Continuation of the Search for the Standard Model Higgs Boson in the $WH \to WWW \to l\nu.jj.jj$ Channel

Anthony Podkowa

August 5, 2010

Abstract

During the summer of 2010, work began on developing intermediary stages of the analysis of the high-mass Standard Model Higgs boson in associated production with a W boson. This channel, $WH \to WWW \to l\nu.jj.jj$, is of particular interest since it has not yet been investigated by any other analysis group. At time of writing, several analysis frameworks have been updated and utilized to produce plots comparing the data samples and Monte Carlo simulations. Continued efforts are being made to develop a customized framework for this analysis and locate candidate events in this channel.

Contents

1 Introduction **3**

2 Materials and Methods **4**

2.1 The $D\oslash$ Detector: 4

2.2 ROOT 7

2.3 C++ 7

2.4 Shell Scripting 7

2.5 Common Analysis Format (CAF) 8

2.6 Analysis Frameworks 8

2.6.1 Data vs. MC Framework 9

2.6.2 Developmental Framework 9

2.6.3 *Starry* Framework 10

3 Results **11**

3.1 Retrieving Signal Monte Carlo from CAB 11

3.2 Processing without a Framework 11

4 Challenges **13**

5 Future Work **16**

6 Acknowledgments **17**

7 Work Cited **18**

1 Introduction

The Higgs boson is perhaps the most sought after particle in the field of High-Energy Physics (HEP) today. The object of Leon Lederman's book, *The God Particle*, the Higgs boson was postulated by Peter Higgs as the particle responsible for mass, and has been the object of many searches at Fermi National Accelerator Laboratory (Fermilab) and at the European Center for Nuclear Research (CERN). Its discovery is key to the verification of the Standard Model (SM), the overarching theory which describes the all the known fundamental particles and their interactions. Though the Standard Model provides much information about the Higgs boson, it provides no information regarding its mass. Therefore, the Higgs mass must be determined experimentally. Recently, combined results from CDF and $D\oslash$ were able to exclude the Higgs mass between 158 to 175 GeV at a 95% confidence level[i] [1]. Standard Model Higgs searches are divided into two main categories: low mass Higgs searches between 115 and 130 GeV, and high mass Higgs searches between 130 and 200 GeV. Depending on the mass of the Higgs, it can decay in a variety of ways. If the Higgs falls in the low mass range, it decays predominantly into bottom quark-antiquark pairs. Conversely, if the Higgs lies in the high mass range, the Higgs will decay primarily into a pair of W bosons. The likelihood of the each type of Higgs decay is called the branching ratio, and is plotted in Figure 1.1.

The Standard Model Higgs can be produced via two methods: gluon fusion, where two gluons collide and produce a Higgs, or through associated production. In associated production, the Higgs is produced alongside either a W or a Z boson. Our channel is concerned with the production of the Higgs in association with a W boson, which has a cross-section around an order of magnitude lower than that of gluon fusion, which can be seen in Figure 1.2. This indicates that there should be more Higgs events from channels involving gluon fusion than those involving associated production. However, gluon fusion channels also feature a higher

[i] As per HEP convention, natural units are utilized throughout this paper. Though mass and momentum units are actually in GeV/c^{2}and GeV/c respectively, the common way of representing these units is to set c=1 to simplify things, leaving the units in GeV.

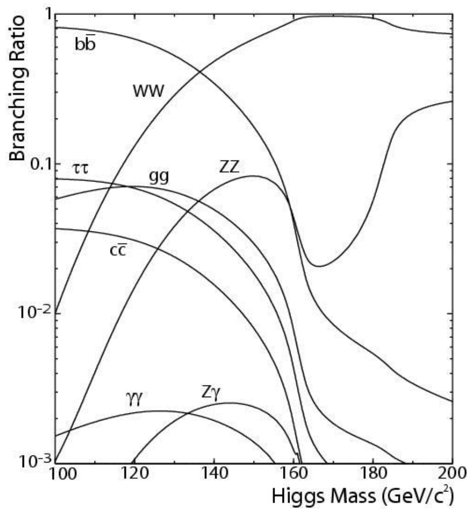

Figure 1.1: Higgs Decay Branching Ratio vs Mass [2]
Retrieved from
http://www-cdf.fnal.gov/physics/exotic/r2a/20050623.lmetbj_wh_tc/higgs_branch.jpg

cross-section for background processes as well, making each event harder to detect [3]. The Feynman diagram of the $WH \to WWW \to l\nu.jj.jj$ channel is shown in Figure 1.3.

The purpose of this paper is to detail the continuing efforts at developing an analysis framework and process the data and Monte Carlo simulations regarding the $WH \to WWW \to l\nu.jj.jj$ channel. Efforts to update existing frameworks to fit our analysis will be discussed, as well as the development of software specific for this analysis.

2 Materials and Methods

2.1 The $D\oslash$ Detector:

The $D\oslash$ detector is a highly sophisticated piece of machinery engineered for the purpose of identifying and tracking particles produced from the proton-antiproton collisions inside the Tevatron. It consists of a central tracking system, a calorimeter system and a muon detector system. The central tracking system is centralized within a 2 T solenoidal field, and consists of two separate tracking subsystems. Closest to the beam collisions is the Silicon Microstrip Tracker (SMT), which consists of four layered barrel silicon detectors and disks

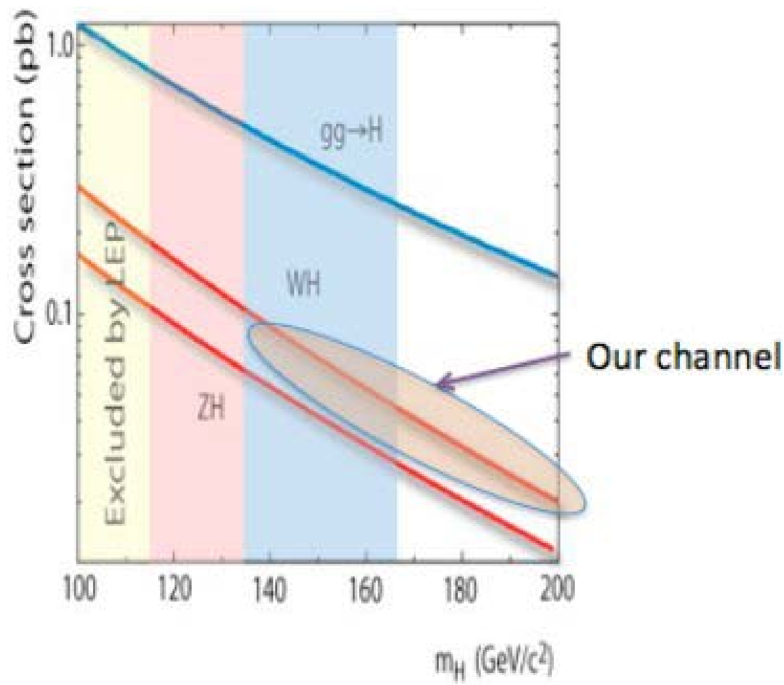

Figure 1.2: Higgs Cross Section vs Mass [3]

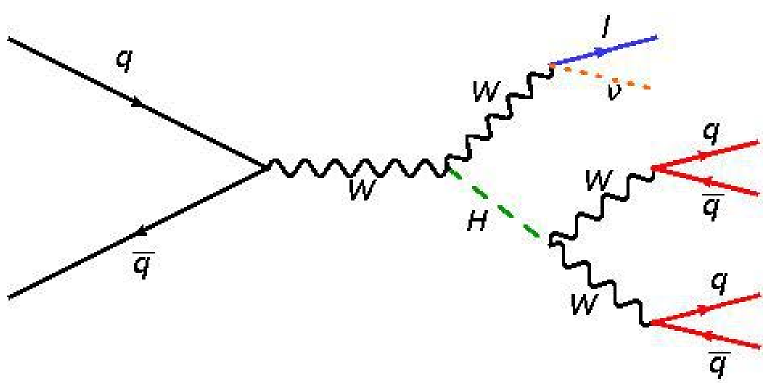

Figure 1.3: Feynman Diagram for $WH \to WWW \to l\nu.jj.jj$
Generated using JaxoDraw (http://jaxodraw.sourceforge.net)

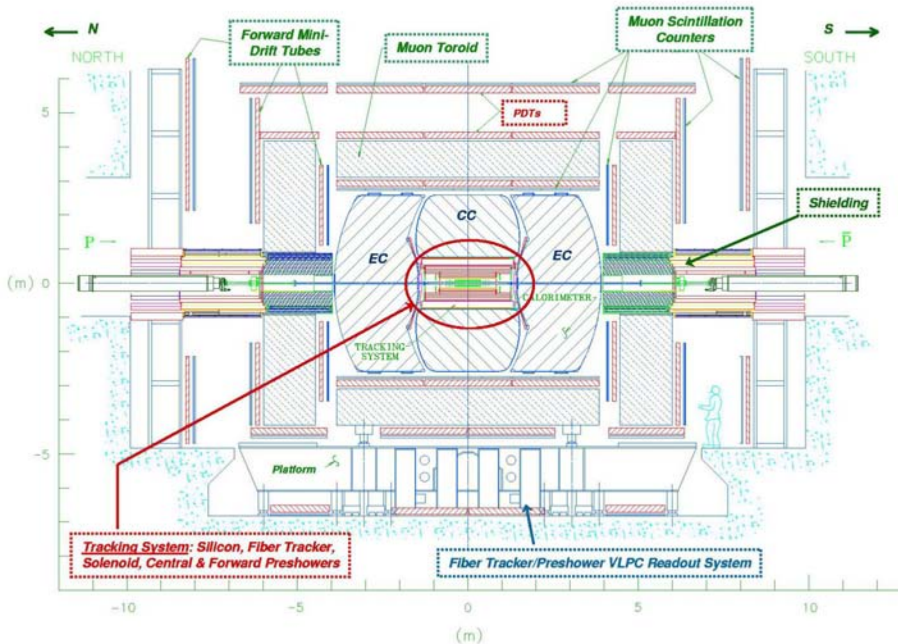

Figure 2.1: The $D\oslash$ Detector

in the central region. In addition, there are large diameter disks positioned in the forward regions of the detector for tracking at high pseudorapidity, η[ii]. Outside of the SMT lies the Central Fiber Tracker (CFT), which is composed of scintillating fibers arranged within eight cylindrical supports wrapped concentrically around the beam pipe. Two doublets of scintillating fibers are overlaid on each cylinder, one parallel to the beam axis (z) and the other at an angle of $\pm 3°$ with respect to z. The liquid argon and uranium calorimeter system consists of three regions: the Central Calorimeter (CC) and the two End Calorimeters. The Central Calorimeter covers the region up to $\eta \approx 1$, while the End Calorimeter covers up to $\eta \approx 4$. The electromagnetic (EM) regions of the calorimeter are located closest to the collisions, followed by layers of fine and coarse hadronic layers outward.

[ii] $\eta = -\ln\left[\tan\frac{\theta}{2}\right]$ where θ is the polar angle as measured from the beam axis.

2.2 ROOT

Developed and maintained by the European Center for Nuclear Research (CERN), ROOT is an object-orienteslated framework utilized by high-energy physicists worldwide. The core of the ROOT framework is based on the development of TTrees, classes that store histograms within binary ROOT files. TTrees and their $D\oslash$-specific derivatives, TMBTrees, provide the basic structure of our analysis output. Other secondary classes such as the TBranch, TClonesArray, and the TLeaf classes contribute to the organizational structure of ROOT files. TTrees for high-energy physicists usually contain information regarding the kinematic information of particles, such as the momentum components (px, py, pz) , the detector angles (ϕ and θ), and particle identification information.

2.3 C++

C++ is an object-oriented programming language which is utilized by ROOT as a basic interface. ROOT utilizes the C interpreter (CINT) to interactively process C and C++ statements from compiled functions or user input. C++ is also regularly used as a basis for Common Analysis Format (CAF) packages, which need to be compiled before submitting a job to the $D\oslash$ Central Analysis Backend (CAB) servers to retrieve data or Monte Carlo samples, to ensure that the samples are retrieved with the right specifications.

2.4 Shell Scripting

Remote utilization of the servers on the $D\oslash$ Clued0 Linux cluster requires the usage of a shell, a simple interactive programming environment which processes commands and returns output. The default shell for servers in the Clued0 Cluster is tcsh (TC SHell), a shell which attempts to mimic C style syntax. Though running the tcsh interactively can be extremely useful, it is often more productive to write scripts to automate repetitive tasks that require several shell commands. Shell scripts are often used in $D\oslash$ to aid in the setup of CAF packages as well as job submission to the $D\oslash$ CAB servers. During my time here at Fermilab, I wrote several scripts in both tcsh and bash (Bourne Again SHell) to aid in a variety of tasks, including a basic output-logging script, a CAB job checker, and assorted CAF package installation and maintenance tools.

2.5 Common Analysis Format (CAF)

$D\oslash$ utilizes the Common Analysis Format to process the data stored in the TMBTrees. CAF is composed of compiled packages called Processors and configuration files. The Processors are groups of packages, usually coded in C++ which contain useful functions to perform over data or Monte Carlo samples. These functions commonly include, but are not limited to, smear corrections in data and Monte Carlo, flavor-tagging of jets, and producing new TTrees and TMBTrees. Since a multitude of analyses need to perform similar functions, most packages are hosted on a CVS server, which are accessible to all users at the $D\oslash$ collaboration. Any given analysis at $D\oslash$ will utilize a variety of different processors, collected from different packages.

Though all of the CAF processing can be done on a single system through ROOT, it is often more useful to send it to the $D\oslash$ Central Analysis Backend system. The CAB servers decrease processing time by splitting an analysis job into several parts, and parallel processing all parts simultaneously. This effectively shortens the processing time to a fraction of what it could have been running it on a single CPU.

2.6 Analysis Frameworks

For our analysis, we had the opportunity to work with three different analysis frameworks. All of these packages were derived from the wh_cafe package, which is in use by the WH Group for their analysis of the $WH \rightarrow l\nu bb$ channel at $D\oslash$. A large portion of my time at Fermilab was spent editing these frameworks to fit our analysis. Since the WH Group is mainly concerned with two-jet channels and low-mass Higgs searches, it remained a constant need to modify these frameworks for our four-jet, high-mass Higgs search.

2.6.1 Data vs. MC Framework

This framework was the framework used by my predecessors Vesselin Velev and Zachary Hynes. It was derived from the publication version of wh_cafe in use by the WH Group for their analysis of the $WH \to l\nu bb$ channel, and was heavily modified by Zach and Vesselin to enable them to produce four-jet, high mass plots.

The core analysis was processed by running a tcsh script called runnewdatavsmc.csh. This script parsed a file called version.txt for several different variables related to the analysis, and passed them as parameters to the ROOT script Run_WH_Analysis.C. This script ran the bulk of the analysis and generated the PostScript files of the plots we needed. These files were then compiled into a pdf file using pdflatex, a Linux command line tool.

Though my predecessors worked hard at editing the framework to fit our channel, it remained very confusing to edit and utilize this framework for our research. Vast amounts of lines were commented out of the code without any hints as to the reason, and the framework was extremely fragile and disorganized. The runnewdatavsmc.csh script created and depended upon many text and LaTeX files that crowded the directory, making it very hard to locate specific files. At time of writing, there were nearly 300 files in the analysis directory, much more than could fit in one terminal window at one time. Any attempts to reorganize these files would often end up breaking the code.

2.6.2 Developmental Framework

At the suggestion of Dr. Michael Cooke, we began to work with a newer, developmental version of wh_cafe. Though the framework was slightly buggy and had no four jet support, it was much more organized and implemented a better PDF output file, complete with a linked Table of Contents. Though there were many source and header files in the framework, each file only contained the relevant classes for each stage of the analysis. In addition, each class only contained specific functions which compartmentalized the framework into smaller sections, making it much easier to debug.

It took some time to produce working plots, since the framework needed to be updated for our four jet analysis. We first needed to adjust the configuration files for the framework to skim for four jet samples. Unfortunately, since the framework was not designed to work with more than 2-3 jets, we ended up needing to make several adjustments to many files, including WH_SampleManager.C (sample selection source file), WH_PlotManager.C (plot selection source file), WH_ResultsManager.h (header file for TTree processing), WH_ResultsManager.C (corresponding source file), and WH_OutputManager.C (plot generation source file). Though we were able to get the framework working for our four jets, we had some difficulty acquiring full samples to run over. Dr. Cooke only had partial samples which he was using for testing. As a result, the Data/Monte Carlo agreement in this framework was fairly poor.

2.6.3 *Starry* Framework

Since we did not have any samples to run over in the developmental framework, Dr. Cooke suggested that we migrate to an even newer framework in development. This framework, tagged *starry* in the $D\oslash$ CVS server, was currently being developed by Dr. Bjoern Penning and Dr. Sebastien Greder of the WH Group. Dr. Penning had already retrieved a working set of samples from CAB, which made it much easier to get the framework running. Since we had already updated the developmental framework for four jets, we were able to update it to four jets much faster. Once we had the framework running, we were able to generate a vast number of plots, though many of them were not applicable to this analysis.

3 Results

3.1 Retrieving Signal Monte Carlo from CAB

Though Dr. Penning was able to provide us with a working set of data and Monte Carlo samples, it was not enough to run our analysis. Since Dr. Penning was working on low mass Higgs analysis, we were missing our high mass signal Monte Carlo. In order to retrieve these samples, we began to work with a tarball of the Dr. Penning's working area. We extracted the tarball into a directory, and proceeded to compile and submit CAB jobs. This proved to be a long and arduous process, for whenever we submitted the jobs to CAB, they tended to fail on the first error message, which added to the confusion. Often they would fail with an undefined symbol error, which usually meant that something had gone wrong in the compilation process, and that the packages had to be edited and recompiled.

Eventually, we were able to retrieve the necessary 160 and 180 GeV Higgs Mass Monte Carlo for electron and muon samples. The release utilized in the successful submission was based on the VJets 5.4 Release [4], along with several other packages. A complete list of CAF packages is listed in Table 3.1.

3.2 Processing without a Framework

At the suggestion of Dr. Penning, we began processing data without a framework. There were several reasons for doing this. First, the *starry* framework is still in active development. Between the changes being made by Dr. Penning and Dr. Greder, it could change vastly in the CVS server without any notice. Therefore, any user utilizing our tutorials on setting up areas would likely be getting different code than we might expect. Second, wh_cafe embedded many things into it which were specific to the $WH \rightarrow l\nu bb$ channel. One such attribute, b-tagging (attributing a jet to a b quark), was integrated firmly into the framework, a feature which we did not need for our analysis and would often interfere with our attempts at research. Third, in processing outside of the framework, we are able to investigate different

Table 3.1: CAF Packages Utilized in CAB Submission

Package	Version/Tag
beamposition	v2010-05-26
btags_cert	v09-00-00
btags_cert_caf	v00-09-03
caf_dq	p21-br-04
caf_eff_utils	p21-br-21
cafe	p21-br-38
cafe_sam	p21-br-06
caf_mc_util	p21-br-157
caf_trigger	p21-br-88
caf_util	p21-br-132
d0root_analysis	v00-09-91
d0root_jlip	v00-02-02
d0root_mva_tagger	v00-00-08
d0root_nnbtag	v00-01-02
d0root_sltnn	v00-00-04
d0root_tmbtree	p21-br-08
dq_defs	v2010-03-11
dq_util	p21-br-05
eff_utils	p21-br-29
emid_cuts	p21-br-25
emid_eff	v09-00-01
EMresolution_cafe	v00-07-02
jetid_eff	v04-01-01
lumi_profiles	v2010-07-12
muid_eff	v05-02-00
tauid_eff	v01-00-00
tmb_tree	p21-br-71
trackcaljet_tools*	v00-00-01
vjets_cafe*	v05-04-00
caf_pdfreweight*	v00-00-02
wh_cafe*	starry
fake_trk_discrim*	v00-01-11
jet_resolution_util*	v00-00-00

* Denotes non-VJets package. Version numbers are taken from [pkg]/VERSION and are not necessarily accurate.

variables which can later be imported into a framework, such as the *starry* framework, once we understand techniques and structure behind them.

To resolve these issues, Dr. Penning wrote a short ROOT script, readtrees.cpp, which would read in the WH_Trees, process the variables we wished to work with, and output the PostScript files we needed. This script was extremely simple, and compatible with the wh_cafe samples we already had. Working with this script has provided us with the ability to constructively develop key features of our analysis. Though we have only been working with this script for less than a week at time of writing, we have been able to successfully plot several mass variables from our signal Monte Carlo.

Using the readtrees script, we have successfully generated 2-jet mass combinations in our plots. These combinations are important because they represent the possible ways that the W boson could have decayed. Since we can only detect the final state of the channel, we have to infer the intermediary steps of the process. As can be seen in Figure 1.3, the W boson can either decay into a lepton-neutrino pair, or a quark-antiquark pair. However, since the jets are indistinguishable, it is difficult to determine which jet pairings resulted from one W boson decay. We began by generating the mass by adding the two jet four-vectors (defined as WHJet in code) and using the inherited TLorentzVector mass function on the sum. This function works by taking the energy and momentum components from the four-vector sum and using the extended form of Einstein's equation

$$E^2 = (mc^2)^2 + (pc)^2 \tag{3.1}$$

to derive the mass[iii]. This should yield a mass of two jet system near the expected W mass of 80.398±0.025 GeV [5]. Figure 3.1 shows one such 2-jet combination.

The 4 jet invariant mass was calculated similarly to the 2-jet combinations. The four WHJet objects were summed together and then the mass function was called. According to what our channel expects to see, the four jet invariant mass should be equivalent to the Higgs mass of the Monte Carlo sample. However, since we have not begun to constrain events in the sample to where only the Higgs independent W boson decays leptonically, events plotted in Figure 3.2 are only accurate about one third of the time, leading to the mass peak discrepancy.

4 Challenges

It has been an exciting period of discovery this summer. Though I had had some prior experience working with $D\oslash$ and the associated tools, I had never before been asked to work on this level on analysis. This summer, I had the opportunity to play a more central role towards the development of a new analysis, which brought with it a multitude of new challenges that I had never faced before. My prior experience in the ZZ diboson production channel had been during it's final stages, when the software had been much more refined

[iii]The reader should note that the mass derived in a nonlinear system, and that simply summing the two jet masses would not yield the correct mass of the two jet system.

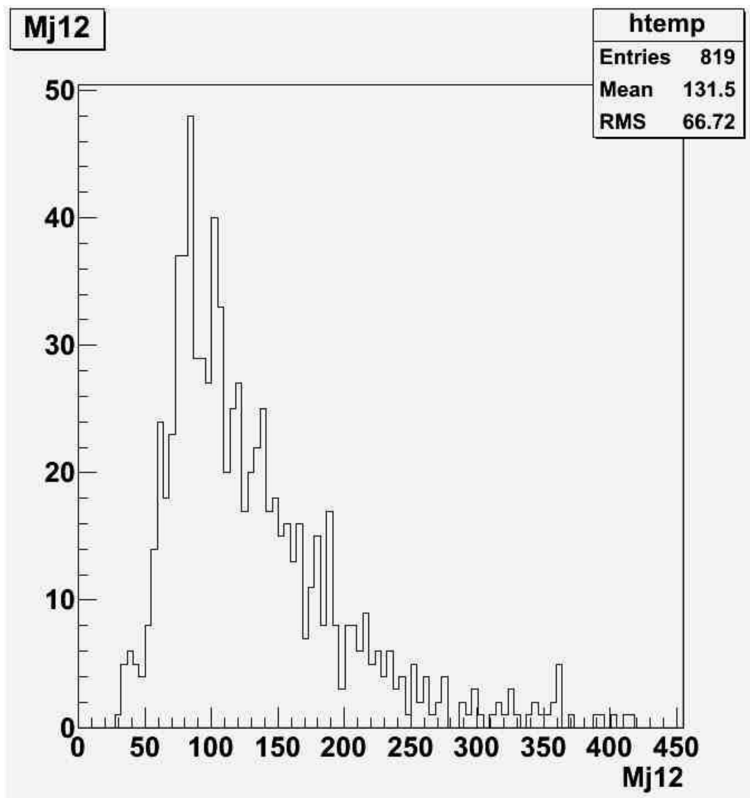

Figure 3.1: Invariant Mass of Jets 1 and 2 in Electron 180 GeV Higgs Mass Signal Monte Carlo.
This graph shows the mass of the combined 2-jet system. Jets 1 and 2 correspond to the respective jets with the highest and second highest transverse momentum (p_t) in each event. The mass in GeV is plotted on the x axis and the number of events is plotted on the y axis. One important feature to note is where the peak is located. This peak should lie at around the mass of the W boson which is 80.398$\pm$.025 GeV [5]. In this plot, the peak lies around 82 GeV without corrections, indicating good correspondence with that prediction.

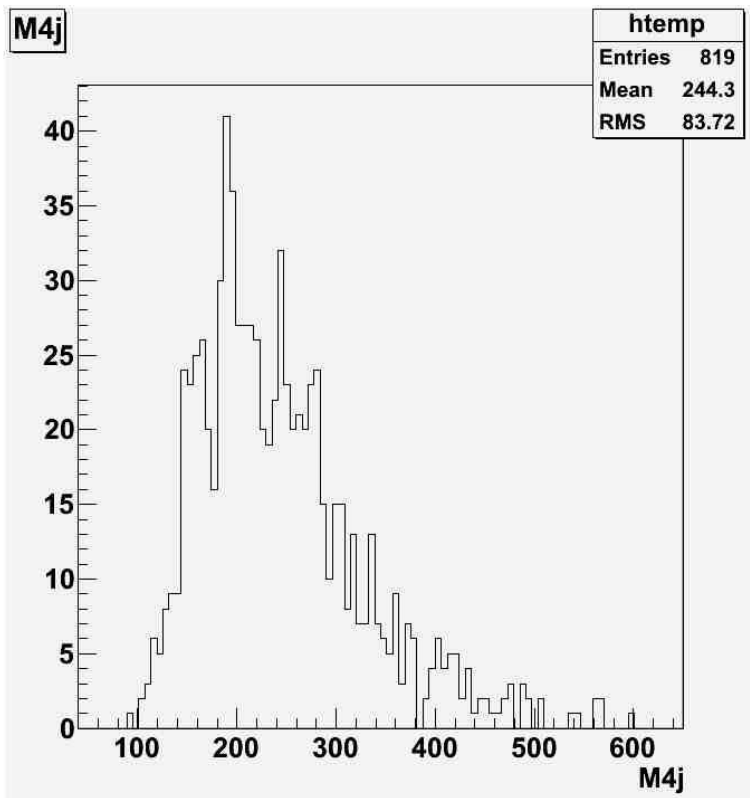

Figure 3.2: 4-Jet Invariant Mass in Electron 180 GeV Higgs Mass Signal Monte Carlo

state.

It took a while for me to learn everything that I thought I once knew about the $D\oslash$ analysis tools. Over the course of the summer, I gradually began to understand the details of the processes that I took for granted. I had the opportunity to learn about CAF and how the CAB submissions actually worked, rather than having it work from a prewritten shell frontend. I learned to utilize many new shell tools, like nm, c++filt, and GNU Screen, and learned some new tricks with other tools I was already familiar with, like grep, find, and vim. As I grew intellectually, little pieces began to fit together, gradually forming the big picture. As I attended lectures, I became more familiar with the terminology and the physics behind the research, and slowly evolved into a competent researcher within the $D\oslash$ Collaboration.

5 Future Work

Since we are just beginning to write our own software from scratch, there is a wealth of future work to be done. Once we are able to reconstruct the W bosons from the $l\nu.jj.jj$ tracks, we can work towards reconstructing the Higgs boson from the W bosons. We also will need to work on isolating the semi-leptonic decaying W boson from the Higgs signal. Hopefully, using the wh_cafe frameworks as guides, we should be able to complete a fully functional framework for our analysis within a year's time.

Following this internship, I hope to continue collaborating with $D\oslash$ remotely, either by working on the $WH \rightarrow WWW \rightarrow l\nu.jj.jj$ channel and/or the ZZ diboson channel. Nevertheless, I hope to remain in contact with my supervisor, Dr. Ryuji Yamada as well as my colleagues from the Illinois Mathematics and Science Academy (IMSA), Wes Bradley and Alex Abbinante throughout the year. This illuminating experience has given me much insight into the world of High Energy Physics, and I hope to continue to grow and contribute to the HEP community in the years to come.

6 Acknowledgments

During my time working for the Summer Internships in Science and Technology program, I have had the privilege of working with Dr. Ryuji Yamada, and my colleagues from IMSA, Wes Bradley and Alex Abbinante. This analysis continues to receive much support from the $D\oslash$ Collaboration. In particular, I would like to recognize Dr. Michael Cooke and Dr. Bjoern Penning for their continued patience and guidance this summer. I would also like to thank Dr. Sebastien Greder, Dr. Joseph Haley, and Dr. Yuji Enari for their contributions to this project. Lastly, I would like to thank Dr. Ryan Hooper, for introducing me to High Energy Physics and providing me with support and encouragement this summer.

7 Work Cited

References

[1] Tevatron New-Phenomena and Higgs Working Group, "Combined CDF and $D\oslash$ Upper Limits on Standard Model Higgs-Boson Production with up to 6.7 fb^{-1} of Data." FERMILAB-CONF-10-257-E (2010)

[2] CDF, "Search for New Particles Decaying into bb and Produced in Association with W Boson." [Online] Available: http://www-cdf.fnal.gov. [Accessed July 25 2010].

[3] Z. Hynes, "Search for the Standard Model, High-Mass Higgs Boson in the $WH \rightarrow WWW \rightarrow l\nu.jj.jj$ Channel," August 2009. [Online]. Available: SIST Presentations 2009, http://sist.fnal.gov. [Accessed July 25 2010].

[4] V. Shary, "vjets_cafe for Summer 2010 v5.4," in *The $D\oslash$ Wiki.* July 2010. [Online]. Available: https://plone4.fnal.gov/P1/D0Wiki/physics/VplusJets/CAFtools/vjets_cafe_v5.4 [Accessed July 26, 2010].

[5] Particle Data Group, "Particle Physics Booklet," July 2008, [Booklet]. Available: http://pdg.lbl.gov/pdgmail. pp. 8.