lf}$ are the systematic uncertainties arising from "heavy flavor" (b and c quarks) and "light flavor" (u, d, and s quarks) events. Thus events with a high systematic uncertainty in the b-ID value will be favored less.

For events with only one b-tagged jet, we simply cut on that value. But for two-tag events, we have a choice of either cutting on the average b-ID value (which is the method used previously) or cutting on each jet separately (which had yet to be tested). The former case is another one-dimensional histogram, but the latter requires a two-dimensional one. All three of these are shown in Figure 4. The best cuts to make are at the maxima of these plots; in fact, for a given data set we use three cuts: the "loose" cut, which is the least restricting, to "medium" and "tight" (the most demanding). More loose cuts allow more events in, so in the event that tight cuts have low statistics we can compare to plots with more events (to the detriment of the signal-to-background ratio). The 2-tag cuts in Table I marked "avg" are cuts on the average of the two highest b-ID values; the ones marked "sep" show the cuts to place on the highest and second-highest b-ID values respectively.

IV. RESULTS & CONCLUSION

The sensitivity of our analysis is calculated by a Confidence Level Limit Evaluator (Collie) [4], which produced a number called the "cross-section scaling factor." Suppose we have S Higgs events in our data; then what this means is that we would need $S \times F$, where F is the scaling factor, events in order to state to 95% confidence level that there is *no* Higgs there. Hence if F is greater than 1, we can't really say anything; if F is 1, then we can exclude the Higgs boson's existence to the 95% confidence level. If F is *less* than one, it is saying that we'd need *less* Higgs events than there actually were in order to exclude its existence—so as F approaches zero, we become more and more

Operating point	Cut value
1-tight	0.92
1-loose	0.45
2-tight (avg)	0.84
2-medium (avg)	0.72
2-loose (avg)	0.45
2-tight (sep)	(0.90, 0.67)
2-medium (sep)	(0.90, 0.45)
2-loose (sep)	(0.45, 0.45)

TABLE I: Optimized cuts to place on BID values.

Channel	PRD	Old	BL*avg	BL*sep	BL*comb
e	6.06	6.28	6.47	6.44	5.70
μ	6.16	6.52	6.28	6.39	5.86
$e+\mu$	4.23	4.42	4.41	4.47	4.03

TABLE II: Collie results after reoptimization. "Channel" is which lepton the W boson decayed into; $e+\mu$ is combined from both channels.

certain that we really are seeing a particle there.

The Collie results from this summer's work are in Tables II. The column labeled "PRD" is a reference to the results that were published in Phys. Rev. D by the DØ Collaboration earlier this year [3], and serve as a benchmark. Between the PRD article and summer, data was reprocessed, which necessitated a reoptimization of the analysis; the "Old" column shows where we were after reprocessing but before reoptimization. "BL*avg" and "BL*sep" are the reoptimized b-ID values using cuts on the average and separate b-ID values respectively. Finally, "BL*comb" is the result when combined with other students' work over the summer.

Altogether we were able to produce a 5% improvement over the PRD result in the space of ten weeks, which is remarkable given that we are adding to over ten years of collected improvements and optimizations! Along with signifcant improvements to usability thanks to automizing the optimzation process and various steps in running the analysis, future improvements and even different analyses will benefit from the work that was completed this summer.

V. ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program. I would like to thank the MVA Team (Stephanie Hamilton, Ben Rabe, Tony Podkowa), mentors (Mike Cooke, Ryuji Yamada), DØ, Fermilab, and the Department of Energy.

REFERENCES

[1] Pushpalatha C. Bhat. Multivariate analysis methods in particle physics. *Annu. Rev. Nucl. Part. Sci.*, 61:281–309, November 2011.

[2] The DØ Collaboration. The D0 silicon microstrip tracker. *Nucl. Ins. Meth. Phys. Res. A*, 634:8–46, November 2010.

[3] The DØ Collaboration. Search for the standard model Higgs boson in $\ell\nu$+jets final states in 9.1 fb^{-1} of pp collisions with the D0 detector. *Phys. Rev. D*, 2013.

[4] Wade Fisher. *Collie: A Confidence Level Limit Evaluator.* Fermilab, 2010. DØ Note 5595.

[5] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. *TMVA 4: Toolkit for Multivariate Analysis with ROOT*, 2009.

[6] Ken Del Signore. Forward and central preshower detectors for the D0 upgrade. In *Workshop on Scintillating Fiber Detectors*, 1997.