Observation and Examination of a Higgs Boson in the ZZ to Four Lepton Channel

by

Duncan Ralph

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2014

Signature redacted

Author .....
Department of Physics
January 17, 2014

Signature redacted

Certified by .....
Markus Klute
Assistant Professor
Thesis Supervisor

Signature redacted

Accepted by .....
Krishna Rajagopal
Associate Department Head for Education

Observation and Examination of a Higgs Boson in the ZZ to Four Lepton Channel

by

Duncan Ralph

Submitted to the Department of Physics
on January 17, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

Results are presented on 5.1 fb^{-1} of proton-proton collisions at a center of mass energy of 7 TeV and 19.7 fb^{-1} at 8 TeV in a search for decays of the Standard Model (SM) Higgs boson to two Z bosons to four charged leptons. We observe a single excess above background expectation with a significance of 7.4 standard deviations at a mass of 125.6 $\pm$0.4(stat) $\pm$0.2(syst) GeV. This excess has a signal strength parameter $\mu = \sigma/\sigma_{SM}$ equal to $1.0^{+0.3}_{-0.2}(\text{stat})^{+0.1}_{-0.1}(\text{syst})$. We examine the transverse momentum spectrum of the new particle and find it to be consistent with the SM expectation, and also determine that the Standard Model $J^{PC} = 0^{++}$ is favored over the plausible alternative hypotheses.

Thesis Supervisor: Markus Klute
Title: Assistant Professor

Contents

Chapter 1

Introduction

At the finest detail we can probe, our understanding of the structure of our surroundings is codified into rules for the interaction of a set of mathematical objects which we typically imagine as particles. In this scheme matter is composed of particles which obey Fermi-Dirac statistics [1, 2], while matter's interactions are mediated by particles which obey the statistics described by Bose in [3]. These two types of particle are thus called *fermions* and *bosons*.

Fermions occur in pairs with an *up-type* and a *down-type* member. For reasons which remain obscure these pairs, or generations, have been duplicated three times, each time changing only the mass. The pairs also come in two varieties depending on the details of their interactions: the *quarks* and the *leptons*.

There are three classes of interaction for which there exists experimental justification for treatment within this framework. The electromagnetic force creates most of our daily experience and is carried by the massless photon. The strong force, mediated by the gluon, is the source of almost all mass. Nuclear decay is caused by the weak nuclear force, which, though less obvious in everyday life, has a structure of surpassing particularity. Its complications are hinted at by its triplet of gauge bosons: the $W^{\pm}$ and the Z. While gravitation is quite familiar, its only established theory is manifestly macroscopic.

The structure outlined above is referred to as the Standard Model of particle physics [4, 5, 6], and is summarized in Figure 1-1. It has both explained and predicted

an extraordinary variety of experimental results. However, as just described it is incomplete: in order to apply to this world it must be modified such as to destroy the symmetries which give rise to the electromagnetic and weak bosons. The simplest means of accomplishing this is typically called the Higgs mechanism and arose from a spasm of coincident papers published in 1964 [7, 8, 9, 10, 11, 12]. As one would hope for any theory which aspires to describe reality, the Higgs mechanism results in experimental predictions in the form of an additional particle not included in Figure 1-1: the Higgs boson. The Higgs was predicted to be, unlike any previously observed fundamental particle, a scalar, and the search for its characteristic signatures eventually came to be seen as the most urgent outstanding task in particle physics. This search commenced with no small measure of indolence, with a considerable uncertainty as to its mass and couplings leading one theoretical eminence to conclude that "we do not want to encourage big experimental searches for the Higgs boson, but we do feel that people doing experiments vulnerable to the Higgs boson should know how it may turn up" [13]. Upon instantiation of the Large Electron Positron collider, however, the pace quickened, and by the time of the perhaps untimely demise of LEP II the particle's occult dwellings had shrunk markedly. The search continued at the Tevatron, with not entirely insubstantial new constraints, until the inception of run I of the Large Hadron Collider (LHC). The Higgs finally succumbed to experimental entreaty after two and a half years of LHC data accumulation (Figure 1-2) with a carefully orchestrated joint announcement by the CMS and ATLAS collaborations on the fourth of July, 2012 [14, 15]. This wondrously historical episode marked the culmination of decades of truly extraordinary effort by generations of physicists and dazzled the world at large with its singular import. Within these pages, we present an account of the analysis in the decay channel to two Z bosons to four charged leptons, whose exquisite sensitivity contributed immeasurably to this momentous occurrence.

We first introduce the theoretical background to electroweak symmetry breaking in Chapter 2, and follow this with a description of the Large Hadron Collider and the CMS detector in Chapters 3 and 4. We next give an exposition of the methods used to interpret the raw detector output in terms of elementary particles in Chapter 6, and

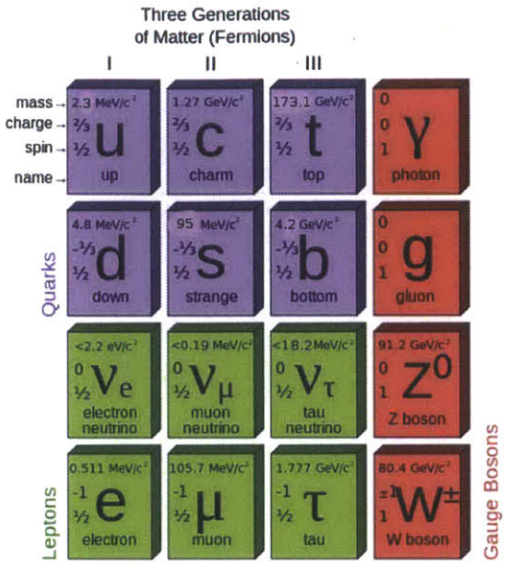

Figure 1-1: Matter and interactions in the Standard Model of particle physics. Matter is composed of the fermions: quarks (purple) and leptons (green), while interactions are mediated by the gauge bosons (red).

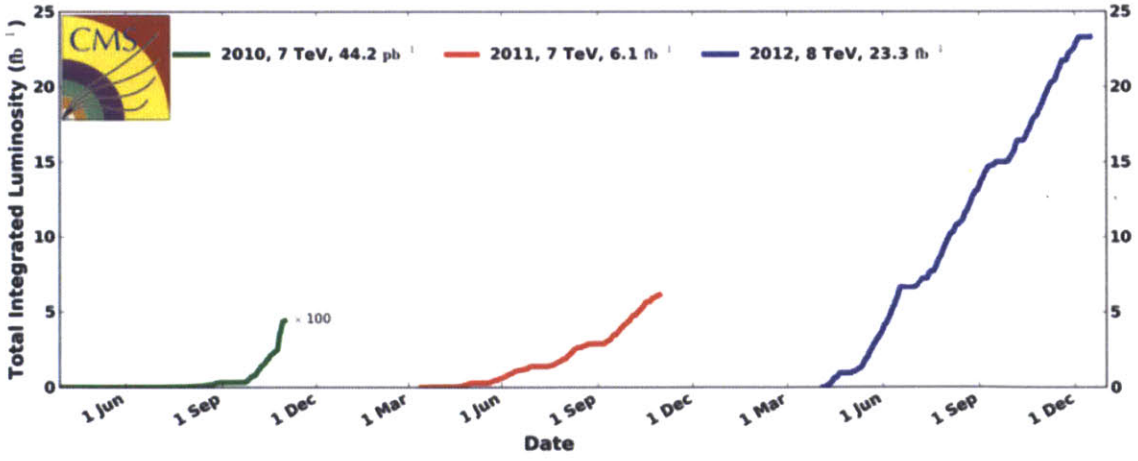

Figure 1-2: Historical view of the luminosity delivered to the CMS detector during the years 2010 through 2012.

of the event selection which we use in order to achieve a large signal to background ratio in Chapter 7. Chapter 8 then describes the models which are used for the signal and background processes. We then proceed to analyze the data: Chapter 9 introduces a machine learning technique to improve signal discrimination, Chapter 10 gives a statistical interpretation of the observed events, Chapter 11 checks the new particle's p_{T} distribution against SM expectation, and Chapter 12 investigates its spin and parity.

Chapter 2

Theoretical Background

2.1 The Higgs Mechanism

2.1.1 Intact Symmetries

The electroweak interaction [16, 17, 18] represents two of the four forces which are in evidence in the universe. Its properties flow naturally from the combination of the underlying $SU(2)_L \otimes U(1)_Y$ gauge symmetry with the experimental existence of left-handed particle doublets and right-handed singlets [1]

$$\mathsf{L} = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L \qquad \mathsf{R} = \ell_R. \tag{2.1}$$

From the $SU(2)_L$ symmetry we acquire a triplet of gauge fields $\vec{B}_\mu$ with attendant field strength tensor

$$F^\ell_{\mu\nu} = \partial_\nu B^\ell_\mu - \partial_\mu B^\ell_\nu + g\epsilon_{jk\ell} B^j_\mu B^k_\nu \tag{2.2}$$

and coupling g. The $U(1)_Y$ meanwhile contributes the single $\mathcal{A}_\mu$ with field strength

$$f_{\mu\nu} = \partial_\nu \mathcal{A}_\mu - \partial_\mu \mathcal{A}_\nu \tag{2.3}$$

[1] While we proceed to draw a cartoon with one generation of lepton, note that both for consistency and to match experiment the theory must be expanded to three generations of leptons and three generations of quarks.

and coupling g'. We are then able to write out the electroweak Lagrangian as

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}F^{\ell}_{\mu\nu}F^{\ell\mu\nu} \\ &-\frac{1}{4}f_{\mu\nu}f^{\mu\nu} \\ &+\bar{\mathsf{R}}i\gamma^{\mu}(\partial_{\mu}+i\frac{g'}{2}\mathcal{A}_{\mu}Y)\mathsf{R} \\ &+\bar{\mathsf{L}}i\gamma^{\mu}(\partial_{\mu}+i\frac{g'}{2}\mathcal{A}_{\mu}Y+i\frac{g}{2}\vec{\tau}\cdot\vec{B}_{\mu})\mathsf{L}\end{aligned} \tag{2.4}$$

where Y is the weak hypercharge, which is -1 for the doublet and -2 for the singlet. We can immediately read off many salient features of the interaction. For example the weak isovector sector is characterized by gauge boson self-interaction and non-participation of fermions which transform in the right-handed representation of the Lorentz group.

The unfortunate circumstance is that the Lagrangian as written clashes violently with experiment. We have made all gauge bosons massless, as required in order to preserve gauge invariance in any theory based on $SU(N)$ [16, 17]. However, in place of these four massless gauge bosons, experimentally there exists only one, while the other three are massive. In addition, simple mass terms for the fermion fields would violate gauge invariance because their left- and right-handed components have different quantum numbers. In reality, of course, there exist massive fermions.

2.1.2 Electroweak Symmetry Breaking

This situation may be remedied by breaking the underlying symmetries using what is frequently called the Higgs mechanism. Using a new complex scalar doublet

$$\phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix} \tag{2.5}$$

we add to the Lagrangian three new gauge-invariant terms

$$(\mathcal{D}^{\mu}\phi)^{\dagger}(\mathcal{D}_{\mu}\phi) - V(\phi^{\dagger}\phi) - \zeta_{\ell}\left[\bar{\mathsf{R}}(\phi^{\dagger}\mathsf{L}) + (\bar{\mathsf{L}}\phi)\mathsf{R}\right]. \tag{2.6}$$

The first term, by virtue of its gauge-covariant derivative

$$\mathcal{D}_\mu = \partial_\mu + i\frac{g'}{2}\mathcal{A}_\mu Y + i\frac{g}{2}\vec{\tau}\cdot\vec{B}_\mu \tag{2.7}$$

specifies the interaction of the scalars with the gauge fields in the original theory. The second term, meanwhile, tells us the self-interaction of the scalars. The potential, for which we use

$$V(x) = \mu^2 x + |\lambda| x^2, \tag{2.8}$$

is at the heart of the matter. With μ^2 negative, the potential has a ring of stable minima away from the origin. This is illustrated in Fig 2-1. Thus the lowest energy state, the vacuum, will have non-zero expectation value in ϕ. The angular location of this minimum, however, is entirely arbitrary: one may imagine that we roll off the summit in a direction chosen at random. Once at rest in the depression, angular invariance is lost and an initial symmetry has been spontaneously broken.

Finally, the third term in equation 2.6 is a Yukawa coupling which is added separately in order to give mass to the fermions.

To see the effects of this scalar field, we for simplicity choose a minimum at

$$\langle\phi\rangle_0 = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \tag{2.9}$$

where $v \approx 246$ GeV. We then expand around this equilibrium in new fields η and ξ which have been obtained by translation of the components of ϕ. Among the terms in this rewritten Lagrangian are those revealing η to be a scalar boson with mass equal to $-2\mu^2$

$$\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) + 2\mu^2\eta^2 \tag{2.10}$$

and ξ to be a massless Goldstone boson

$$\frac{1}{2}(\partial_\mu\xi)(\partial^\mu\xi). \tag{2.11}$$

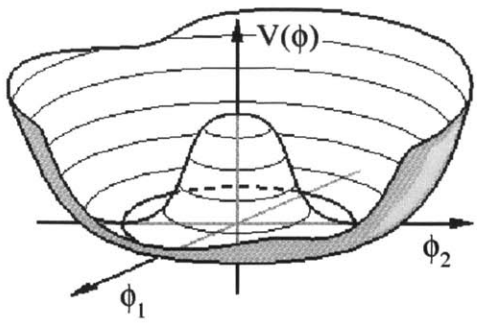

Figure 2-1: The Higgs potential in ϕ space. Note that the lowest energy state (the vacuum) is not trivial, i.e. ϕ has a non-zero vacuum expectation value.

We also find that, as we originally required, the gauge bosons have acquired mass terms

$$\frac{q^2\mu^2}{4|\lambda|}A_\mu A^\mu. \tag{2.12}$$

There are also, however, terms *bilinear* in two fields – a hallmark of particle-particle oscillations and thus that we are not working in the appropriate basis. Noting also that the Goldstone boson ξ has no known experimental counterpart, we exploit gauge invariance to rotate between the two components of ϕ by an angle

$$\theta = -\tan^{-1}(\phi^+/\phi^0), \tag{2.13}$$

and thereby eliminate ξ entirely. We also perform a rotation by what is called the electroweak mixing angle, θ_W, between $\mathcal{A}$ and B_3, and recombine B_1 and B_2. We have then instead of $\vec{B}$ and $\mathcal{A}$ a different set of four fields. There are now three massive fields

$$W^\pm = \frac{1}{\sqrt{2}}(B_1 \mp B_2) \tag{2.14}$$

$$Z = B_3\cos\theta_W - \mathcal{A}\sin\theta_W, \tag{2.15}$$

corresponding to the broken symmetries. Since a $U(1)$ symmetry remains unbroken,

the fourth boson stays massless

$$A = \mathcal{A}\cos\theta_W + B_3 \sin\theta_W. \tag{2.16}$$

These are the four gauge bosons which we observe in experiments. With this rotation, the mass terms which we created above become

$$M_W = g\frac{v}{2}, \tag{2.17}$$

$$M_Z^2 = \frac{M_W^2}{\cos^2\theta_W}, \tag{2.18}$$

$$m_\ell = \zeta_\ell \frac{v}{\sqrt{2}}, \tag{2.19}$$

where we see that the fermion masses are set by the dimensionless constants ζ_ℓ, which from the point of view of this theory are free parameters to be determined by experiment.

Thus during the symmetry breaking, one scalar degree of freedom is consumed in the creation of the longitudinal polarizations which newly-massive gauge bosons now require. The other remains as an observable particle, one which has for many years eluded all searches and is typically called the Higgs boson. With this particle are we here concerned.

2.2 Proton Collisions

We will be searching for the Higgs boson in the environment of high energy proton-proton collisions, and so we pause here to describe a few of their general properties.

When two clouds of the order of 10^{11} protons impinge on each other, the vast majority of collisions will be off-center or glancing. Such cases are characterized by small momentum transfer between the constituents of each proton, and thus tend not to involve interesting higher-energy processes. We will want to focus on the rare instances in which two protons encounter each other almost head on.

These head on collisions may be viewed as possessing two separate physical scales.

The internal dynamics of the proton occur at around a GeV and determine the identity and kinematics of the various parton constituents. This is the regime of non-perturbative QCD. The hard interaction of two partons, on the other hand, proceeds at the much higher energy scale of the particle collider, and can be approached with a straightforward perturbative expansion. This division is codified in the QCD factorization theorem, where we write the differential cross section $d\sigma$ as a convolution over momentum fraction x of the partonic cross section $d\hat{\sigma}$ with the phenomenological parton distribution functions (PDFs) f

$$\frac{d\sigma}{dp_{\mathrm{T}}^2} = \sum_{a,b} \int_0^1 dx_1 dx_2 f_{a/h_1} f_{b/h_2} \frac{d\hat{\sigma}}{dp_{\mathrm{T}}^2}, \tag{2.20}$$

for parton a from hadron h_1 and parton b from hadron h_2. PDFs are determined by fitting elaborate functions to a wide variety of experimental data from different experiments.

Within the hard interaction, on the other hand, resides the potential for substantial new physical processes, and it thus holds much of our interest.

The dominant hard process which one observes in proton-proton collisions is the production of two back to back jets. As can be seen in Figure 2-2, a number of well-characterized and thus uninteresting processes populate the roughly ten orders of magnitude which separate dijet production from the total Higgs boson cross section. The main task of a search for the Higgs boson is thus to devise ways to reduce the number of events which contribute from these background processes.

2.3 Higgs Production

The Higgs boson may be created in several different channels during proton-proton collisions. The various mechanisms are summarized at 8 TeV as a function of mass in Figure 2-3, while their dependence on center of mass energy at a mass of 126 GeV is listed in Table 2.3. We discuss the dominant gluon fusion first, and also briefly describe vector boson fusion and the associated production channels WH, ZH, and ttH.

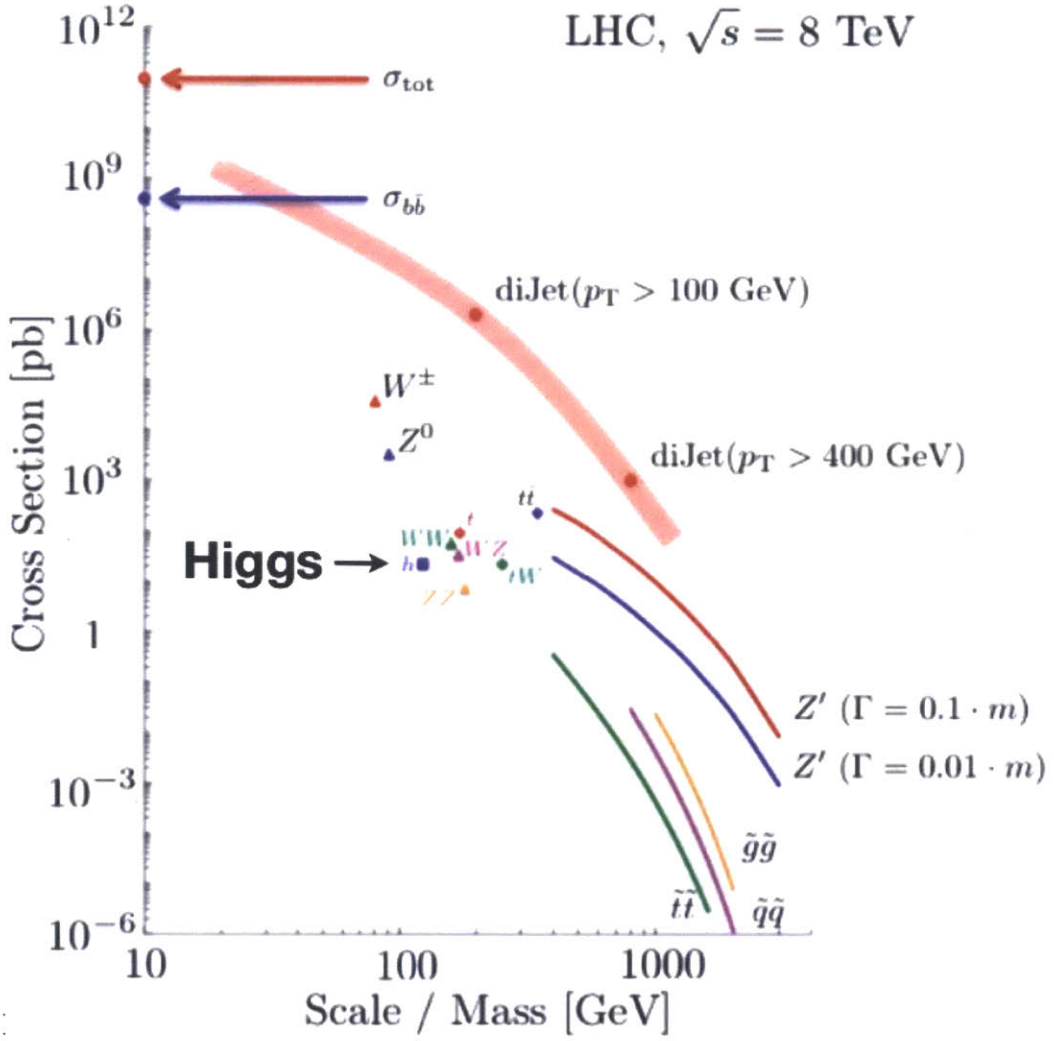

Figure 2-2: Cross sections for a selection of processes in proton-proton collisions at 8 TeV. It is to be noted that the total cross section, which consists mainly of two-jet events, is roughly ten orders of magnitude larger than the total Higgs boson cross section.

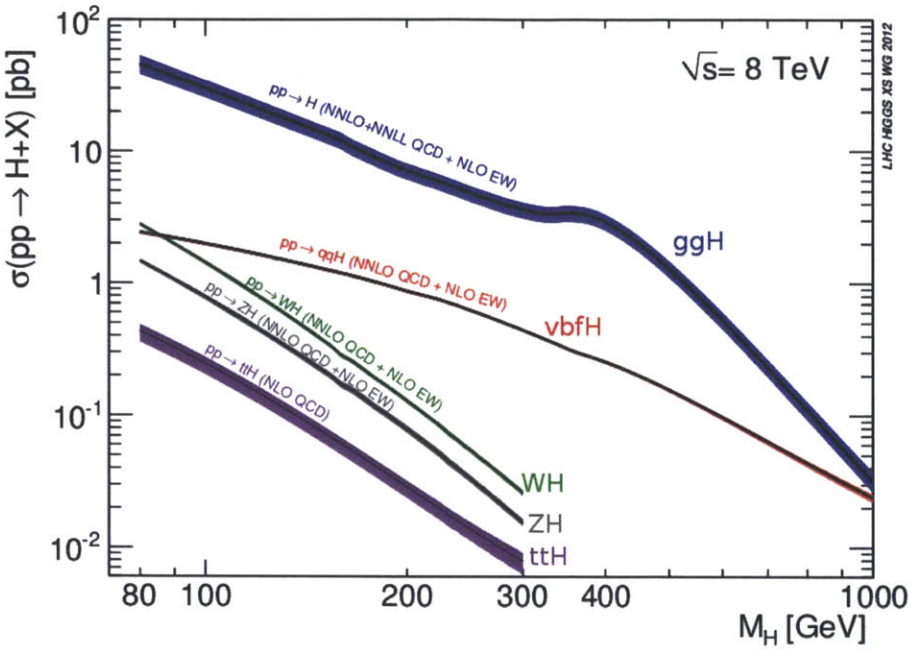

Figure 2-3: Higgs boson cross sections as a function of mass at 8 TeV for the five production mechanisms described in the text.

Cross sections (pb) vs center of mass energy

	ggH	vbfH	VH	ttH
7 TeV	14.89	1.211	0.8896	0.08426
8 TeV	18.97	1.568	1.0852	0.1262
14 TeV	49.85	4.180	2.387	0.6113

Table 2.1: Higgs boson cross sections in pb at a mass of 126 GeV for the three different center of mass energies relevant to the LHC.

2.3.1 Gluon Fusion

The dominant Higgs boson production mechanism at hadron colliders is the so-called gluon fusion (ggH) and is illustrated in Fig 2-4.

In this process gluons from each incoming proton interact via a heavy-quark loop, which then radiates a Higgs boson. Since the magnitude of the Higgs-fermion Yukawa coupling is governed by the fermion mass, loops including top quarks are the most important. The bottom quark, however, must also be included to achieve acceptable

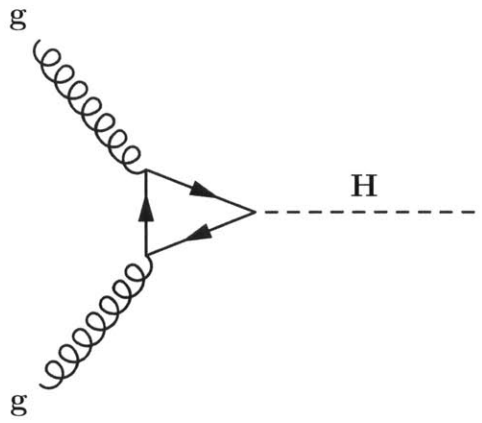

Figure 2-4: The lowest-order Feynman diagram for $gg \to H$.

accuracy.

The main difficulty encountered in making theoretical calculations for this process are that its dynamics are principally governed by the strong interaction. The fact that the next-to-leading order (NLO) contributions are within twenty percent of the magnitude of those at leading order gives some idea of the slowness with which even the most well-behaved perturbative calculation can be expected to converge.

The judicious choice of approximations is thus an imperative for the achievement of decent accuracy. The first such simplification is that the quarks in the loop are infinitely massive. This is numerically close to reality for the top, if we are discussing a light Higgs, but is also not a terribly poor choice for the bottom quark. Although the bottom is not nearly as heavy, its contribution to the total cross section is also much smaller, so on balance the effect is small. In the end this approximation is good to a few percent.

At next-to-next-to leading order (NNLO) the new contribution, although so far done only in the heavy-quark limit, is around one fourth the magnitude of the sum of the previous terms [19]. We thus see that the series appears to move towards convergence. This NNLO result can be further improved by resummation of soft gluon contributions up to the next-to-next-to-leading log level, which provides an upwards correction of between five and ten percent [20].

At the current level of accuracy, electroweak corrections must also be included. They are currently known to the two-loop level [21], and are roughly a five percent

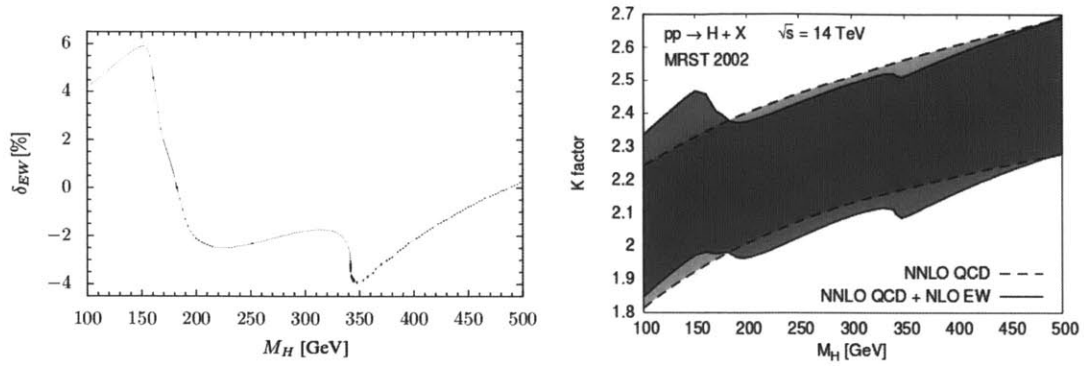

Figure 2-5: Higher order corrections to the LO Higgs boson gluon fusion cross sections as a function of Higgs boson mass. On the left are the NLO electroweak corrections, while on the right are the K-factors, which are defined as the ratio of the cross section including higher order corrections to the LO result. The dashed band shows the K-factor including only NNLO QCD corrections, while the solid band also includes the NLO electroweak corrections. Both are from [21].

effect for a low mass Higgs. Their full mass dependence is shown in Figure 2-5. Although the electroweak theory is very well understood, and thus these corrections can be calculated with a high level of confidence, the proper method which should be used to combine them with the QCD corrections above is not yet known. The basic choice is whether one should apply the electroweak corrections as a multiplicative factor only to the LO QCD result (this is called *partial factorization*), or as a multiplicative factor to the fully-corrected strong cross section (*full factorization*).

Attendant with the theoretical cross section estimate is of course a full budgeting of the contributing uncertainties. Given the nature of the strong interaction it should come as no surprise that the leading uncertainty is from uncalculated higher-order terms in the partonic cross section. When one makes a fixed-order perturbative calculation in QCD there emerge two arbitrary parameters: the renormalization scale and the factorization scale. The former is introduced as a cut-off in order to finesse away divergent integrals stemming from loop diagrams, while the latter arises during the factorization of the proton-proton cross section into a convolution of hard partonic cross sections and parton distribution functions. Both of these quantities are, at first, set to be of order the scale of the process, which for our purposes is the Higgs boson mass.

Since both the renormalization and factorization scale are essentially arbitrary, an ideal calculation, i.e. one made to infinite order, would not depend on them at all. The effect of their non-physical inclusion can thus be estimated by observing how much a change in their values modifies the final result. This should give an idea of the difference between our approximate calculation and the ideal one. In practice both parameters are varied up and down by a factor of two. It should be noted that this procedure only gives, strictly speaking, an estimate of the perturbative uncertainty. However, it is phenomenologically plausible that the factor of two gives decent coverage of the full possible phase space of variation. Furthermore, and more convincingly, the NLO, NNLO, and NNLL predictions are consistent when compared using these uncertainties. Since the perturbative contributions which are not included in each of these cases are to a large extent non-overlapping, this inspires confidence that we have in fact properly estimated this uncertainty.

The electroweak corrections also contribute appreciably to the overall uncertainty. As mentioned above this stems largely from our lack of knowledge as to whether to use full or partial factorization. The size of this uncertainty can thus be approximated by the difference between the two methods, and is 2-3% for Higgs boson masses between 120 and 130 GeV.

The effect of the large-m_t approximation has been evaluated by calculating the sub-leading contributions to the large-m_t limit, for instance in [22], and has been found to be less than one percent for a light Higgs and within a few percent for a Higgs boson heavier than 300 GeV. While in principle the scheme which is used for renormalization (for our purposes typically the modified minimal subtraction, or $\overline{\text{MS}}$, scheme) has no effect on observable quantities, the use of unphysically heavy quark masses in the case of the top and bottom in practice introduces some scheme dependence. This uncertainty is, however, easy to estimate by simply performing the calculation with a number of different values for the quark masses. This procedure gives a value of around 1%.

Finally, the choice of a particular parton distribution function (PDF) also has an associated uncertainty. In a typical implementation the fit from which PDFs are

m_H	Cross Section (pb)	QCD scale %	(PDF + α_s) %
90	36.23	+8.4 -8.8	+7.8 -6.6
120	20.86	+7.3 -7.9	+7.5 -6.9
126	18.97	+7.2 -7.8	+7.5 -6.9
130	17.85	+7.1 -7.7	+7.5 -6.9
300	3.594	+5.7 -6.1	+7.7 -7.9

Table 2.2: Cross sections at 8 TeV for Higgs boson production via gluon fusion at a few representative Higgs masses. The upper and lower uncertainties for the QCD renormalization and factorization scales, and for the PDFs and value of α_s, are reported separately. Results are at NNLL in QCD and NLO for electroweak uncertainties. Taken from [23].

extracted involves of order 50 free parameters, and as such the PDF's uncertainty can be estimated by independently varying the fitted values of each of these degrees of freedom.

The cross section values for gluon fusion along with their associated uncertainties are reported in Table 2.3.1 for a few representative Higgs boson masses.

2.3.2 Vector Boson Fusion

In vector boson fusion (VBF), the subleading contribution to Higgs boson production at the LHC, the Higgs boson is accompanied by jets. In the leading order diagrams two vector (W or Z) bosons, which have been radiated by the incoming quarks, fuse to form a Higgs boson (Fig 2-6). Because the final state quarks retain much of the colossal initial momentum of those in the initial state, we find in the far forward part of the detector two jets with very large momentum. While the cross section is roughly an order of magnitude smaller than that for the gluon fusion process, the VBF mode is possessed of great importance because it probes non-fermionic Higgs boson couplings, and because its unique jet topology provides for significant additional background elimination.

It can be seen from Figure 2-6 that at leading order VBF is a purely electroweak process, involving QCD only in the form of the quark and antiquark PDFs. In addition, while the s, t, and u channel diagrams are shown at leading order, both the

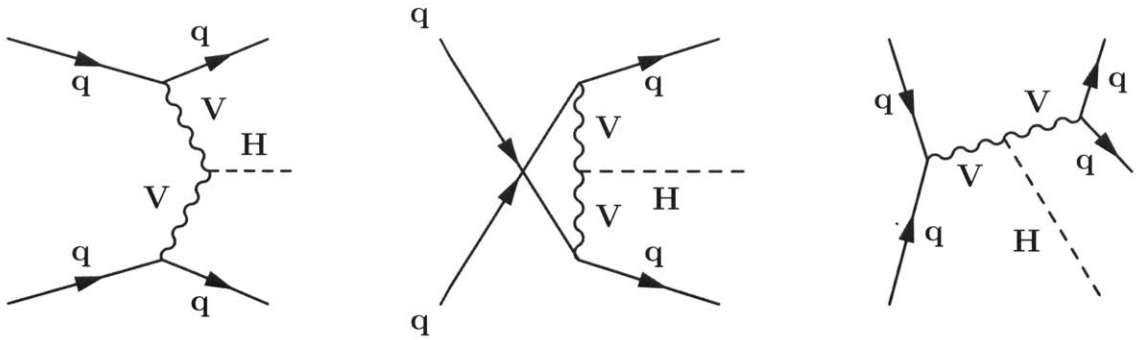

Figure 2-6: Leading order diagrams for $qq \to qqH$ in the t, u, and s channels. Note that q here represents any quark or antiquark and V stands for the W or Z bosons. The s channel diagram is included in associated production and also, in practice, contributes very little to the VBF channel.

s channel and interference between the three diagrams are suppressed, particularly when imposing typical VBF selection criteria. The squared t and u channel diagrams alone therefore provide an excellent approximation to the full cross section. The remaining QCD corrections reduce to vertex corrections on the vector boson - quark vertices, and are of the order of 5 to 10% [23].

An additional difficulty from which the VBF channel suffers stems from the contribution of ggH, which is now viewed as a background. While the total ggH cross section is quite well known, after the application of VBF cuts we find ourselves in a region of phase space which is both somewhat poorly understood and thinly populated in typical Monte Carlo event samples.

2.3.3 Production with Weak Vector Bosons

The Higgs boson can also be produced along with a W boson or a Z boson. The leading order diagrams are shown in Figure 2-7. In this mode, two quarks produce a weak vector boson in the s channel which then radiates a Higgs boson, hence the moniker Higgstrahlung. While this was the main production mode at the previous colliders with significant Higgs boson sensitivity, the Large Electron Positron collider and Tevatron, it is less important for the present case. Its importance in the current analysis stems mainly from the VBF region, since the presence of weak vector bosons frequently results in extra jets in the final state. In an analysis of Higgs boson decays

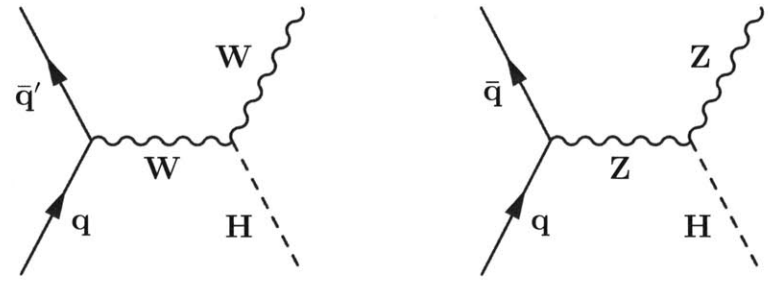

Figure 2-7: Leading order diagrams for production of a Higgs boson in association with weak vector bosons.

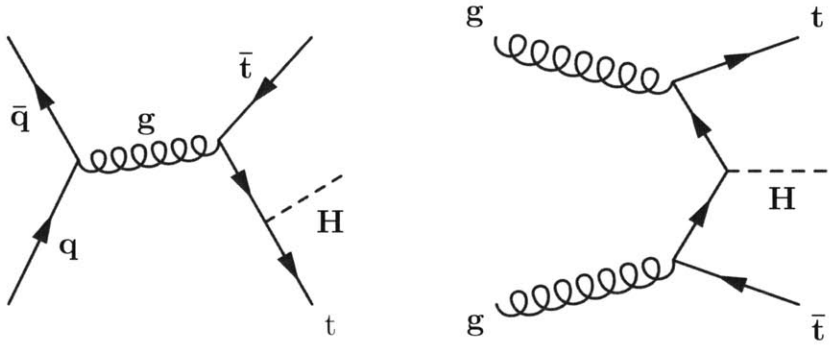

Figure 2-8: Leading order diagrams for production of a Higgs boson in associated with top quark pairs.

to a b quark pair at high p_{T} at the LHC, this mode is, however, of somewhat greater interest.

2.3.4 Production with Top Quark Pairs

For Higgs boson masses below 150 GeV the radiation of a Higgs boson from top quarks can also play a role, and provides information on the Yukawa coupling between top quarks and the Higgs. This production mode also contributes to this analysis mainly in the VBF region because of the additional jets which can result from top quark decays into W bosons and b quarks. The leading order diagrams for this process are shown in Figure 2-8.

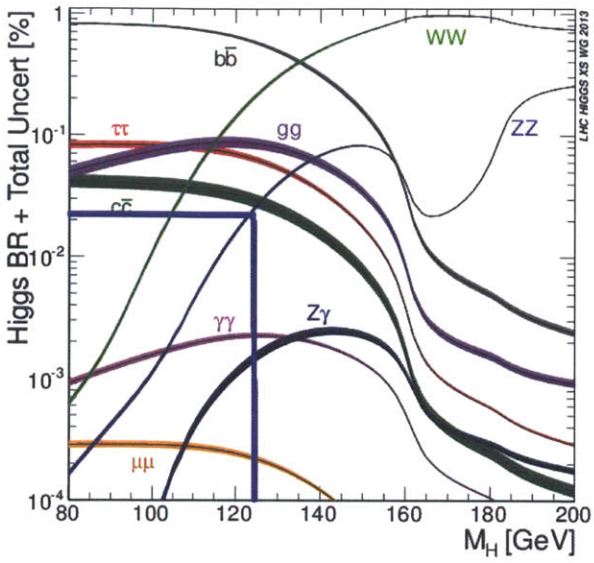

Figure 2-9: Branching fraction for Higgs boson decays to various final states. The blue lines mark the decay relevant for the present analysis at a mass of 125 GeV.

2.4 Higgs Decay

When searching for a new particle one would of course rather look in a decay channel with a large branching ratio in order to maximize the number of signal events. In practice this concern is balanced against the prevalence and distinguishability of the backgrounds. In the H→ZZ→4ℓ decay mode we have chosen a very small branching fraction of about 10^{-4} in exchange for background rates which are relatively trivial. The branching fractions on offer are illustrated in Figure 2-9.

These branching ratios are calculated using the two programs HDECAY [24, 25] and PROPHECY4F [26]. The information from the two are combined in order to arrive at an estimate of the total width by replacing the WW and ZZ widths from HDECAY with the corresponding value from PROPHECY4F

$$\Gamma_{tot} = \Gamma_{tot}^{\text{hdecay}} - \Gamma_{WW+ZZ}^{\text{hdecay}} + \Gamma_{4f}^{\text{prophecy}}. \tag{2.21}$$

The branching ratio uncertainties are separated into two parts [23]: parametric, which stem from experimental inputs; and theoretical, from unknown, mainly higher order,

parameter	central value	uncertainty
α_s	0.119	$\pm$0.002
m_{charm}	1.42 GeV	$\pm$0.03 GeV
m_{bottom}	4.49 GeV	$\pm$0.06 GeV
m_{top}	172.5 GeV	$\pm$2.5 GeV

Table 2.3: Central values and associated uncertainties for α_s and the charm, bottom, and top quark masses, which are used as inputs to the branching ratio calculation [23]. Other inputs such as the Fermi constant and the W and Z boson masses contribute below the per mil level and are thus not included.

	% branching ratios								
m_H	$b\bar{b}$	$c\bar{c}$	$\tau\tau$	$\mu\mu$	gg	$\gamma\gamma$	Zγ	WW	ZZ
126	56$\pm$3	2.8$\pm$12	6.2$\pm$6	0.021$\pm$6	8.5$\pm$10	0.23$\pm$5	0.16$\pm$9	23$\pm$4	2.9$\pm$4

Table 2.4: Higgs boson percent branching ratios and associated uncertainties for a mass of 126 GeV.

parts in the theoretical prediction. The parameters which contribute significantly to the former are α_s and the charm, bottom, and top quark masses. Their central values and uncertainties are summarized in Table 2.4. The branching ratio (or decay width) is calculated using the upper and lower bound for each parameter in order to obtain an envelope on the final value. The envelopes from each parameter are then added in quadrature to arrive at the total parametric uncertainty.

The portion of the theoretical uncertainties stemming from QCD are estimated by varying the scale up and down by a factor of two. The effect of higher electroweak orders, on the other hand, is based on the known NLO electroweak corrections. For the ZZ branching fraction the uncertainty is $< 0.5\%$ for QCD, while the electroweak uncertainty is 0.5% ($0.17\left[\frac{m_H}{1\text{TeV}}\right]^4\%$) for Higgs boson masses less than (greater than) 500 GeV.

The final branching ratio results are summarized for a mass of 126 GeV in Table 2.4.

As mentioned above, the decay kinematics of the four final state leptons reveal a great deal about the intermediate states from which they arose. We utilize these in two separate contexts: to distinguish Higgs events from background events, and to

distinguish among a variety of hypotheses for the spin and parity of the new boson. Since the background processes do not have an intermediate Higgs, and because of the different balance of s, t, and u channel production between signal and background, one finds that the angles between final decay products and the invariant masses of the intermediate Z bosons differ between signal and background.

When decaying to a pair of vector bosons, the Higgs boson gives information on its CP value [27, 28]. A CP-even particle will decay to a mixture of longitudinal and transverse polarization, whereas the bosons from the decay of a CP-odd state will have purely transverse polarization. This difference will, again, manifest itself in different kinematic distributions of the final lepton decay products. For example, the distribution of the azimuthal angle ϕ^* between the decay planes of the two vector bosons may be calculated to be proportional to

$$1 + a_1 \cos\phi^* + a_2 \cos 2\phi^*, \tag{2.22}$$

for particles with $J^{PC} = 0^{++}$, where a_1 and a_2 may be found in [27]. Meanwhile, in the case of 0^{-+} decays the distribution is proportional to

$$1 - \frac{1}{4}\cos 2\phi^*. \tag{2.23}$$

The particular parametrization which we choose for the lepton kinematic information, and their predictions, are described in detail in Chapter 9.

2.5 Background Processes

The initial goal of this analysis is to estimate the number of observed events in which a Higgs boson was produced. In general our final selection will contain a (perhaps non-zero) number of Higgs boson events as well as events which are attributable to previously known Standard Model processes. It is thus of primary importance that we have a reliable description of these background processes in order to enable a precise measurement of the number of signal Higgs boson events.

While we discuss the final event selection in more detail in Chapter 7, its main characteristic relevant for a discussion of backgrounds is the requirement of four charged leptons consistent with provenance from a pair of Z boson decays. This immediately reduces to but a few the contributing background processes and limits their yields to a manageable level.

The most important background, continuum ZZ production, contains two Z bosons which each decay to two charged leptons. The other significant backgrounds contain fewer than two Z bosons, and make their way into the signal region by means of additional leptons from non-prompt sources such as heavy meson decays or from misidentified hadronic jets. The main contributor to these fake backgrounds is events with a single Z boson accompanied by two jets, while the remainder is made up of top quark pair and WZ events.

2.5.1 Continuum ZZ production

The dominant background for this analysis at all masses is due to the continuum production of two Z bosons. This state is produced via two mechanisms, initiated either by two quarks (the predominant one, illustrated in Figure 2-10) or by a pair of gluons (shown in Figure 2-11). Although a continuum ZZ event results in a set of particles indistinguishable from the Higgs boson signal, the differing production mechanisms between signal and background and the absence of an intermediate scalar state in background result in differences in both the final lepton decay kinematics and in the p_{T} of the entire ZZ system.

Monte Carlo event generators exist for both continuum ZZ production modes, and more importantly the passage of the final state particles through the detector for these final states is easy to model. For this analysis we thus use straight simulation, POWHEG [29] in the case of quark initiation and gg2zz [30] for gluon fusion.

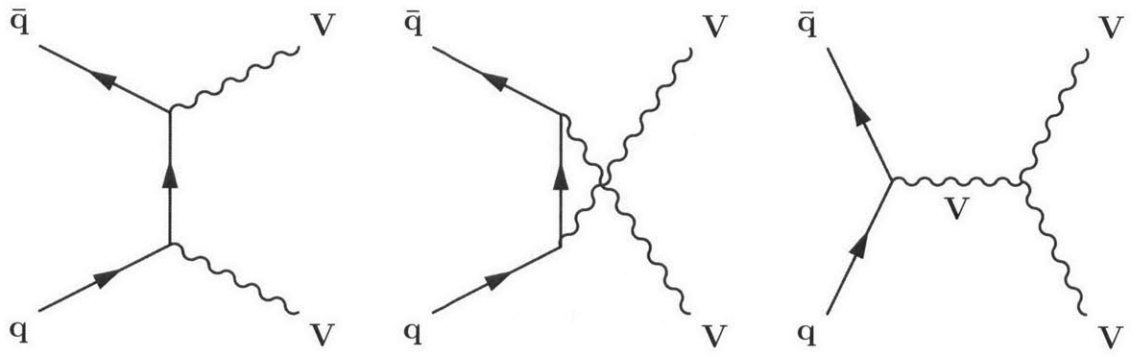

Figure 2-10: Leading order diagrams for the dominant background in the analysis, continuum ZZ production via quark anti-quark pairs.

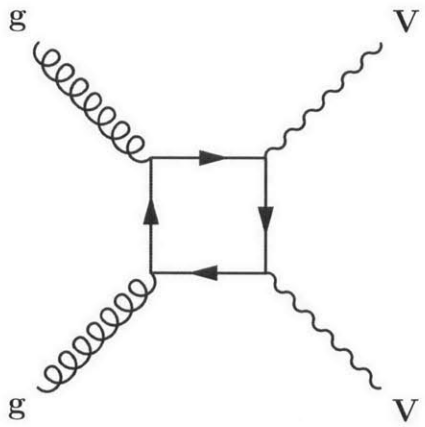

Figure 2-11: Leading order diagram for continuum ZZ production via gluon fusion [31]. This background contributes 5 - 10%, depending on mass range in the analysis.

2.5.2 Fake Backgrounds

In certain cases hadronic jets or non-prompt leptons can be mistaken for electrons or muons from Z boson decays, giving rise to what we call fake backgrounds. While numerical implementations exist which provide an accurate simulation of the fundamental physical processes, the means by which a converting photon or the constituents of a jet pass through the many layers of the CMS detector in such a way as to resemble a prompt charged lepton is quite difficult to model with much accuracy. In practice it is more reliable to base estimates of these backgrounds on events from data rather than on theoretical predictions. The procedure by which this is implemented is described in in Section 8.2.2.

2.6 Monte Carlo Event Generators

The link from the theory described above to the world of real particle detectors is spanned by the computer codes known as Monte Carlo event generators. Significant improvements are a constant in this realm due both to increasing theoretical understanding and to the inexorable march of Moore's law. The Monte Carlo event generation process may be divided into several sequential steps. The hard interaction is characterized by decently perturbative behavior and as a consequence is perhaps the most accurate stage in the simulation. The colored products of this hard process are then showered using parton branching models, and finally the particles are hadronized to create the input particles for the detector simulation. Throughout, particles are treated as stable if appropriate, or decayed if their lifetimes are short compared to the relevant time scales.

The pairing of the POWHEG hard event generator [32] with PYTHIA [33] for decays and parton showering is a common example of these Monte Carlo techniques, and forms the basis of most of the analysis which we describe.

The process's hard component corresponds roughly to interactions, decays, and radiation events occurring within a time $1/\Lambda$, where Λ is the typical QCD scale of a few hundred MeV. We are at the moment concerned with processes for which the scale

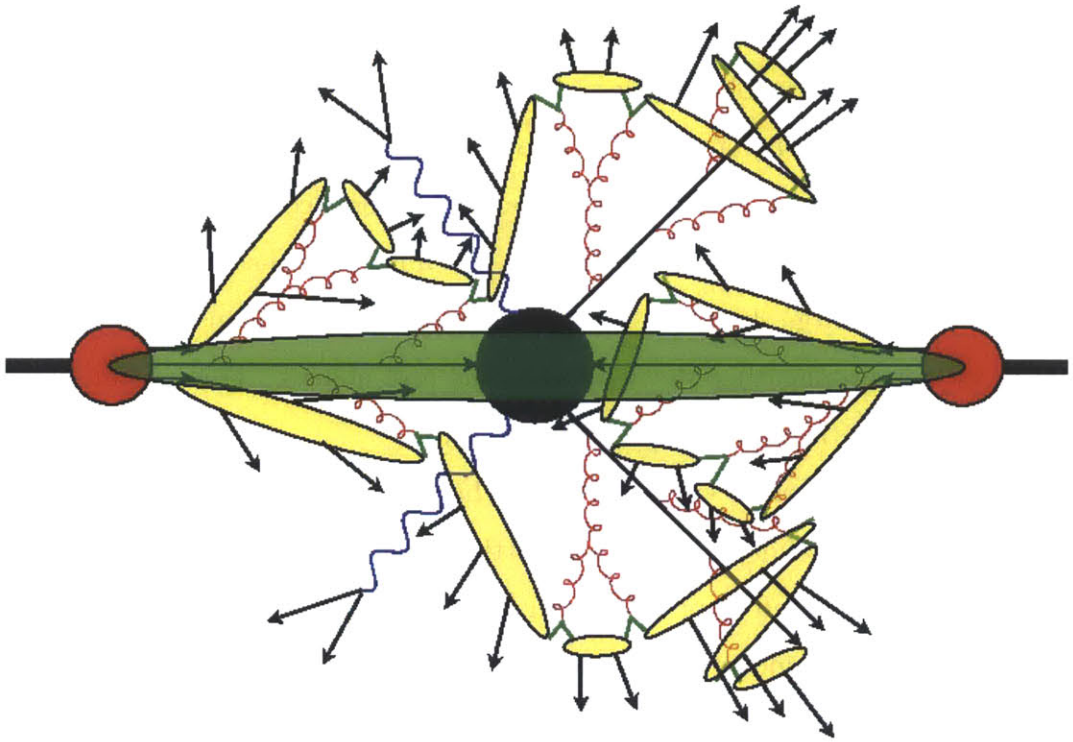

Figure 2-12: Stylized diagram of the steps in Monte Carlo event generation. Two partons (long horizontal black lines) are selected for the hard collision, one from each of the two incoming red protons. They both produce initial state radiation (ISR) in the form of gluons (curlicues) as they propagate toward the site of the hard interaction (large black circle). The four main products of the collision are two electroweak gauge bosons going to the left at forty-five degrees (which subsequently each produce fermion pairs), and two quarks going at forty-five degrees to the right. Each of these quarks showers as they split off gluons. Finally, in the process of hadronization (yellow ovals) all of the colored objects which were produced in showering and ISR are grouped into mesons and baryons (outermost black arrows). Note that additional small black horizontal arrows emanate from the protons. These are soft peripheral collisions between other partons in the protons: the underlying event. Image courtesy Scholarpedia

Q is much greater than Λ. Thus we are in a region where a fixed order truncation of the perturbative calculation will allow us to achieve a reasonable level of accuracy.

Care must, however, still be taken to avoid problems with the singularities which occur in QCD in association with soft gluon emissions and collinear splitting. These infinite quantities can be avoided in inclusive quantities such as total decay widths, but in more exclusive observables make themselves known by means of divergences which must then be carefully treated to ensure that they cancel.

The products of this hard process are in general colored QCD objects, and as in QED they can radiate gauge bosons. Because of anti-screening and the gluon's nonzero color charge, however, instead of the small corrections which are present in QED we get propagating showers of gluons and quarks. This parton showering is simulated using the QCD splitting functions.

Hadronization begins when the shower has reached the point at which all particles are below the threshold for further splitting. We are now firmly in the strong coupling region for α_s and so proceed with a model-based approach. The most prevalent is known as the Lund string model [34]. If we imagine a color-connected quark-antiquark pair at the close of showering, their mutual potential is approximately linear $V(r) = kr$ if we neglect the short-distance Coulomb type term. This linear potential describes an elastic string with tension of 1 GeV/fm. This picture is justified intuitively by the fact that confinement and the gluon color charge restrict the QCD field lines to a narrow string- or tube-like region between the quarks. As we metaphorically stretch the string, the system's potential energy increases seemingly without bound until relativity intervenes to allow the creation of new quark-antiquark pairs. This hadronization continues until there remain only colorless mesons and baryons.

2.6.1 The Higgs boson transverse momentum spectrum

Given a measure of certainty that a signal has been observed, focus naturally shifts to how these signal events are distributed in various kinematic variables and whether these distributions agree with SM expectation. We characterize the Higgs boson's kinematics with transverse momentum p_{T}, rapidity y, and azimuthal angle ϕ. Pro-

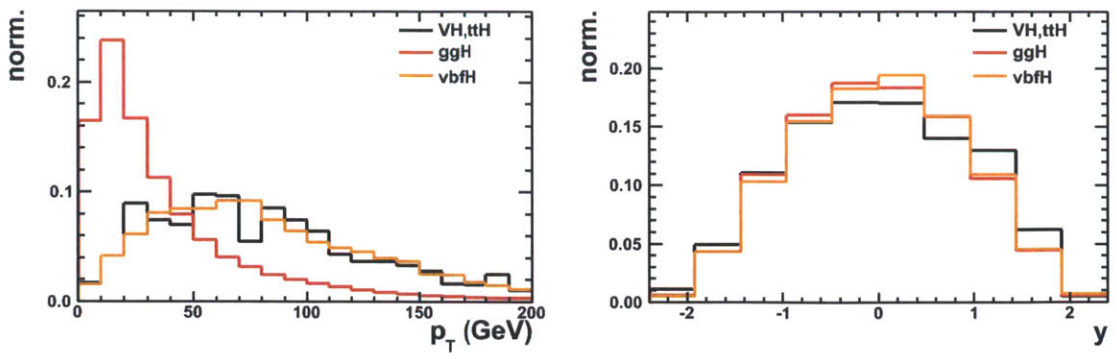

Figure 2-13: The expected kinematics of Higgs bosons produced in various production modes, shown for $m_H = 126$ GeV.

duction is expected to be uniform and thus uninteresting in ϕ. Although rapidity can be of some interest, for instance in [35], it is of minimal use in determining if production has proceeded as for a SM Higgs. For this task we want to know if the main production channels expected for the SM Higgs boson are present in the proper ratios, and whether any additional channels are present. As can be seen in Figure 2-13, p_{T} provides the main means of distinguishing the production modes.

Some care must be taken in order to arrive at a reliable prediction for the Higgs transverse momentum. While a fixed order calculation could in some cases suffice for p_{T} of order the Higgs boson mass, this region is of limited utility as most events fall in the lower reaches of the p_{T} spectrum. For the region in which we encounter the bulk of events, below around 40 GeV, one finds that terms in a given power of α_s have become enhanced by powers of a large logarithm: $\alpha_s^n \log^m(M_H^2/p_{\mathrm{T}}^2)$. The series must thus be resummed to all orders in these logarithms in order to achieve convergence.

As discussed in Section 2.2, the QCD factorization theorem expresses our understanding of the proton as a composition of quarks and gluons. For the purposes of the p_{T} spectrum, the relevant point is that for the purposes of resummation the partonic cross section can be written as a sum of two parts, the first of which contains all of the problematic low-p_{T} logarithmic contributions and which must be summed to all orders in α_s. The second, however, being free of these logarithms, can be evaluated

at fixed order

$$\frac{d\hat{\sigma}}{dp_{\mathrm{T}}^2} = \frac{d\hat{\sigma}^{\mathrm{res.}}}{dp_{\mathrm{T}}^2} + \frac{d\hat{\sigma}^{\mathrm{fin.}}}{dp_{\mathrm{T}}^2}. \tag{2.24}$$

In addition to these perturbative concerns, one also introduces a non-perturbative transverse momentum smearing parameter. Technically speaking, the above calculation is performed after making a Bessel transformation to the conjugate impact parameter space rather than in transverse momentum space. In this conjugate space, in practice, a Gaussian smearing factor is applied, with width to be tuned to data.

It should be noted that the parton showering approach described previously, which is the basis for the event generators, in effect approximates the full resummation. Thus although the actual Monte Carlo events which we use in the analysis do not incorporate higher order resummation, and thus are in general corrected to the differential distributions from a cross section calculator such as HRes [36], in practice the parton showering from PYTHIA does not do an unacceptably terrible job of approximating resummation.

The concerns above are relevant for making predictions of the Higgs boson p_{T} spectrum in any theory. The experimental question of most relevance is whether an observed Higgs boson corresponds to that expected under the SM or to some more exotic theory. Due to its sensitivity to the particulars of production modes, the Higgs boson p_{T} spectrum has significant discriminating power between different Higgs boson hypotheses [37]. Its further constraint can also improve the total cross section measurement and reduce theoretical uncertainties in coupling measurements.

Chapter 3

The Large Hadron Collider

There have been a number of previous searches for the Higgs boson during the forty-odd years since its prediction. The main impact has been from direct searches at the Large Electron Positron (LEP) collider, from indirect constraints from global electroweak fits, and from direct searches at the Tevatron.

After operating for a number of years with a center of mass energy at the Z boson pole, LEP's energy was increased above 200 GeV in order to study WW scattering and to search for the Higgs. This run resulted in a 95% lower bound on the Higgs boson mass at 114.4 GeV [39]. A wide variety of precision measurements of electroweak observables also enable the placement of constraints on the Standard Model Higgs boson mass [40, 41]. The Higgs boson makes itself felt in these quantities via radiative corrections, and fits were performed which found the value of the Higgs mass which is most compatible with the measurements to be 91^{+30}_{-20} GeV. Finally, direct searches at the Tevatron proton-antiproton collider resulted in an additional exclusion region between 155 and 180 GeV [42].

The Large Hadron Collider (LHC) is a particle accelerator located at CERN in Geneva, Switzerland encompassing roughly sixty square kilometers of area. It occupies the old LEP tunnel, and was designed to deliver proton-proton collisions at a maximum center of mass energy of 14 TeV and maximum instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ [43]. At design specifications it was expected to provide sensitivity to the Standard Model Higgs within the entire plausible mass range and thus definitively

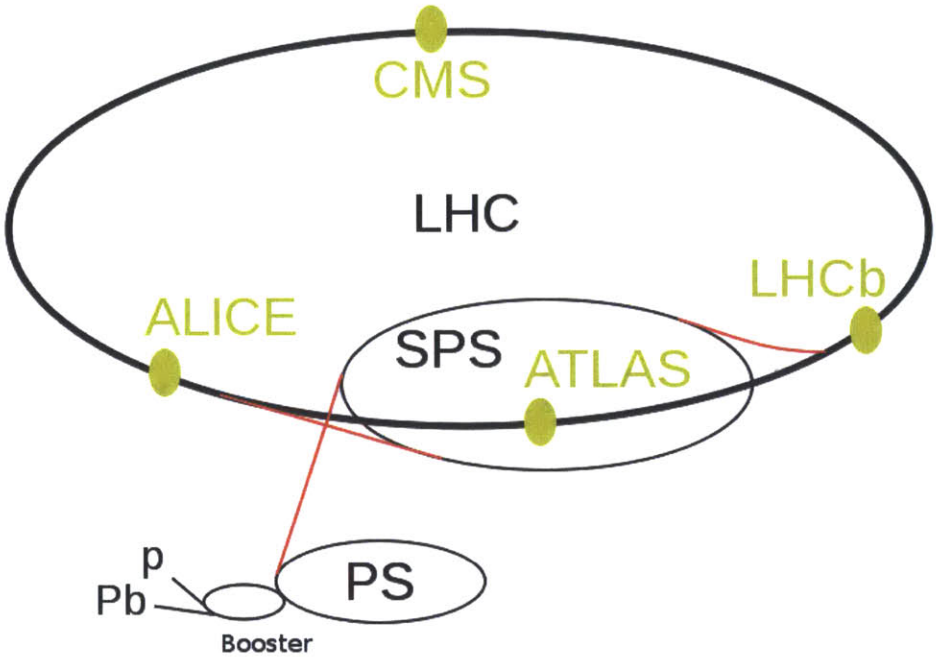

Figure 3-1: Schematic representation of the LHC accelerator complex at CERN in Geneva, Switzerland. Protons originate at the p, then accelerate through the Linac 2 linear accelerator (small straight line), the Booster synchrotron, Proton Synchrotron, Super Proton Synchrotron, and finally the main LHC ring. Collisions of counter-rotating proton beams occur at the four labeled yellow points.

settle the matter of electroweak symmetry breaking. The whole of the LHC accelerator complex which contributes to this effort is shown in Figure 3-1. The complex consists of an initial linear accelerator followed by four synchrotrons, and acceleration begins from a single bottle of hydrogen gas. An applied electric field removes the atomic electrons before insertion to the initial linac, called Linac 2 [44]. Upon leaving the linac the protons have achieved an energy of 50 MeV and move on to the Proton Synchrotron Booster [45]. The booster consists of four synchrotron rings on top of each other in order to increase the potential luminosity, and results in a final energy of 1.4 GeV. After exiting the booster, protons are accelerated first to 25 GeV in the Proton Synchrotron [46] and then to 450 GeV in the Super Proton Synchrotron (SPS) [47].

Finally, the protons are extracted from the SPS into the main LHC ring. The ring

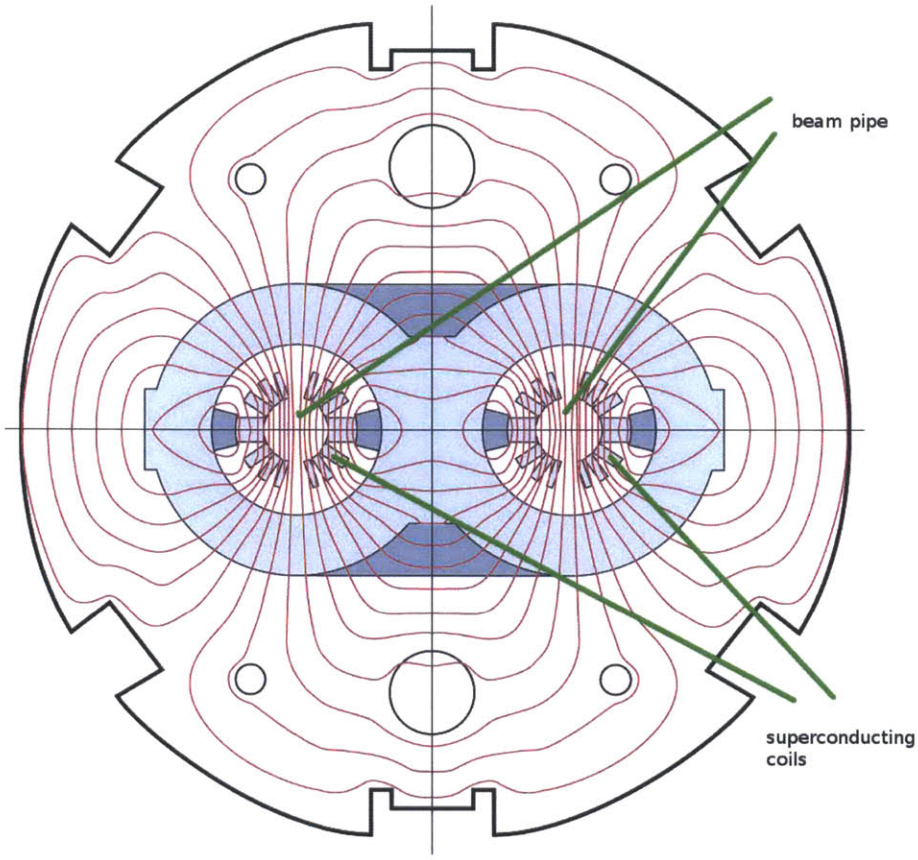

Figure 3-2: Cross section of an LHC dipole magnet with superimposed red magnetic field lines. In order to create a uniform 8.3 Tesla magnetic field in opposite directions in each beam pipe, superconducting coils are arranged as indicated by the labeled blue squares. Superconductivity is maintained using liquid helium cryogenics.

is close to the shape of a circle with a radius of four kilometers, but is really a sequence of eight alternating curved and straight sections. The former house superconducting dipoles to keep the beams on track, while the latter contain beam services such as radiofrequency cavities, cleaning, and dumping, as well as the four collision points. The LHC uses a pair of beam pipes, one for the proton beam going in each direction. This gives rise to the particular design for the LHC dipoles shown in Figure 3-2, which utilizes shared services and a shared magnetic return yoke for the two pipes. The LHC parameters from the main years of run I are summarized in Table 3.

At each of the four interaction points around the ring the counter circulating beams are crossed, creating collisions at the focus of a large particle detector.

year	2011	2012
bunches	1381	1381
protons per bunch	1.3×10^{11}	1.5×10^{11}
bunch spacing	50ns	50ns
β	1m	0.6m
emittance	2.5μm	2.5μm
max. luminosity	4×10^{33}	7×10^{33}

Table 3.1: LHC beam parameters during 2011 and 2012.

Chapter 4

The Compact Muon Solenoid

4.1 Overview

At the interaction point labeled "P5" on the LHC ring lies the Compact Muon Solenoid (CMS), one of the two general purpose detectors which have been designed to provide sensitivity in the greatest possible variety of physical analyses [48]. As such its main design goals are to provide good particle identification and momentum or energy resolution over as large a fraction of 4π as possible.

4.2 Magnet

CMS takes its name from the 13 meter by 6 meter solenoidal superconducting magnet which provides the magnetic field that enables momentum measurements. The solenoid encloses an area of uniform 4 Tesla magnetic field which contains the tracking and calorimetry. The return flux is confined within an extensive iron yoke outside of the solenoid at a magnitude of 2 Tesla. The muon chambers are mounted within the yoke in order to take advantage of the return field.

The magnet parameters were largely set by the need for a resolution $\delta p/p < 0.1$ and unambiguous charge identification for muons up to 1 TeV. They are summarized in Table 4.2.

central magnetic induction	4 T
total Ampere-turns	41.7 MA-turns
nominal current	19.14 kA
inductance	14.2 H
stored energy	2.6 GJ
operating temperature	4.65 K

Table 4.1: Parameters of the CMS superconducting solenoid.

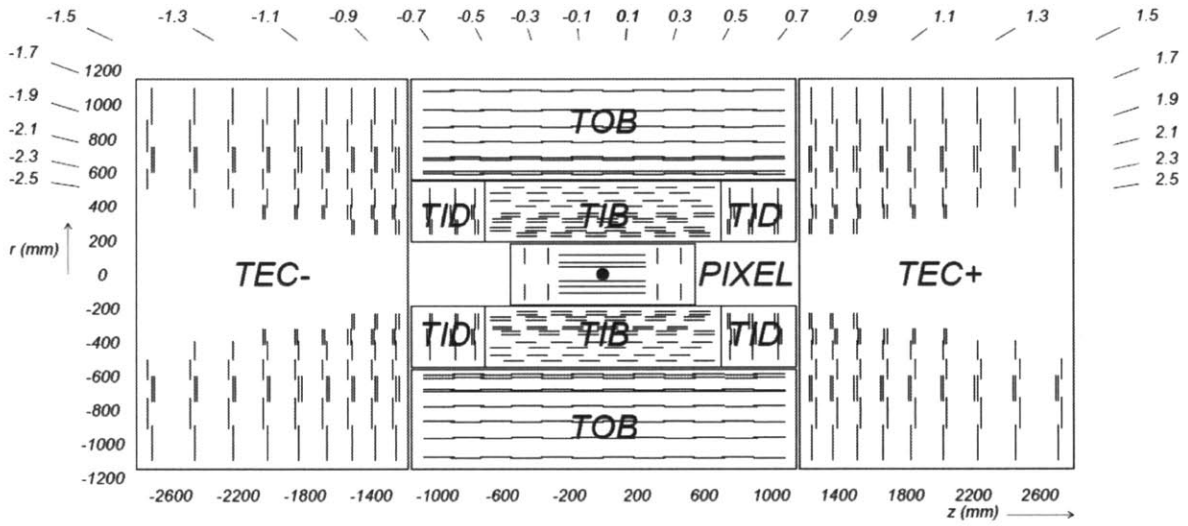

Figure 4-1: Cross section of the CMS tracker in the r-z plane with coordinates indicated in cm along z and r, and in pseudorapidity along η. The silicon pixels are in the shoebox-sized region nearest the interaction point, while the various strip subdetectors are labeled with an intriguing variety of three letter acronyms whose intricacies do not concern us.

4.3 Tracking

The first CMS layers outside of the beam pipe provide all-silicon charged particle tracking. These semiconductor devices are arranged in a combination of pixels and strips which is shown in Figure 4-1. The innermost layers, nearer than 15cm to the interaction point, are composed of very small 100μm x 150μm pixels in order to keep occupancy per bunch crossing as low as possible. Further out but still within 55cm the particle flux has dropped enough to enable the use of strips measuring 10cm x 80μm. Beyond 55cm, even larger strips of 25cm x 180μm are used. The tracking

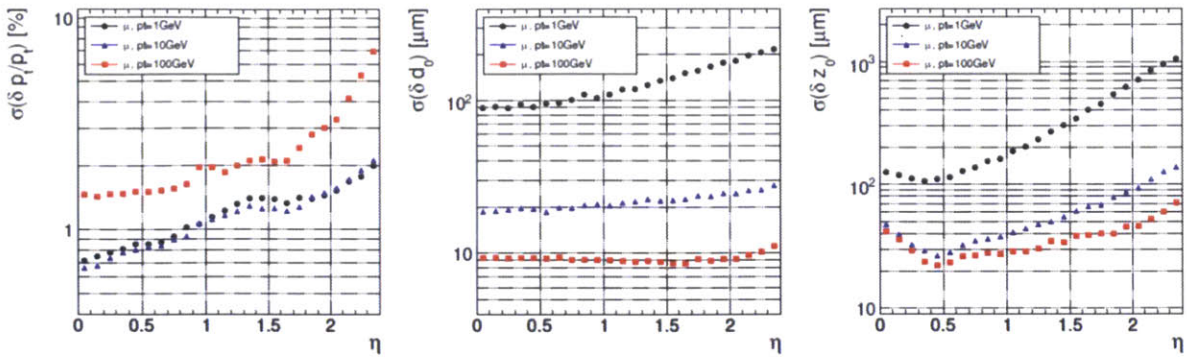

Figure 4-2: Tracker resolution for, from left to right, p_{T}, transverse impact parameter, and longitudinal impact parameter, for muons between 1 and 100 GeV.

system's performance is summarized in Figure 4-2. At momentum scales typical for Higgs boson decays of 10 to 100 GeV momentum resolution is a few percent, and impact parameter resolution is of order 10μm.

4.4 Electromagnetic Calorimeter

After passing through the tracking volume, particles reach a homogeneous $PbWO_4$ electromagnetic calorimeter (ECAL) which is designed to stop electrons and photons. The use of a high density inorganic crystal serves to minimize both radiation length (0.90cm) and Moliere radius (2.2cm), thereby allowing a generous 26 radiation lengths to fit within the solenoidal volume. Most of the light is collected within the design LHC bunch spacing interval of 25ns, enabling unambiguous bunch crossing identification.

Because the ECAL's performance is intimately linked to its light sensitivity, a crystal-by-crystal monitoring system is used in order to measure and correct for changes in crystal transparency over time. Laser light is injected through optical fibers into each crystal, and the transparency to laser light is then extrapolated to scintillation light transparency using a predetermined phenomenological relationship.

The performance of the ECAL is principally codified in the fractional electron/photon energy resolution, which can in general be written as a quadratic sum of three com-

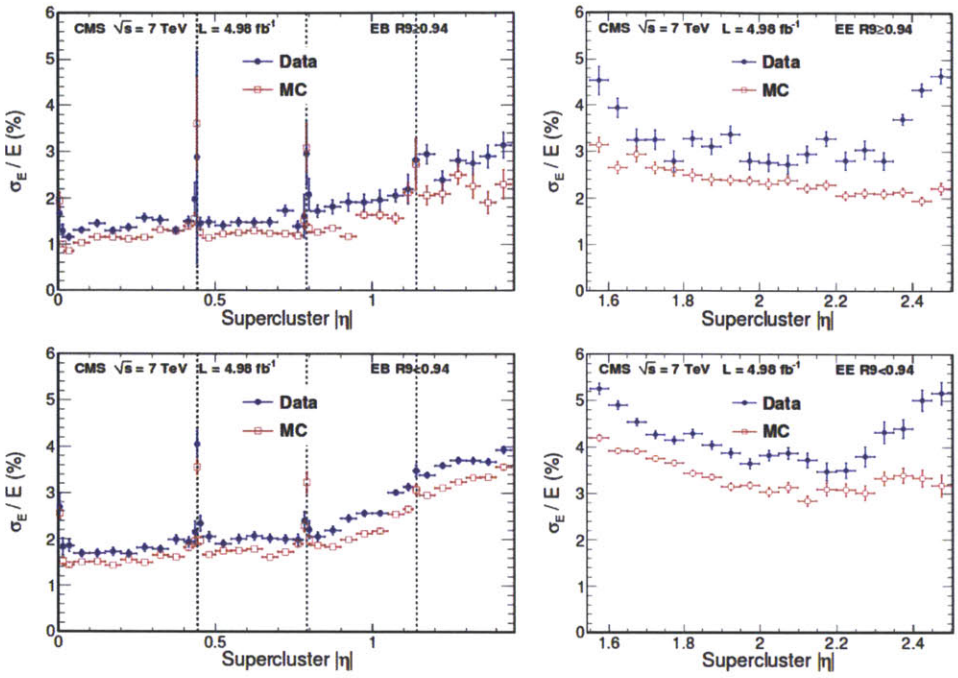

Figure 4-3: Fractional electron energy resolution as a function of η for electrons with minimal (top) and high (bottom) levels of bremsstrahlung.

ponents

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \tag{4.1}$$

for E in GeV. The first term is stochastic and stems from shower containment, from the number of photoelectrons and from fluctuations in the gain process. The second term is from pileup, digitization, and noise in the electronics. The final, constant, term is due to non-uniformity in longitudinal light collection, leakage out the back of the calorimeter, and lack of single-channel response stability. This formula, however, is derived from test-beam data and is consequently somewhat idealized. In practice the resolution is determined by fitting for a width parameter in a sample of Z→ee decays [49]. The results of this latter procedure are shown in Figure 4-3. It is unknown why the Monte Carlo simulation systematically underestimates the ECAL resolution. However, the effect is highly correlated with the amount of tracker material inside the ECAL, so is suspected to be due to an inaccuracy either in the material budget in simulation or in the details of the bremsstrahlung process in the tracker material.

4.5 Hadron Calorimeter

The hadron calorimetry is realized by means of alternating layers of brass absorber and active plastic scintillator tiles. The light from scintillation is downshifted in wavelength by thin fibers in order to enhance collection. The hadron calorimeter's main goals are tau identification and jet and missing transverse energy measurement.

The jet energy bias and resolution are usually determined either by the dijet balancing technique, which takes advantage of the inherent balance in p_{T} between jets in dijet events, or by measuring the apparent missing transverse energy (MET) in events with no real MET, for example Z+γ events [50]. The latter type of events also serve as a means of calibrating the MET resolution.

4.6 Muon Systems

Particles are identified as muons by their unimpeded passage through the entirety of the CMS calorimetry. The variety of gas detectors which comprise the muon systems to a large extent simply tag muons in order that they can be matched to tracks in the inner tracker. They also, though, provide a momentum measurement by means of the return field.

In the central detector regions, where particle fluxes and the residual magnetic field are low, drift tubes are used to take advantage of their good coverage. Cathode strip chambers are utilized in the higher flux and larger magnetic field environment of the more forward regions. Resistive plate chambers are also found interspersed throughout all rapidity ranges because of their excellent timing resolution.

The drift tubes give a drift time of roughly 380ns, and a position resolution in the r-ϕ directions of about 100μm. The cathode strips have position resolution between 75 and 150μm and timing resolution of a few nanoseconds. The resistive plate chambers, meanwhile have very poor position resolution of about 1.5cm, but time resolution of order a few nanoseconds.

As muon system performance is dependent on alignment accuracy of a few hundred

microns in the r and ϕ directions among its various components and with respect to the inner tracker, an optical alignment system has been installed. This uses a network of LED lights and lasers monitored by photodetectors and analog sensors in order to measure position relative to rigid reference structures. This information is combined with data from muon tracks originating from cosmic rays and proton collisions in order to arrive at a final calibration.

The overall momentum resolution below 100 GeV is 1-6%, depending on pseudo-rapidity range, and about 10% for 1 TeV muons [51].

4.7 Triggering

At the LHC's design luminosity inelastic proton-proton collisions will occur in CMS at a rate of about 10^8-10^9 Hz. The permanent storage and data transfer systems, however, can handle a rate only of order 100 Hz. The trigger system acts to span this gap. First comes the largely hardware-based Level 1 (L1) trigger, consisting of programmable custom electronics, which achieves a reduction to 100 kHz. Events are then filtered through a ten thousand core commercial processor farm called the High Level Trigger (HLT).

The L1 trigger builds up a very approximate picture of important quantities in the event starting from a hierarchy of subsystems which feed it data from the calorimeters and muon chambers. If the global L1 trigger reaches a "pass" decision, the detailed event data, hitherto stored in electronic buffers, is released and passed on to the HLT.

The HLT carries out much of the type of detailed event reconstruction which is later performed on data for the final analysis, but with two main differences. First, the HLT reconstruction code is optimized for speed and as such uses many approximations which would not be appropriate for final data. Second, reconstruction proceeds in a sequential manner, such that quantities are computed only if they are needed. Thus although all paths are executed for all events, computation on a given trigger path is halted once a failing result is reached [52].

Chapter 5

Regression and Decision Trees

A number of problems are encountered in this analysis which are ideally suited to the use of regression and decision trees, and we thus pause to describe some general features of these techniques [53]. For both tree types, the basic problem is that one would like to go from a set of inputs X_i to an output Y. We have both a sample where we know the proper mapping from the X_i to Y, which we can use to train our method, and then a separate sample where we do not know the correct mapping, on which we hope to apply the method.

In the simplest toy instance of this problem, we could begin with a one-dimensional data set for X_i which we happen to know is linearly related to Y. So we fit a linear function to the X_i, and can use this line to predict the response on other similar samples. The two complications which we encounter in practice are multidimensional inputs and nonlinear mappings between X_i and Y. It turns out, however, that the method can be generalized quite simply to these cases.

For the sake of concreteness, we can picture a typical case where the X_i are a set of variables describing an electron in the detector: raw calorimeter energy and its uncertainty, track momentum and its uncertainty, and variables describing the exact position in the detector and the shape of the shower. We wish to derive from this the energy of the original electron. A global model of the relationship between these inputs and the true energy, however, would be hopelessly complicated. The basic idea of a regression tree is that instead of trying to fit the entire phase space at once with a

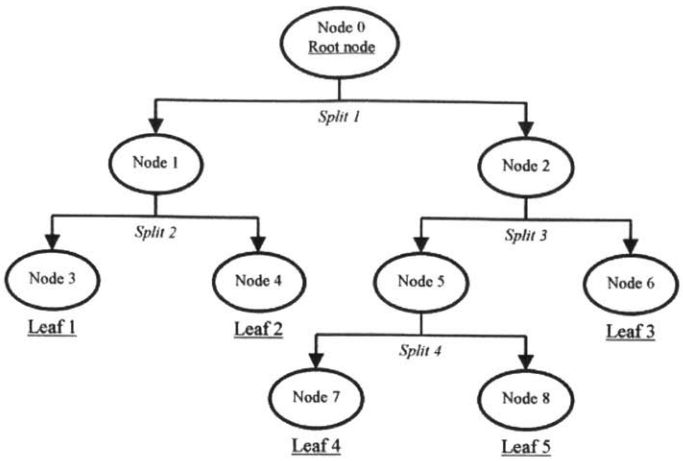

Figure 5-1: Diagram of the general procedure employed in regression and decision trees. The sample is split at the root node by cutting on the variable which at that point yields the best performance. This procedure is then repeated at each splitting until the final leaves. At each leaf we have a sub-sample which is either close to homogeneous in electron energy (regression trees) or highly enriched in one class (decision trees).

complex model, we instead recursively partition the space into chunks that are small enough that within each one a simple model will suffice. It is generally best to put most of the effort into constructing an optimized tree (i.e. an optimized partitioning of the data), such that the "model" can be a simple average over the X_i which remain in each final leaf, rather than to attempt to refine the model.

The procedure for constructing such a tree is depicted in Figure 5-1. We begin at the root node with the entire training sample and must choose the best variable to use to divide the sample in two such as to provide the most accurate estimate of the electron energy. The "most accurate" is taken to be that which minimizes the sum of the squared differences between the estimated (i.e. mean) value within the subdivision and each true value. The process is repeated on each of the two subsets, and in turn on the subsets created from them, until either the improvement in the sum of squares or the number of points in a node is very small.

A single tree implemented like this will, however, suffer from instability with respect to statistical fluctuations in the training sample from which its structure has been derived. Two main techniques are used to improve stability and performance in

this regard. First, having trained one tree, we examine the events which were very poorly measured, and give them larger weight in a new tree. This is called boosting [54, 55], and after repeating it a number of times we have a forest of trees. The final result is obtained by taking a weighted majority vote of all the trees. Secondly, we can train each tree on a different randomly chosen subset of our training sample. By thus bagging the whole sample into pieces we effectively smooth over the statistical fluctuations of the original sample.

If instead of a regression tree we would like to construct a decision tree, for instance to distinguish signal events from background events, only two changes are necessary. First, the criterion for which is the "best" variable to cut on at each branching is now that which provides the greatest increase in signal and decrease in background, which is typically implemented as a decrease in the Gini coefficient. Also, instead of the end result of our tree at each leaf being a very simple model of electron energy, it is a sample which is almost entirely either signal or background.

The implementation which we use in this analysis for decision trees is the TMVA framework [56]. For regression trees we use a private implementation [57].

Chapter 6

Event Reconstruction and Object Identification

6.1 Introduction

We next interpret the raw detector outputs in terms of physical objects on which later steps of the analysis will act. This reconstruction proceeds with information from all the constituent subdetectors at its disposal, and incorporates each piece as needed in order to arrive at an optimal result.

6.2 Primary Vertices

At the luminosities reached by the LHC (see Table 3) one encounters a significant number of inelastic but low momentum-transfer events coincident with each interesting hard collision. These additional collisions are modeled in Monte Carlo (Figure 6-1), but they require that we take care in selecting the proper vertex from which the tracks resulting from the primary collision actually originate.

In order to make a collection of all vertices in the event, we first select tracks which are compatible with having originated at the beam line. These tracks are then clustered together in groups which come from the same region. Finally, within each cluster we fit for the actual vertex position, and keep all vertices which are within

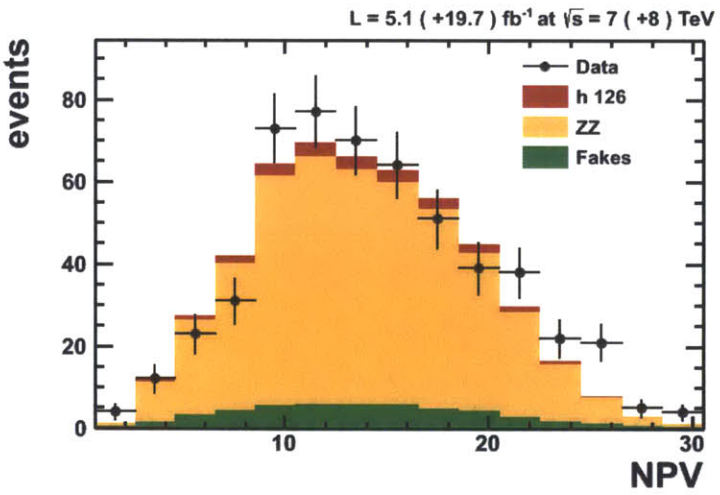

Figure 6-1: Comparison of the number of reconstructed vertices in data and simulation.

2 cm in the transverse direction and 24 cm in z of the interaction point. The fit is required to have at least four degrees of freedom. If more than one vertex satisfies these requirements, we choose the one whose tracks' p_{T} sum to the highest value.

6.3 Isolation

Because most background processes at the LHC involve copious production of hadrons within jets while the leptons from electroweak decays typically have little surrounding activity, we can increase signal purity by requiring that our leptons are isolated in the detector. To quantify this we introduce ΔR as the distance between two particles in η-ϕ space, $\Delta R^2 = \Delta\eta^2 + \Delta\phi^2$. We call a particle isolated if the sum of the energy deposits and track p_{T} in a cone with some radius in ΔR is small. Specifically, we define the isolation as a sum of the p_{T} of charged hadrons from the primary vertex and of the energy of all neutral hadrons and photons

$$I = \sum p_{\mathrm{T}}^{\mathrm{charged}} + \sum p_{\mathrm{T}}^{\mathrm{neutral}} + \sum p_{\mathrm{T}}^{\gamma}. \tag{6.1}$$

However, some of the neutral hadrons and photons in this sum will come from pileup vertices rather than the hard collision. To correct for this we subtract from the sums an estimate of the energy from pileup interactions which was deposited in the area of the isolation cone. In addition we want the isolation to be positive, and thus in total we replace the second two terms with

$$\max\left[0, \sum p_{\mathrm{T}}^{\mathrm{neutral}} + \sum p_{\mathrm{T}}^{\gamma} - \sum p_{\mathrm{T}}^{\mathrm{pileup}}\right]. \tag{6.2}$$

For electrons, $p_{\mathrm{T}}^{\mathrm{pileup}}$ is calculated as the product of the neutral particle energy density and the area of the isolation cone, while for muons it is one half the sum p_{T} of charged particle tracks from pileup vertices within the cone. In the latter case, the factor of one half is a phenomenological factor which has been found to give a good estimate of the relative abundances of charged and neutral particles.

Electrons and muons in the final selection are required to have a ratio of I to p_{T} of less than 0.4.

6.4 Muons

The core of identifying a muon is the idea that no known charged particle will propagate through the considerable thickness of the calorimeters. At its simplest, then, muons are roughly equivalent to combinations of muon chamber segments. In practice, however, the momentum measured in the tracker is much more accurate than that measured in the muon chambers except at the highest momenta, and some degree of hadronic energy can punch through from the calorimeters. We thus require some level of compatibility between a prospective muon's trails in the tracker and muon systems.

Two methods are used to accomplish this. We can start from a muon chamber track and follow its likely trajectory backwards into the silicon tracker. Once there we look for a track with compatible parameters; if one is found, a fit is performed for the two tracks together using the Kalman filter technique [58]. Alternatively, we

take all tracker tracks with $p_{\mathrm{T}} > 0.5$ GeV and $p > 2.5$ GeV and extrapolate their path outward to the muon chambers, and there look for muon chamber tracks with compatible parameters.

Because very low p_{T} muons are both poorly measured and dominated by background processes, we only consider muons with p_{T} greater that 5 GeV. Similarly, muons must be within the detector acceptance, with η magnitude less than 2.4.

We also require that either the muon track is matched to a calorimeter deposit consistent with the track's identity as a muon, or that the muon candidate has few surrounding particles. The lack of surrounding particles is implemented as the requirement that the sum of the p_{T} of all tracks within a radius $\Delta R = 0.3$ are less than 10% of the muon candidate's p_{T}.

To eliminate muons from pileup interactions, b quark decays, and cosmic rays, we also require the muon's track to pass within 0.5 cm in the transverse plane and 1 cm in the z direction of the selected primary vertex, and that the three dimensional distance to the vertex be less than four times its uncertainty.

The muon momentum determination is corrected using muons from Z boson, J/ψ, and Υ decays by comparing line shapes between data and Monte Carlo. Fits to data and Monte Carlo are performed in order to extract additional scale and smearing corrections, and these are applied to data and Monte Carlo respectively.

6.5 Electrons and Photons

Reconstructed electrons and photons begin as energy deposits, called superclusters, in the ECAL. These deposits are typically of significant extent in ϕ because bremsstrahlung photons or pair-produced electron-positron pairs are streamed off while the initial particle curves along ϕ before entering the ECAL. Starting from each supercluster, an attempt is made to match the supercluster to a tracker track. If such a match is found, the object is assumed to be an electron; otherwise it is a photon.

The dominant challenge in electron and photon reconstruction in CMS is the prevalence of bremsstrahlung and pair production during traversal of the few radiation

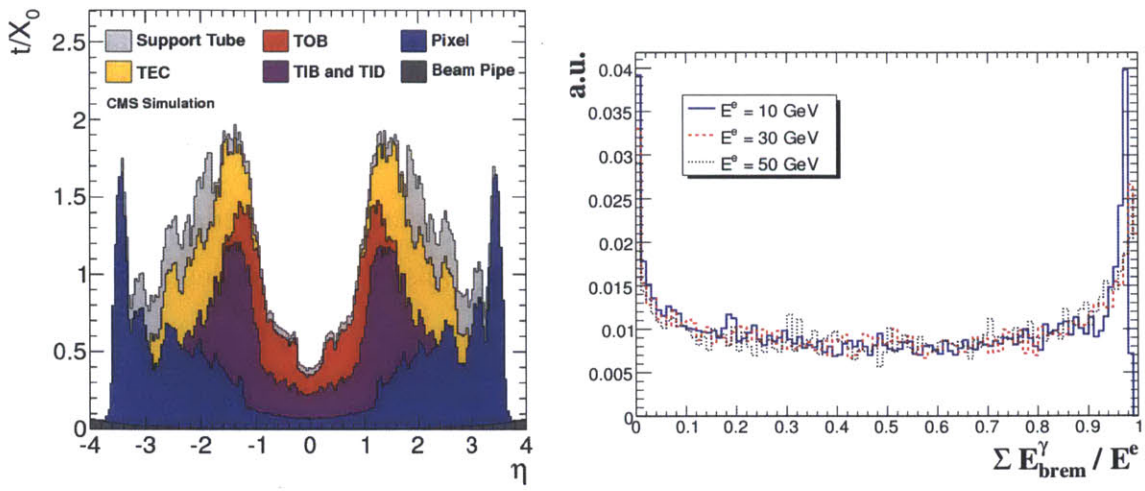

Figure 6-2: The quantity of tracker material, in units of radiation lengths, which precedes the CMS electromagnetic calorimeter (left), and its consequence: the radiated energy fraction for electrons with p_T of 10, 30, and 50 GeV (right).

lengths of tracker material which precede the ECAL. As can be seen in Figure 6-2, significant radiation of even the majority of an electron's energy is more the rule than the exception. As such the reconstruction algorithms are optimized to cluster together all energy deposits in ϕ which are likely to have come from a single electron or photon. The result is that the picture of an electron in the ECAL is more a large swathe in ϕ than a localized deposit.

A supercluster in the ECAL barrel is reconstructed from a deposit in an individual crystal above the seed energy threshold. This is expanded to the two adjacent crystals on either side in η to form a 5 x 1 array. Then an attempt is made to expand this grouping by incorporating neighboring 5 x 1 arrays by searching in ϕ for crystals with additional energy deposits. The result is a cluster of clusters, or supercluster, with width of five crystals in η and variable ϕ extent.

The procedure is similar in the ECAL end cap, except that due to different crystal arrangements, particle multiplicities, and quantities of tracker material, the cluster ϕ width is five rather than one crystal.

Once a supercluster has been created, an attempt is made to find hits in the silicon tracker which are compatible in ϕ and z with the ECAL deposits. Although

the radiative energy losses in the tracker are well described by the Bethe-Heitler theory [59], the resulting non-Gaussian energy loss distribution is not properly handled by the Kalman filter track reconstruction algorithm. A specialized track fitting procedure for use with electrons has thus been devised which utilizes sums of Gaussian distributions to approximate the Bethe-Heitler formalism. This is called the Gaussian sum filter (GSF) technique.

Because the apparent curvature direction of an electron's track can change quite dramatically upon hard photon radiation, some care must be exercised in choosing the proper electron charge hypothesis. The apparent charge from the electron candidate's GSF track, the apparent charge of the nearest Kalman filter track, and the charge inferred from the change in ϕ between the first tracker hit and the supercluster give three complementary charge determinations. In practice the majority opinion of the three methods provides an effective combination of the information at hand.

The ability to correctly determine the kinematics of observed electrons is one of the main determinants of the final analysis sensitivity because of its effect on peak width in the invariant mass distribution. Although an electron's direction presents few complications and can simply be taken from its track at the point of closest approach to the beam line, significant effort is put into a variety of techniques for correcting and checking the electron energy determination.

The electron energy measurement is at heart a procedure for using the myriad variables which describe the particle's qualities in the detector to inform a decision on how to combine the supercluster's energy with the track momentum. These two determinations are to a large extent complementary because the tracker measurement suffers at high momentum as the track loses curvature, whereas the ECAL resolution is best at high energies when the effects of detector noise and intrinsic shower fluctuations are minimized.

We use a boosted regression tree to estimate the true electron energy based on raw supercluster energy, position in the detector, and a large number of variables describing shower shape and energy deposition in the electromagnetic and hadron calorimeters. In order to obtain the uncertainty on this energy, it is necessary to train

an additional regression tree on the uncertainty for each electron, in other words on the difference between the regression energy and the true energy.

A further regression tree is trained on the ECAL energy, tracker p_{T}, and their associated uncertainties in order to find the optimal combination of the two.

The two prior steps are purely Monte Carlo-based, because they depend on knowledge of the electron's true energy. It is thus necessary to apply a further set of corrections to account for residual differences between simulation and data. These stem mainly from tracker misalignment and from imperfections in the crystal transparency corrections. These discrepancies are evaluated by comparing the mass line shapes of the Z boson, J/ψ, and Υ resonances between data and Monte Carlo. The scale in data is then corrected to match that in simulation, and an additional Gaussian smearing is applied to Monte Carlo such as to match the resolution seen in data.

For electrons we also apply a basic minimum p_{T} cut, at 7 GeV, and require that the η magnitude be less than 2.5. We also apply the same impact parameter cuts as for muons.

Because muons have a tendency to plausibly fake the signature of an electron, we also require that electrons are at least a distance $\Delta R = 0.05$ from all muon candidates.

Finally, we must distinguish real electrons from both photons which have converted to electron-positron pairs and from jets of hadrons which happen to mimic electrons. We combat the first by requiring that the tracks associated with electron candidates not skip, or miss hitting, more than one tracker layer, since if the candidate started life at the interaction point as a photon it would not have produced hits in any layers until conversion. For the latter we use a boosted decision tree with input variables describing the degree of matching between supercluster and track, the radiative energy losses in the tracker, and a description of the electromagnetic shower shape. The decision tree is trained with electrons from simulated Z$\rightarrow$ *ee* events as signal, and jets in W plus jet events in data as background.

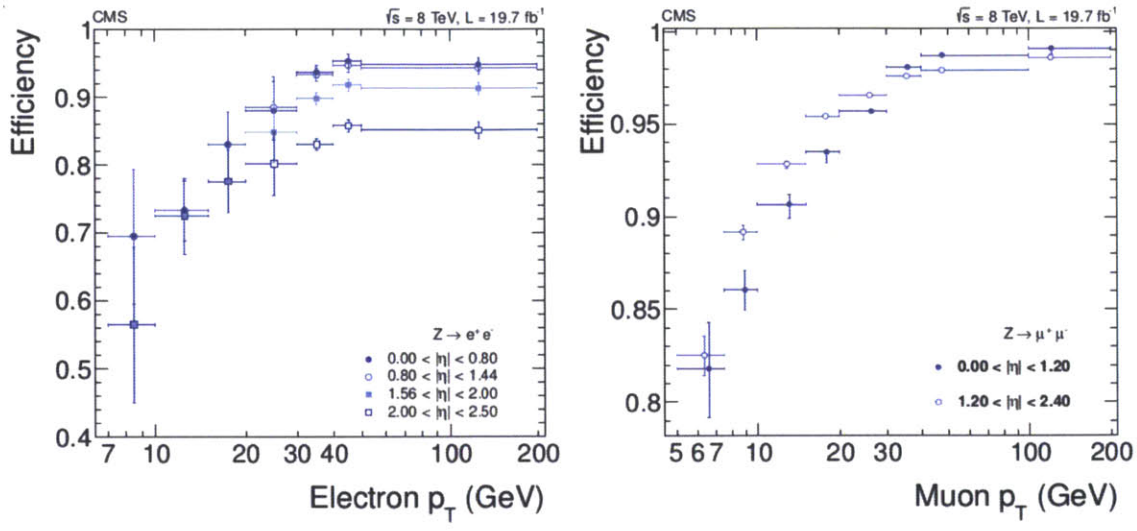

Figure 6-3: Lepton identification efficiency as measured in Z boson decays in data for electrons (left) and muons (right).

6.6 Lepton Efficiency Corrections

In order to account for inaccuracies in the simulation of the passage of leptons through the detector we measure the efficiency of our full lepton selection with the tag and probe method [60] in both data and Monte Carlo. A reasonably pure sample of Z$\rightarrow \ell\ell$ events is obtained by requiring one very well-identified lepton (the tag) and one lepton with minimal identification requirements (the probe) which together have an invariant mass near to the Z pole. By fitting the resulting dilepton mass distributions for the signal (Drell-Yan) yields in the cases where the probe has passed and where it has failed the lepton identification requirements one is able to extract the precise efficiency. The results of this procedure are shown in Figure 6-3.

Chapter 7

Event Selection

In this analysis we have the good fortune to be concerned with a region of phase space which is comparatively devoid of large backgrounds: the requirement of four charged leptons alone does much of the work of our selection by itself.

The first step in the selection is in the online triggering system (see Section 4.7). For the four lepton selection we mainly use double lepton triggers with asymetric p_{T} thresholds of 17 and 8 GeV which admit two muons, two electrons, or one of each. These triggers are supplemented with a triple electron trigger with p_{T} thresholds of 15, 8, and 5 GeV. The efficiency of this triggering strategy is found to be 98-99%. For lepton identification efficiency measurements (see Section 6.6) with low-mass resonances we also use triggers with lower p_{T} thresholds.

We then select events which contain four well-identified, isolated leptons with η magnitude less than 2.4 (2.5) and p_{T} greater than 5 (7) GeV for muons (electrons). At this stage we have eliminated essentially all QCD production processes and are left with predominantly electroweak processes, which as can be seen in Figure 2-2 affords a huge advantage in relative cross sections.

Given at least four such leptons, we first endeavor to build a single on-shell Z boson by finding the pair of oppositely charged, same-flavor leptons whose invariant mass is closest to the Z pole at 91.187 GeV. Here Z bosons are taken as on-shell for masses between 40 and 120 GeV. We then attempt to find a second Z from the remaining leptons. Since this Z will in general be off-shell, its mass is only constrained

to be between 12 and 120 GeV. If more than one is found, we choose that with the highest lepton sum p_{T}.

In roughly ten percent of the Z decays which we consider in this analysis, a final state lepton radiates a photon, and in order to correctly reconstruct the original invariant mass of the system we want to recover this photon's energy. We therefore identify photons nearby to selected leptons as candidates for inclusion in the lepton's four-momentum if, when included, they move the invariant mass of the Z candidate closer to the Z pole. However, because this final state radiation (FSR) is typically low energy and almost always collinear with the lepton, the standard supercluster reconstruction techniques for electrons do an excellent job of recovering virtually all of it. The FSR recovery algorithm thus has an impact only in the case of muons, which are however much less prone to radiate than are electrons, so the effect on the mass resolution is negligible. Because photons which are tagged as FSR are removed from the isolation sum of their associated leptons, however, the main effect of the algorithm is to systematically decrease the isolation values of selected leptons. The net effect of this is to increase yields by 2-3% for signal and 3-4% for background.

Once we have acquired two Z boson candidates, we require that at least one lepton has p_{T} greater than 20 GeV and at least one other has p_{T} greater than 10 GeV. In order to eliminate leptons from low-mass hadronic resonances, we also require that every oppositely-charged pair which can be made from our four leptons has invariant mass greater than 4 GeV.

Chapter 8

Signal and Background Models

8.1 Signal

We base our model of the Standard Model Higgs boson signal on the Monte Carlo techniques outlined in Section 2.6. The hard processes for gluon and vector boson fusion are modeled with the POWHEG NLO event generator [32]. The intermediate products from this step are then passed to PYTHIA [33], a general-purpose LO generator, for decay, showering, and hadronization, before being run through a GEANT [61] based simulation of the CMS detector. The control of high-p_{T} radiation emission in POWHEG has been tuned to reproduce the p_{T} distribution seen in higher-order calculations (see Section 2.3), and the underlying event parameters in PYTHIA have been tuned to the particular environment observed in LHC collisions. The vector boson and top quark associated production modes, meanwhile, use PYTHIA for both the hard event generation and for subsequent steps. The uncertainties discussed in Section 2.3 have been incorporated as normalization and shape uncertainties where appropriate.

8.2 Background

As discussed in Section 2.5, we model the backgrounds which are of electroweak origin, quark- or gluon-induced continuum ZZ production, using Monte Carlo simulation.

The processes which stem largely from QCD and involve hadronic phenomena, on the other hand, are modeled using data-based techniques.

8.2.1 ZZ Continuum

Quark-initiated ZZ continuum production is simulated with POWHEG, and all subsequent steps proceed as for the Higgs boson signal. The hard collision for production via gluon fusion is simulated with gg2zz [30], while the remainder proceeds as for Higgs boson signal.

8.2.2 Fakes

As discussed in Section 2.5.2, processes in which hadronic jets or non-prompt lepton sources mimic, or fake, a prompt electron or muon are more difficult to model. While the underlying hard interactions in Z plus jet, W and Z, or top quark pair production are theoretically well-understood, the process by which objects in these events can fake a real, prompt lepton while passing through the detector is much more difficult to simulate with a high level of fidelity. We thus take advantage of a convenient source of events in which these processes have been in effect perfectly modeled: real data events.

In order to obtain a model for these backgrounds, we begin with a sample which differs from the signal region only in that instead of four well-identified leptons, it has two well-identified leptons and two very loosely identified leptons, which are in practice usually jets. We then calculate the probability for these jets to pass our final lepton selection, and use this probability as a weight on the jet events in order to make a prediction for the number of events which have real jets identified as leptons and which thus end up in the final signal region.

The first task is to calculate the probability for, or rate at which, jets or other non-prompt sources fake leptons. This fake rate is obtained from the source of pure single jets which is nearest to hand and which most closely resembles the signal region: Z plus one jet events. In practice, in order to use these we must decide on a looser lepton

selection which, in events with a well-identified Z plus on loose lepton, will result in a sample dominated by jets. It must do the same in the region with one well-identified Z and two jets, which we will use to extrapolate into the signal region. While the definition of this selection is in principle arbitrary, it must be optimized (separately for electrons and muons) such as to be loose enough to give a large event yield for a reliable calculation, but must also be tight enough such that we do not encounter regions of phase space which are drastically different from the signal region in either kinematics or quark flavor content. In practice, we have a number of control regions in which to test the closure of the fake rate methods, and we increase yields as much as possible, by loosening the fake lepton requirements, until we observe significant disagreement in the closure tests. We then re-tighten the requirements to allow a decent safety margin.

We thus require the leptons from the single Z in this region to satisfy the selection requirements discussed in Chapter 7. The loose leptons must pass the same p_{T}, η and impact parameter cuts as tight leptons. Loose muons are also required to pass the same requirements as tight muons for compatibility between tracker and muon chamber tracks, and loose electrons must have fewer than two missing tracker hits (see Section 6.5). In order to further enrich the sample in Z plus fake lepton events, the invariant mass of the Z must be within 10 GeV of the Z pole, and, if the loose lepton can be paired with a lepton from the Z to make a pair which is flavor- and charge-compatible with J/Ψ decay, their dilepton mass must be greater than 4 GeV. In addition, the negative magnitude of the vector sum of all particles in the event, or missing transverse energy, is required to be less than 25 GeV. The fake rates resulting from this procedure are shown as a function of p_{T} in Figure 8-1.

We now want to use these fake rates in order to extrapolate from a sample with two real leptons and two fake leptons into the signal region, which has four leptons. We thus select events in data which have a single on-shell Z , made from well-identified leptons as described in Chapter 7, and two or more additional leptons which are only loosely identified (the 2P2F region). These loose leptons must have the same electric charge, and must satisfy the requirements which we used above in order to calculate

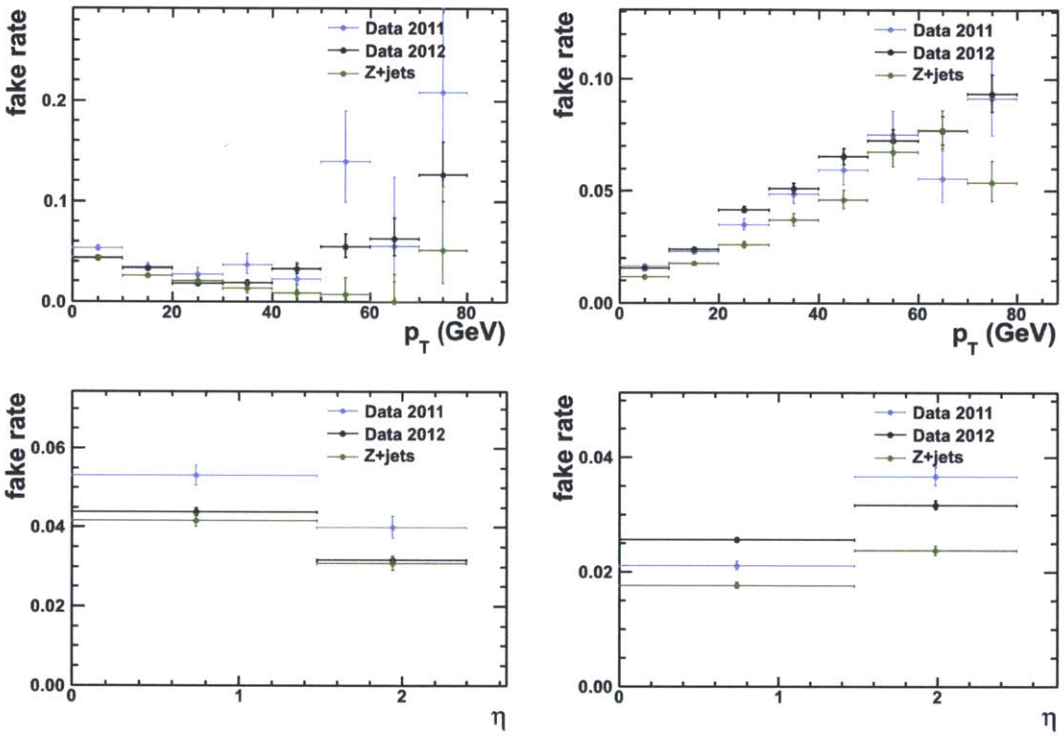

Figure 8-1: Fake rate f for fake leptons to pass the tight lepton selection as a function of p_{T} (top) and η (bottom), where "lepton" is in this context taken to mean an object passing the full lepton identification requirements outlined in the text. They are shown for muons (left) and electrons (right) and for 2011 data (7 TeV, blue), 2012 data (8 TeV, black), and Z + jets Monte Carlo (green). The Monte Carlo does not enter into the final analysis, and is shown only to emphasize the difference to and motivate the use of data-based methods for fake background estimation.

the fake rates, and we thus know the probability for each of these loose leptons to pass the final selection requirements. We then apply this probability as an event weight, with one power for each loose lepton, $f_3/(1-f_3)$ and $f_4/(1-f_4)$, where the denominator accounts for the fact that these fake leptons must fail the tight selection requirements. We also take from Monte Carlo the ratio of same-sign to opposite-sign events ($R_{\mathrm{os/ss}}$), and thereby arrive at the number of predicted fake events in the signal region

$$N^{\mathrm{fake}}_{\mathrm{signal\ region}} = R_{\mathrm{os/ss}} \sum_{i}^{\mathrm{2P2F_{ss}}} \frac{f_3^i}{1-f_3^i} \frac{f_4^i}{1-f_4^i}. \tag{8.1}$$

This method provides an estimate of backgrounds with a single Z plus jets and top quark pair production. As mentioned above, we do not expect the Monte Carlo samples for these processes to accurately model the fake lepton probability. Furthermore, because these processes must have two powers of the fake rate to appear in the signal region, a number of order a few parts in a thousand, it is computationally prohibitive to generate sufficient Monte Carlo events to provide an accurate estimate. Nevertheless, if we simply apply the full signal region selection to Z plus jet and top quark pair Monte Carlo samples, we obtain an estimate which agrees within its statistical errors, which are of order unity, with the data based approach.

Events with a W and Z , on the other hand, are neglected in this method. Because they require only one power of the fake rate to enter the signal region, however, we can use Monte Carlo to provide an estimate of their contribution. This Monte Carlo-based approach gives an expected yield consistent with zero, which is corroborated by the results of an alternative method outlined below.

A number of tests are performed in order to assess the systematics on this fake rate extrapolation. These fall in two categories, based on whether they depend on Monte Carlo simulation or are purely data-based. An example of the former is shown in Figure 8-2, where we evaluate the composition of the fake extrapolation sample. An example of the latter is shown in Figure 8-3, where we show comparison of prediction and observation in a region which differs from the signal region in the inversion of charge and flavor requirements of the leptons from the second (off-shell) Z.

We also perform a cross check with an alternative fake estimation method which uses orthogonal data samples. This method extrapolates to the signal region mainly from a region with three passing and only one failing lepton (the 3P1F region), and thus only one power of the fake rates. Here there are two tight leptons compatible with an on-shell Z decay, and one passing and one failing lepton which are charge- and flavor-compatible with an off-shell Z decay. This region, however, includes processes which must be removed from the final fake background estimate. First, the 3P1F sample contains ZZ continuum events in which one real lepton fails the tight selection. This must be removed because it is already included in our Monte Carlo samples. We estimate the number of ZZ continuum events in the 3P1F region, N^{ZZ}_{3P1F}, from Monte Carlo. Second, events from processes with two fake leptons (the 2P2F region) will contribute twice to the 3P1F region, once for each of the failing leptons. Their contribution must thus be divided by two in order to remove this double counting. We accomplish this by applying one power of the fake rate to each 2P2F event and subtracting this sum from the 3P1F estimate. The final fake estimate with this method is thus

$$N^{\text{fake}}_{\text{signal region}} = \left(1 - \frac{N^{ZZ}_{\text{3P1F}}}{N_{\text{3P1F}}}\right) \sum_{i}^{N_{\text{3P1F}}} \frac{f_4^i}{1 - f_4^i} - \sum_{i}^{N_{\text{2P2F}}} \frac{f_3^i}{1 - f_3^i} \frac{f_4^i}{1 - f_4^i}. \tag{8.2}$$

This estimate, as for the first method, includes contributions from Z plus jet and top quark pair events. It also, however, includes events with both W and Z bosons. The fact that it gives the same overall estimate for fake contributions to the signal region provides additional evidence that the Monte Carlo based WZ estimate of zero can be trusted.

The synthesis of the various control studies is a set of shape systematics which are propagated through the entire analysis (see Figure 8-4), and a set of normalization uncertainties which are applied to the fake background in all subsequent steps which correspond to 20% (4e), 40% (4μ), and 25% ($2e2\mu$).

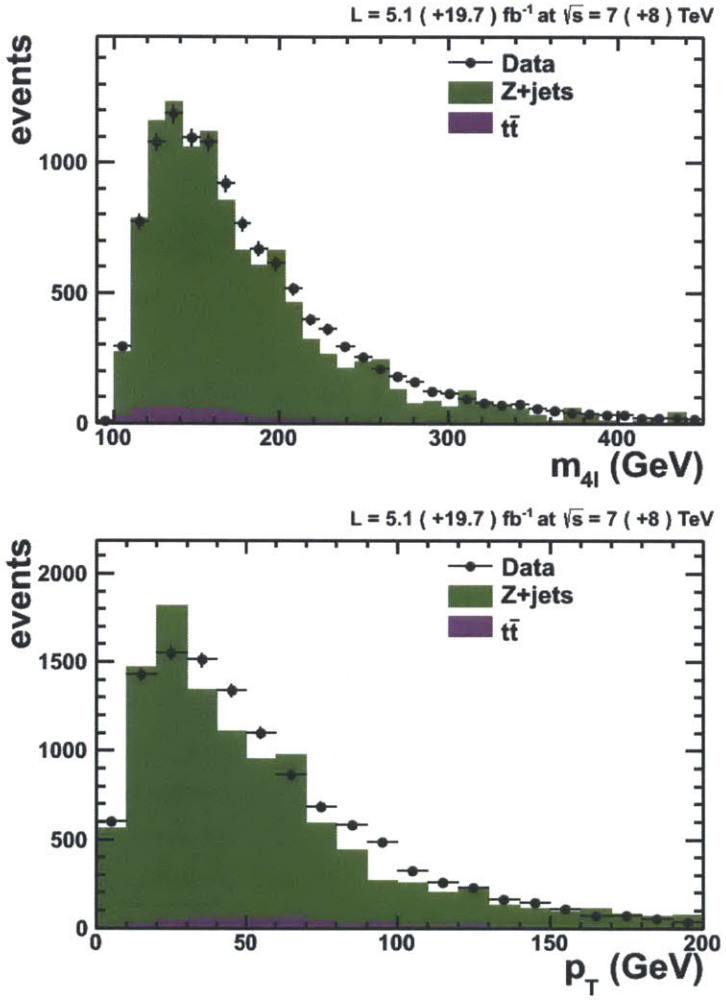

Figure 8-2: Kinematic distributions in the fake control regions for data and Monte Carlo for four lepton mass (top) and four lepton p_{T} (bottom). For a discussion of the systematics which are used to cover the discrepancies see the text.

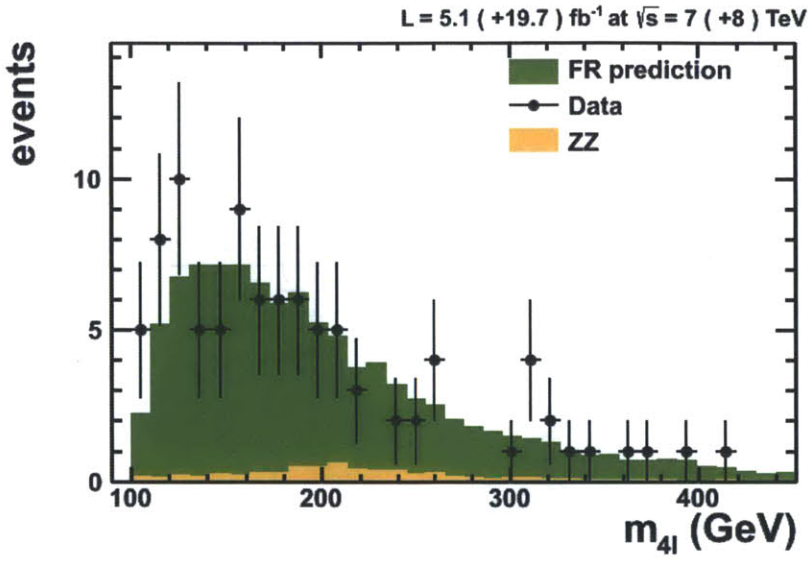

Figure 8-3: Fake method closure test in a sample which differs from the signal region in the inversion of charge and/or flavor requirements such as to exclude the signal region. The fake rate (FR) prediction is an extrapolation from the region with two loose leptons to that with two tight leptons, while the ZZ contribution is the Monte Carlo based prediction of ZZ continuum out of the signal region due either to charge mismeasurement or to lepton misidentification.

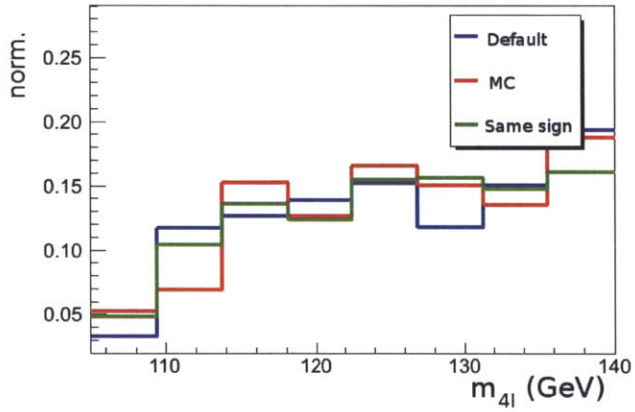

Figure 8-4: Example of the shape differences between the standard fake rate calculation (blue), a fake rate extrapolation done on Monte Carlo (red), and that performed on a sample with reversed lepton charge requirements on the off-shell Z (green). Similar shapes are calculated for ZZ p_{T} and Boosted Decision Tree output (see later sections), and propagated through each step of the analysis as shape uncertainties.

	Yield
ZZ continuum	23.9
Fakes	8.6
ggH	19.3
vbfH	1.6
VH	0.9
ttH	0.1

	Yield
Backgrounds	31.4
Signals	22.1
Data	50

Table 8.1: Expected and observed yields in the four lepton mass range 106 to 141 GeV for each individual subprocess (left) and for all backgrounds, all signals, and data (right).

8.3 Distributions and Event Yields

We now show the signal and background predictions after application of the full event selection overlaid with the observed data for a number of kinematic distributions. Four lepton mass, $m_{4\ell}$, is the most immediately interesting. As can be seen in Figure 8-5 a prominent excess is observed near 126 GeV. The significance of this excess and its compatibility with the predicted Standard Model cross section and properties are discussed in the following chapters. The expected and observed yields are shown in Table 8.1. We also show the four lepton mass region around the Z pole in Figure 8-6. This is populated with Z$\rightarrow 4\ell$ events and the agreement between Monte Carlo and data gives additional confidence that we have sufficient understanding of the background normalization and lepton energy scale corrections.

We also show the p_{T} and η of the ZZ system in Figure 8-7, and the single boson masses in Figure 8-8. The p_{T} distribution is investigated in more detail in Chapter 11.

Although we see in the previous plots that the observed signal has a mass near 126 GeV, we perform a cursory mass measurement in order to more accurately estimate this parameter. In order to do this we fit the full signal and background models to data with separate widths in the 4e, 4μ, and 2e2μ channels but with common floating signal strength and Higgs boson mass. The results of this fit are shown in Figure 8-9, from which we obtain a Higgs boson mass of 125.6 $\pm$ 0.6 GeV, where the uncertainties are purely statistical. This compares favorably to the value

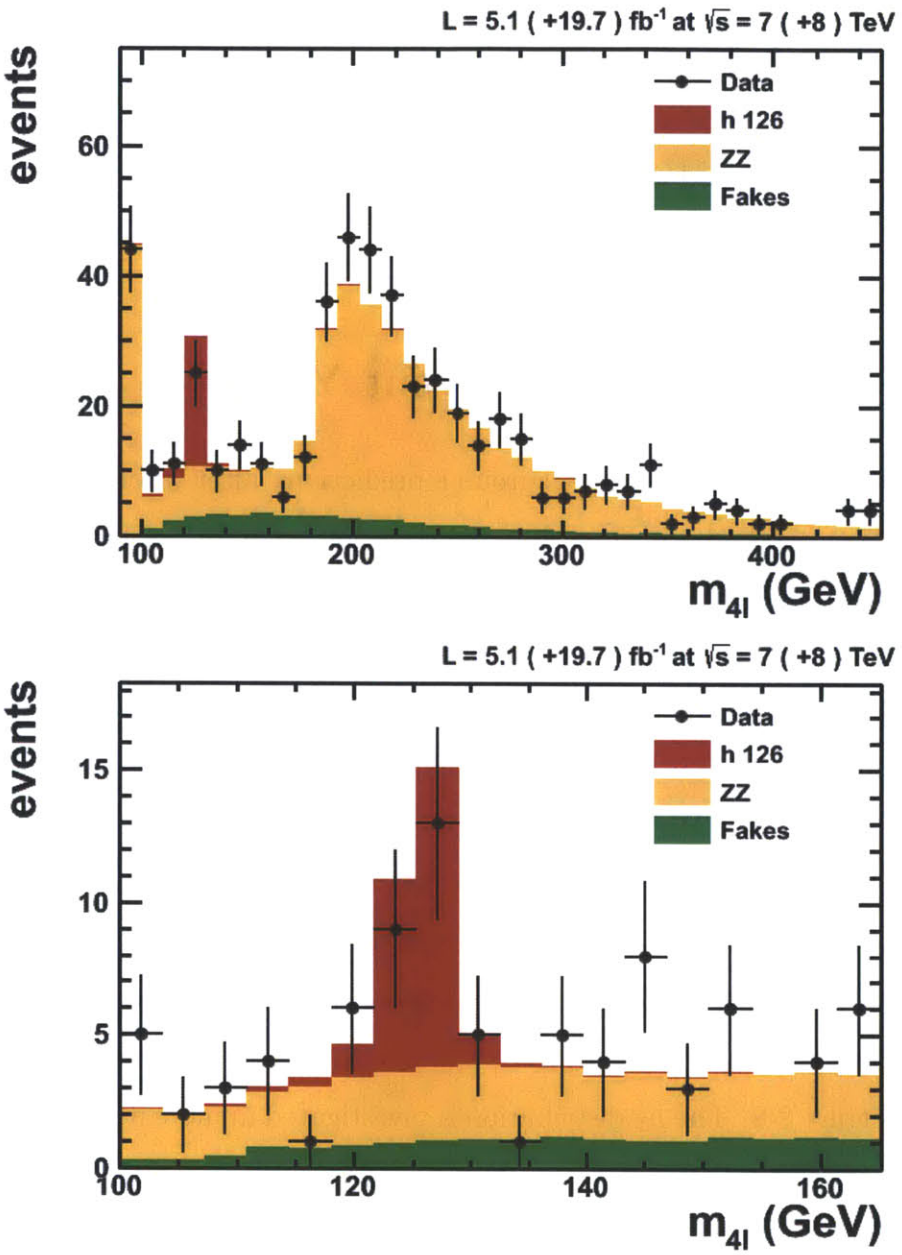

Figure 8-5: Four lepton mass distributions after the full event selection from 100 to 600 GeV (top) and zooming in to the excess around 126 GeV (bottom). The signal and background models described in the text are shown as filled histograms behind the combination of 7 TeV and 8 TeV data.

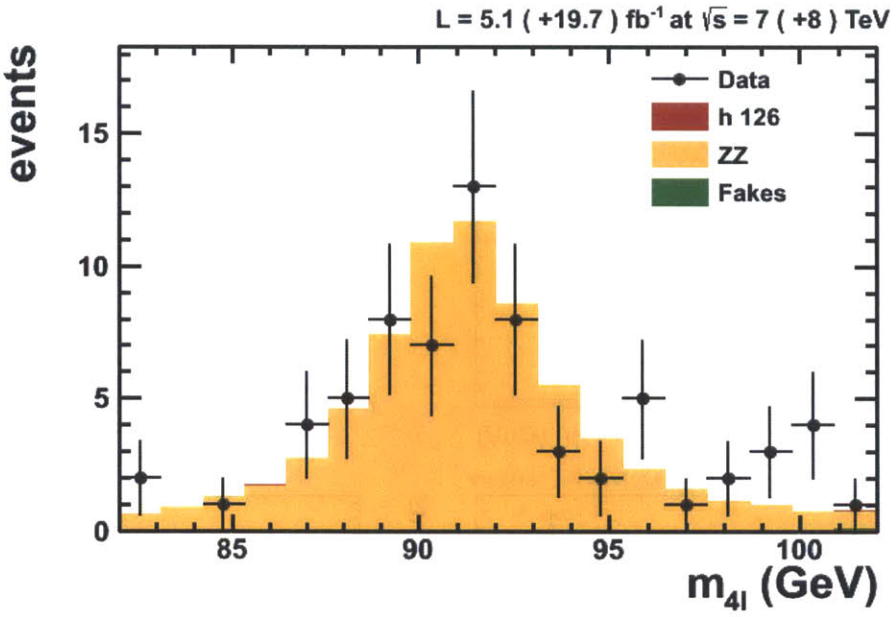

Figure 8-6: Four lepton mass distribution near the Z pole after the full event selection. This region is populated with Z$\rightarrow 4\ell$ events and it is used as a control region due to its proximity to the signal region, which is at slightly higher mass.

of 125.6 $\pm$0.4(stat) $\pm$0.2(syst) GeV obtained with a three-dimensional fit including per-event mass uncertainties in the official CMS publication [62].

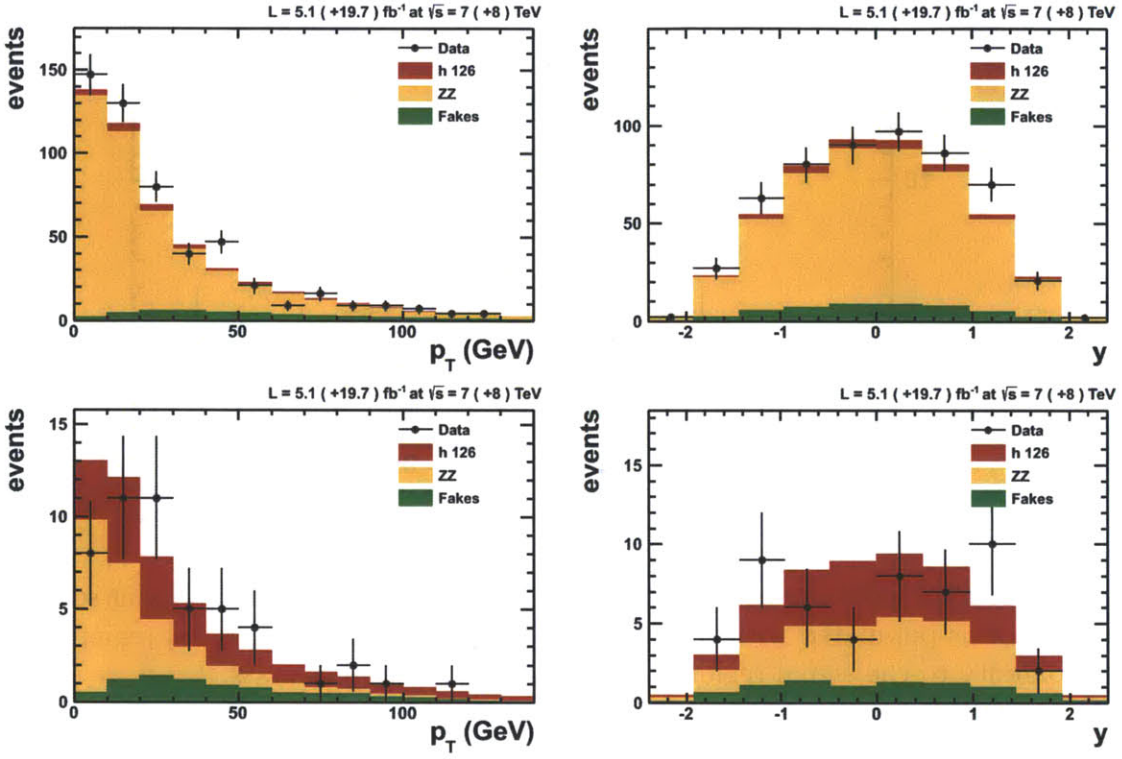

Figure 8-7: Four lepton p_T and rapidity distributions after the full event selection in the mass range 100 to 1000 GeV (top) and 106 to 141 GeV (bottom).

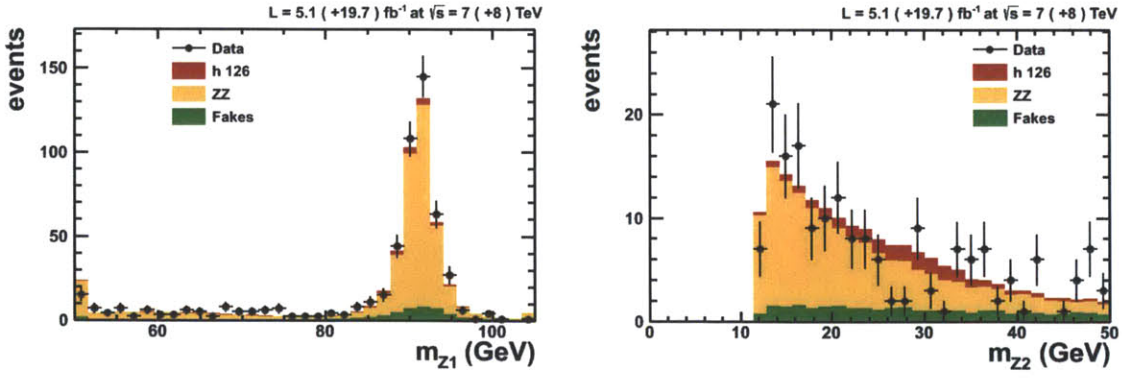

Figure 8-8: Single boson mass distributions after the full event selection.

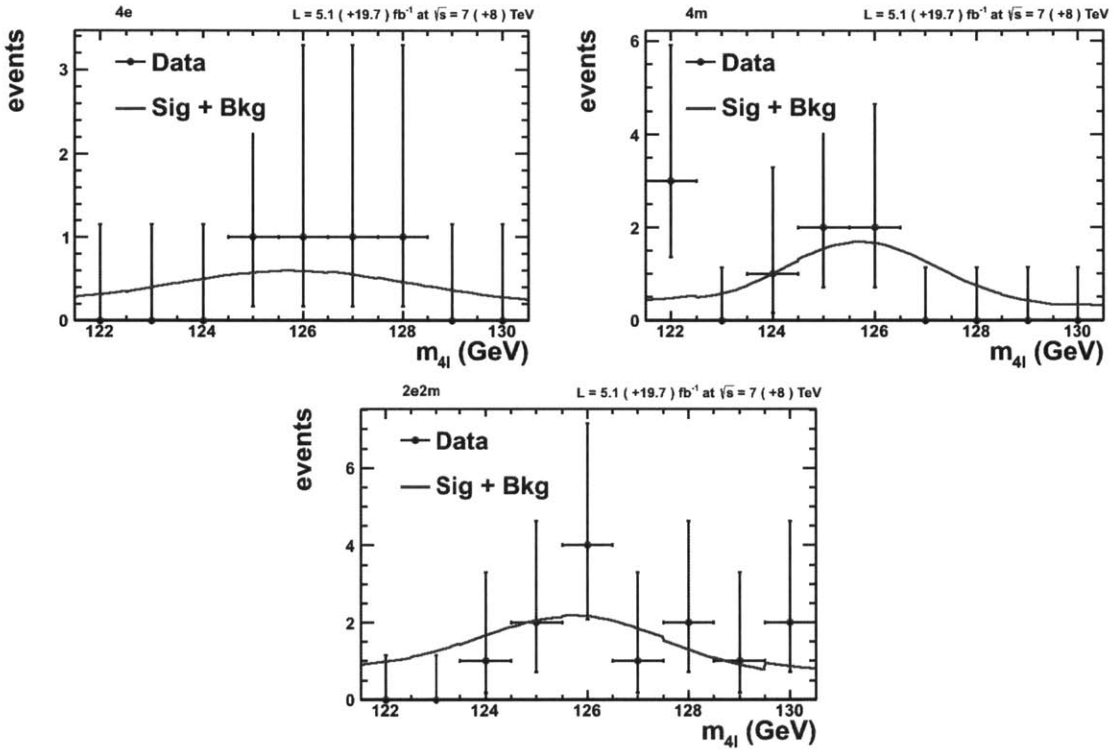

Figure 8-9: Results of the simultaneous mass fit in the 4e (top left), 4μ (top right), and 2e2μ (bottom) channels. Each channel has a different width, while the Higgs boson mass and signal strength are the same across all three channels.

Chapter 9

Signal Discrimination

The distributions in Figure 8-5 indicate the excellent discrimination between Higgs boson signal and both the ZZ continuum and Fake backgrounds which the four lepton mass provides. However, there is a wide variety of extra information in the event which we can use in order to further separate the processes.

9.1 Full Event Kinematics

This discrimination stems from two main sources. First, the Standard Model Higgs boson will be produced via different mechanisms than either of the two backgrounds, and this will manifest itself as different kinematics for the four lepton system as a whole. Secondly, the difference in intermediate states causes differences in the decay kinematics of the four leptons. Taken as a whole, this kinematic information provides a powerful means of distinguishing between the various processes.

We choose only one possible parametrization of this information into measurable quantities, however, the particular choice of variables is in the end immaterial as we will pass them to a boosted decision tree. We make a rough distinction between those variables which are associated with the diboson rest frame and are shown in Figure 9-1:

- $\cos\theta_1$: angle between Z_1 decay axis and the ZZ system (in Z_1 frame)

- $\cos\theta_2$: angle between Z_2 decay axis and the ZZ system (in Z_2 frame)
- Φ: angle between the Z_1 and Z_2 decays planes (in ZZ frame)
- Φ_1: azimuthal angle between direction of Z_1 and the pp collision axis (in ZZ frame)
- $\cos\theta^\star$: polar angle between direction of Z_1 and the pp collision axis (in ZZ frame)
- m_{Z1} and m_{Z2}

and those which depend on the boost of the ZZ system as a whole, which are shown in Figures 9-2 and 9-3:

- pT of Z_1, Z_2, and ZZ system (rescaled by $m_{4\ell}$)
- Three-vector dot products: $\vec{Z_1} \cdot \vec{ZZ}$ and $\vec{Z_2} \cdot \vec{ZZ}$ (rescaled by $m_Z m_{4\ell}$)
- $\Delta\phi$ (azimuthal) angle between ZZ and Z_1, Z_2

The boost variables are rescaled by the mass scales indicated above in order to reduce correlations with the four lepton mass.

One of the main advantages of decision trees is that they excel at disentangling complex correlations between variables. For the sake of documentation, however, we show the correlations between input variables in Figure 9-4

These figures contain only the three processes which dominate the analysis: Higgs boson production via gluon fusion (ggH), quark-induced ZZ continuum (qqZZ), and fake backgrounds. The behavior of the other processes in the analysis (Higgs production via vector boson fusion and associated production, gluon-induced ZZ continuum production) has been checked and determined to be of negligible impact on the performance.

We leave the four lepton mass out of these discriminants and instead incorporate it as a second dimension in the final fits for two reasons. First, the mass by itself provides an extremely robust cross check for our background normalization procedure:

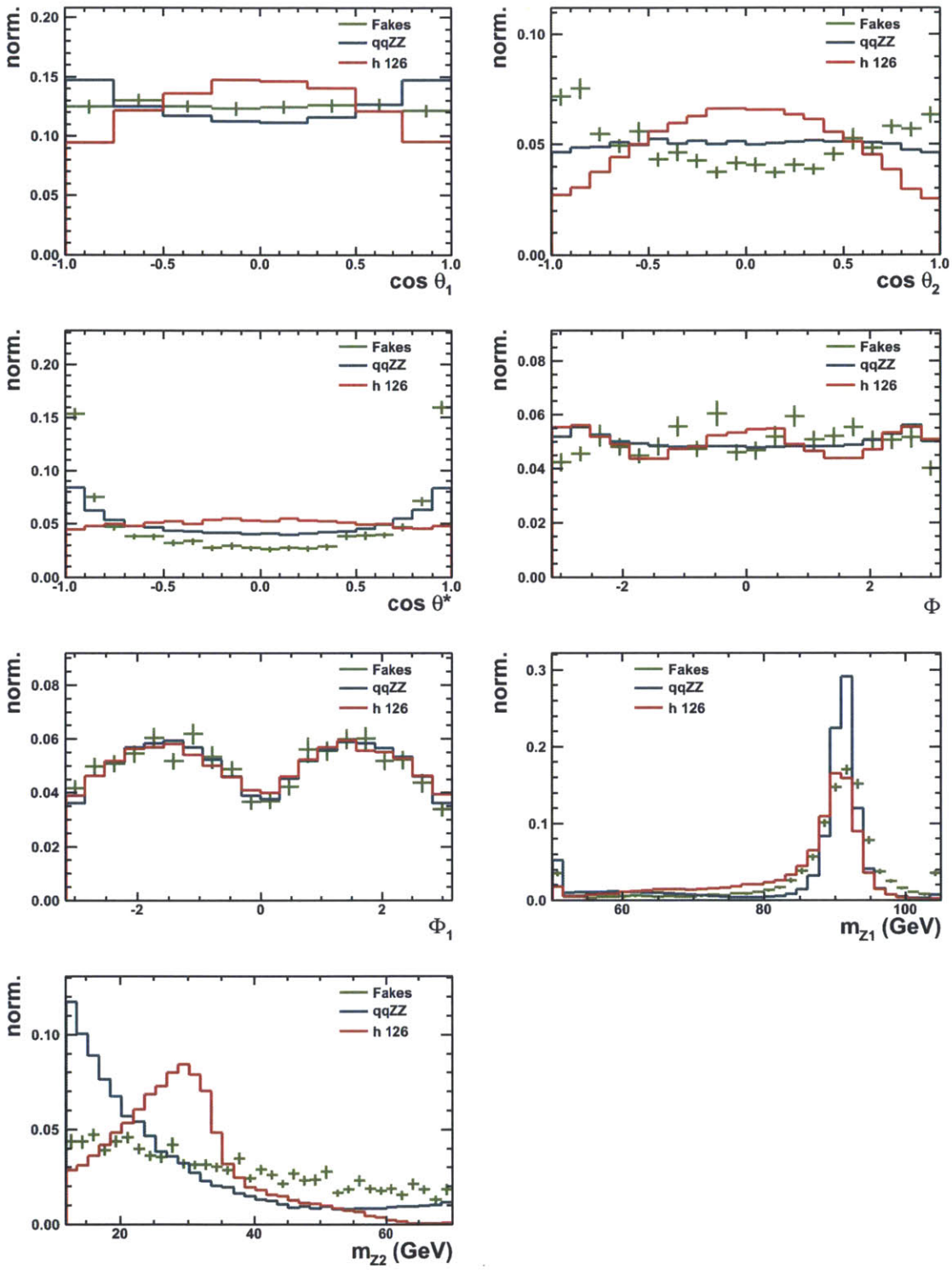

Figure 9-1: Kinematic variables which depend primarily on the decay in the diboson rest frame.

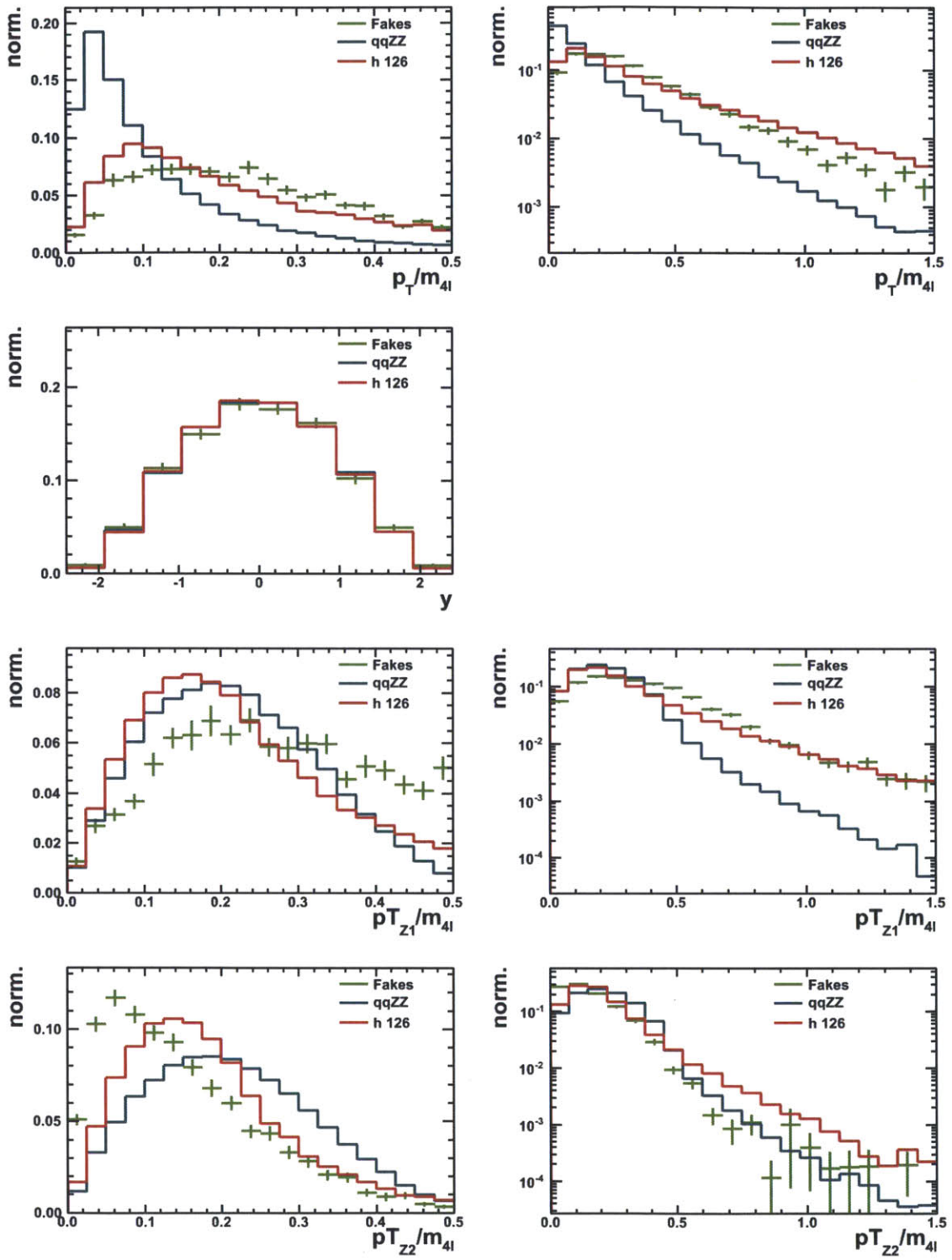

Figure 9-2: Kinematic variables which depend on the boost of the diboson system with respect to the detector in linear (left) and logarithmic scale (right). Continued in Figure 9-3.

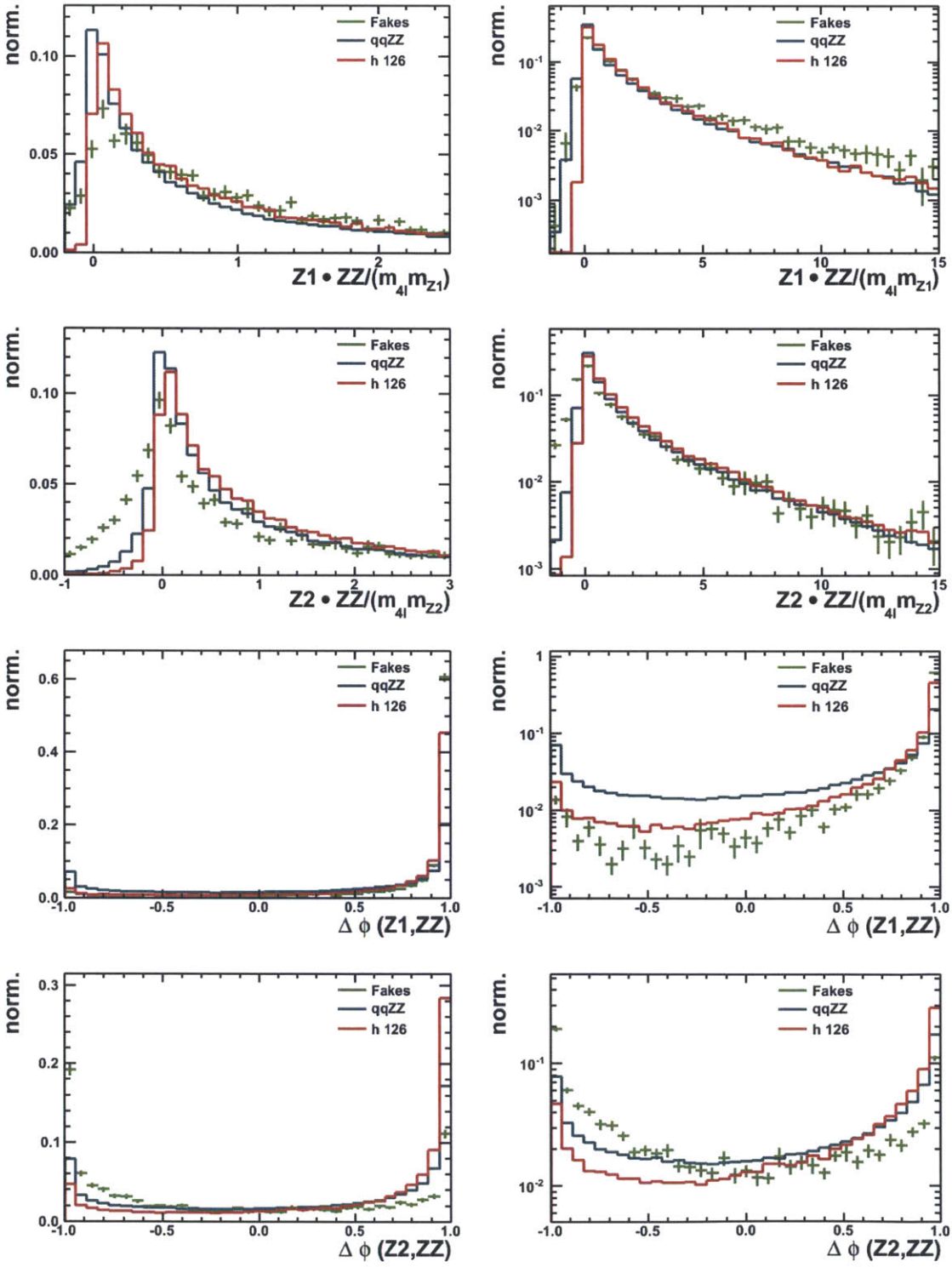

Figure 9-3: Kinematic variables which depend on the boost of the diboson system with respect to the detector in linear (left) and logarithmic scale (right). Continued from Figure 9-2.

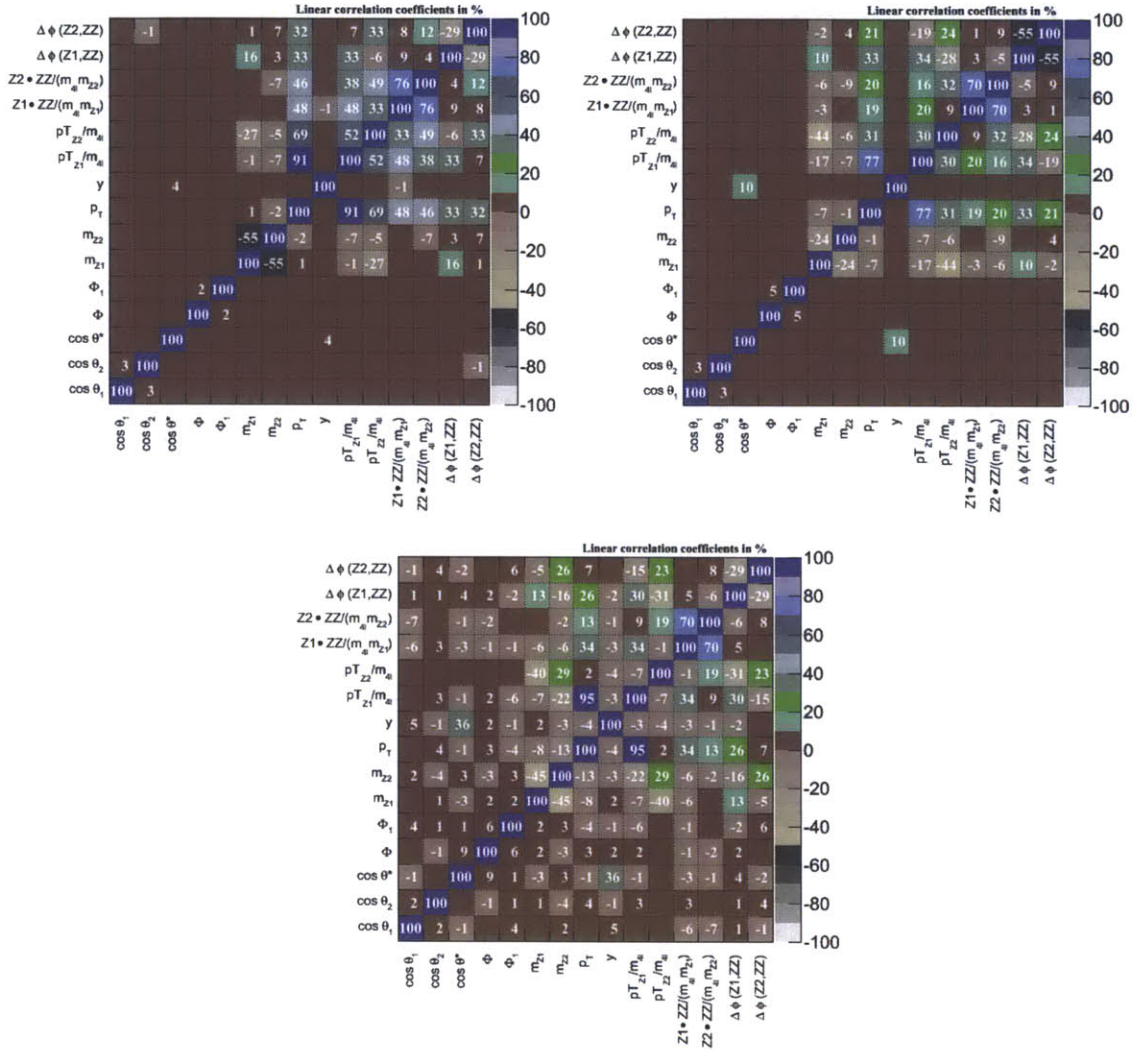

Figure 9-4: Linear correlation coefficients between each of the input variables used in the BDTs for signal (top left), qqZZ (top right), and fakes (bottom).

we have great confidence that the background processes have no significant structure underneath the signal peak. Thus if our background yield estimates were to be incorrect, this would be rectified when they are floated in the final fit. While we also have great confidence that our Monte Carlo accurately describes the full event kinematics and their correlations, it is deemed prudent to leave the mass separate. Second, since fitting to the mass as a second dimension incorporates the full mass information and its correlations to the other variables, we do not anticipate a significant performance reduction in taking this approach. We have tested this by including the four lepton mass in a discriminant and verified that the performance does not improve. In Chapter 12, on the other hand, we include the mass in the signal-background discriminant because we are at that point comparing signal hypotheses and are thus less interested in background control, and to avoid the practical difficulties inherent in moving to a three dimensional fit.

9.2 Decision Trees

Given this wealth of additional information, we must find a way to incorporate it into the rest of the analysis. The simplest way to do this is to combine all of the new variables into one additional variable. Instead of performing the final fit in one dimension with the four lepton mass, we will then fit in two dimensions using the new variable as the second axis. As described in Section 5, boosted decision trees (BDTs) provide a highly-evolved means of combining a number of variables into one final discriminating variable, and here we take this approach.

As described in the previous section, the three most important processes are ggH, qqZZ, and fakes. In building the decision trees, we can either train against the three processes simultaneously, in which case we obtain three discriminants, each optimized to push one process to the right side of a plot and the other two to the left, or we can lump the two backgrounds together into one category such as to obtain a single discriminant between all signal and all background processes. We investigate both approaches below.

For the ggH and qqZZ processes we produce dedicated training samples in order to provide maximum statistical power and to avoid training on a sample which is used for the statistical analysis. The training is performed in mass windows of about 10 times the larger of the detector resolution and the Higgs width. For a Higgs mass hypothesis of 126 GeV, for instance, we require four lepton mass between 106 and 141 GeV. Within each of these windows the dedicated training samples have several hundred thousand events. Data-based methods, however, are necessary in order to arrive at accurate predictions for fake backgrounds. This leads to the central problem with training against fakes: the limited number of events available in fake control samples. We have investigated several approaches to this problem. First, we can use the sample with two failing leptons which is used to extrapolate to the signal region (see Section 8.2.2). This sample has the advantage of giving excellent representation of the kinematics of the signal region, but has only a few thousand events, and its use in training necessitates reducing its statistics even further by separation into a training sample to train the BDT and a testing sample to use for the rest of the analysis. Another approach is to use the sample with two failing leptons with inverted charge and flavor requirements. This avoids both problems with the first approach, giving roughly three to five times as many training events, and any differences in the kinematic distributions are smaller than the current statistical uncertainties. In practice we find that both of these approaches yield similar results, so we show only the latter for illustration purposes in Figure 9-5. We also note that a poor choice of training samples affects only the optimality of the BDT and not its propriety. In other words if one were to train on samples which were utterly inappropriate the analysis would lose sensitivity but would not be subject to any additional biases.

We must also decide how to partition the various signal and backround processes into categories for the BDT output. The simplest choice is to have a signal category and a background category. However, we may also configure the BDT implementation to train between more than two categories and thus to produce more than one output discriminant. This latter possibility is generally called a multiclass BDT, and for instance if we give three categories, the output will consist of three discriminants, one

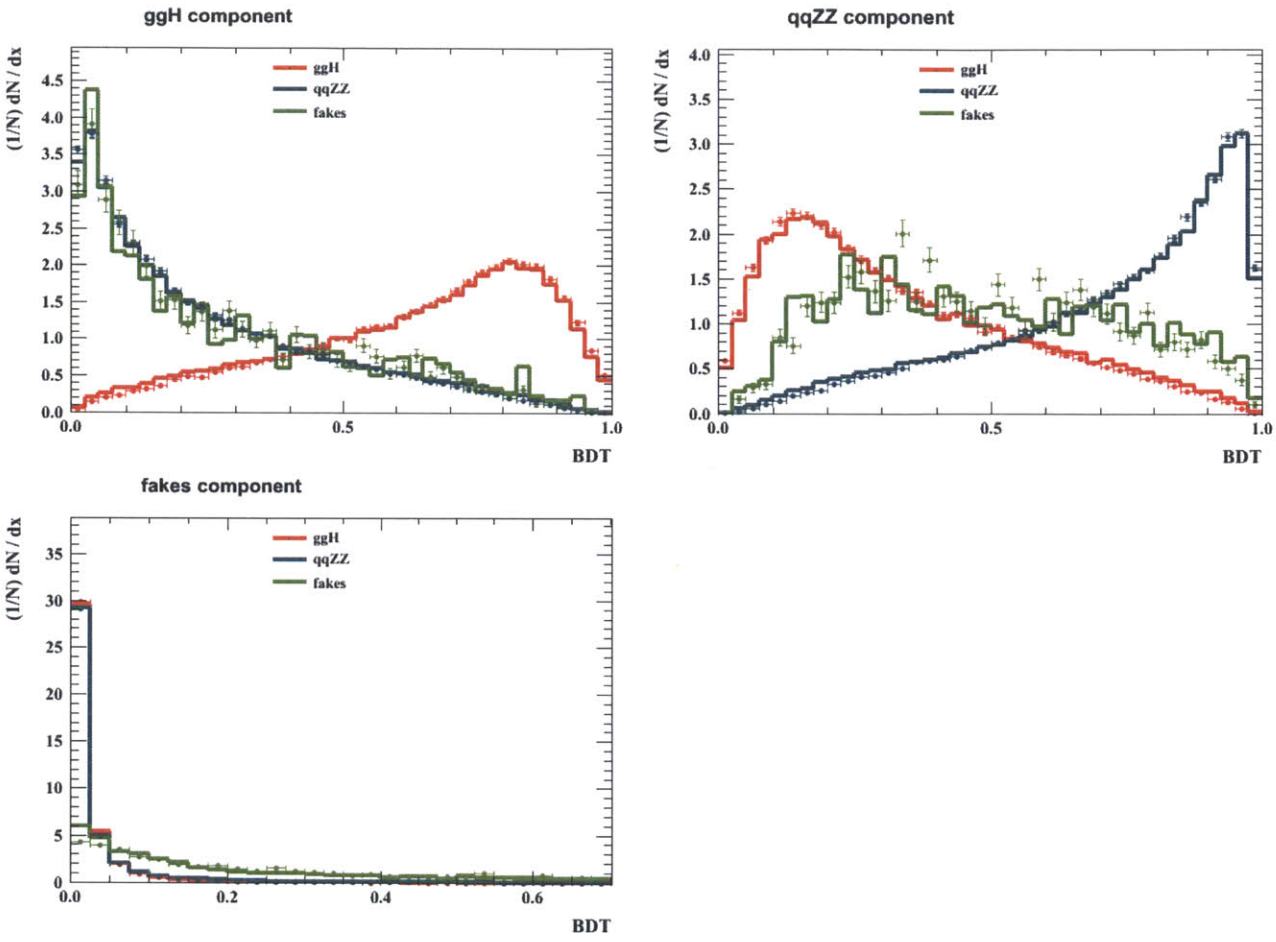

Figure 9-5: BDT output when training a multiclass BDT simultaneously with ggH, qqZZ, and fakes. The procedure results in three discriminants, optimized for separation of ggH (top left), qqZZ (top right), or fakes (bottom left).

optimized for discriminating between each pair of categories. Note that only two of these three are independent. This approach is shown in Figure 9-5.

Since we are only interested in separating signal from all backgrounds, in order to incorporate this multiclass BDT into the final statistical analysis it is in principle enough to use the component with both backgrounds pushed toward the left of the plots in Figure 9-5, in other words the top left in the figure. However, in order to use the full information in the discriminants one must choose at least two of the three discriminants, so we implement both a two-dimensional fit with four lepton mass for the former case, and a three-dimensional fit for the latter.

As mentioned above, the other means of incorporating both fake and qqZZ back-

grounds into the training is to mix them together into one sample. In order to do this we weight the events in both types of background to achieve the proper ratio of the two. We compare the performance of this method to the first method in Figure 9-6. The figure also shows a BDT trained only on qqZZ background. We see that those trained explicitly against fakes perform 10-20% better against this background. We also note that the three-class BDT performs somewhat more poorly against fakes than does the two-class BDT where qqZZ and fakes are mixed together. From extensive testing of combinations of training parameters and samples we can say that this is likely due to differences in training sample sizes. In other words the fact that the ggH and qqZZ samples are much larger than the fake sample complicates their partitioning during the tree-building process, with the result that the multiclass BDT in general performs a few percent worse. We would in any case prefer the two-class method based purely on its simplicity, so in further chapters we use the two-class BDT exclusively.

The curve labelled "mela" in Figure 9-6 gives the first indication that we take a different approach to that in the official CMS paper [62]. The authors of that paper use matrix element methods to find a leading order analytic approximation to the angular decay distributions in order to build a likelihood ratio-based discriminant. We use decision trees because within the matrix element method it is difficult or impossible to add boost information, which necessitates discarding much of the available kinematic information. Also, the analytic expressions in the matrix element discriminant cannot be modified to account for the full detector simulation, and thus in practice this discriminant achieves its initial optimality only on generator level quantities. Finally, as fake processes are not included in the matrix element discriminant, its performance against this background depends on the extent to which qqZZ and fakes happen to resemble each other. As can be seen in Figures 9-1, 9-2, and 9-3 this is a poor assumption.

We note that although efficiency curves as shown in Figure 9-6 give an excellent indication about relative performance between different discriminants, the only performance which we care about is in the final p-values and thus in ultimate signal

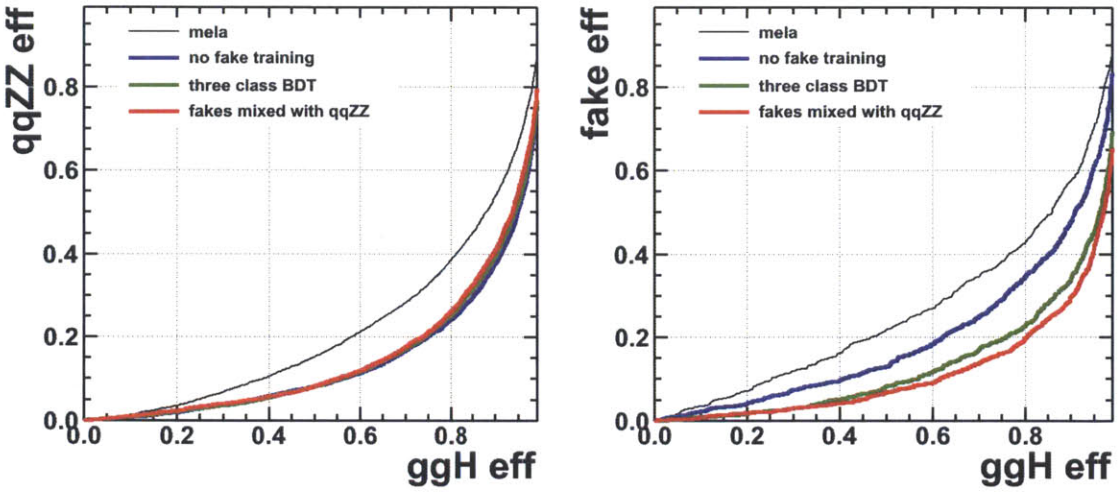

Figure 9-6: Signal (ggH) vs background efficiency for various discriminants against qqZZ (left) and fake backgrounds (right). BDTs are shown which do not train explicitly against fakes (blue), which use three classes in a multiclass training (green), and which mix fake training events with qqZZ (red). The kinematic discriminant (mela) used in the CMS publication [62] is also shown (black).

separation. As such, while we show efficiency curves for illustration, all decisions about which method to pursue are based on propagating the discriminants through the full statistical analysis. For example, the gains of 10-20% from using a BDT trained explicitly against fakes in Figure 9-6 translate into 5-10% gain in expected significance in the statistical analysis.

Similarly, the improvement of about 20% in the efficiency curves compared to the mela matrix element discriminant translates to an 8% improvement in the final significance. This improvement in performance stems partly from the explicit inclusion of fake backgrounds in the BDT training, and partly from the information in the additional boost variables which are fed into the BDT.

The end result of this method is then one additional variable with which we will supplement the four lepton mass in the statistical fit. The shapes of signal and background processes for this final variable are shown in Figure 9-7, and stacked histograms of the same can be found in Figure 9-8.

In the remainder of the analysis, we use the BDT which is trained with the fake and qqZZ backgrounds mixed together.

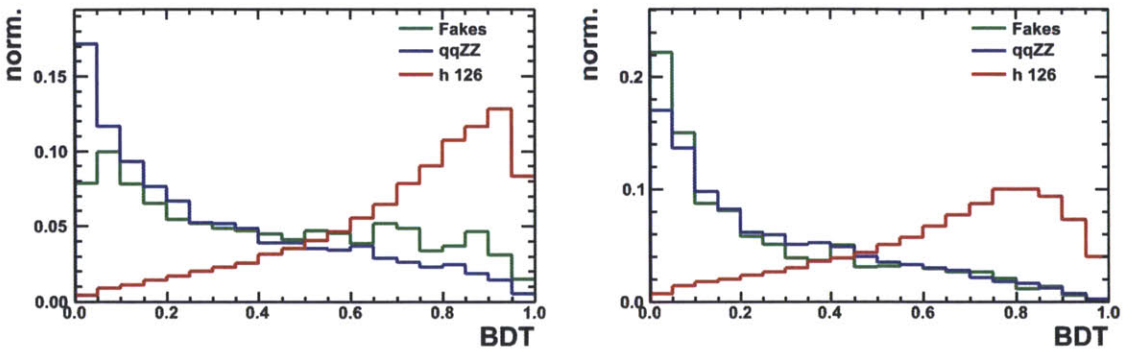

Figure 9-7: BDT output shapes for the BDT which includes only the qqZZ background in training (left) and that which includes fakes as well (right) in the mass window 106 to 141 GeV. The low-yield processes which were excluded from training are included here: the ZZ (blue) line contains ggZZ, and the h 126 (red) line includes vbfH and associated Higgs production.

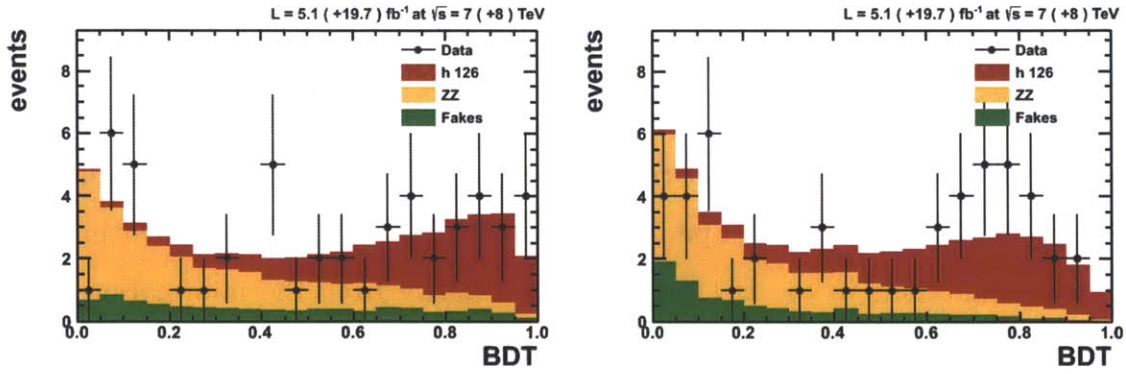

Figure 9-8: BDT output distributions for the BDT which includes only the qqZZ background in training (left) and that which includes fakes as well (right) in the mass window 106 to 141 GeV. The low-yield processes which were excluded from training are included here: the ZZ (blue) line contains ggZZ, and the h 126 (red) line includes vbfH and associated Higgs production.

Chapter 10

Statistical Interpretation

10.1 Introduction

The next step in the analysis of the distributions shown in Chapter 8 is to quantify the statistical significance of the excess seen near a mass of 126 GeV. To do this one in general defines a test statistic which can discriminate between the background-only and signal plus background hypotheses. The simplest, and indeed an optimal, choice is known as the likelihood ratio [63]. For the case of a single counting experiment, the likelihood ratio would take the form of a ratio of Poisson distributions in the expected number of signal events s, the expected number of background events b, and the observed event count d

$$q = \frac{e^{-(s+b)}(s+b)^d}{d!} \; / \; \frac{e^{-b}b^d}{d!}. \tag{10.1}$$

In order to generalize this to the combination of experiments and to distributions of events in various discriminatory variables, one simply multiplies together the ratios for each experiment and bin, such that the total likelihood ratio is the product over all experiments and bins.

In the case that we fail to observe a large excess, for instance, we would then calculate our confidence level in excluding the signal plus background hypothesis CL_{s+b} as the probability for the data, under the signal plus background hypothesis,

to fluctuate down to the level which we observe, q_0,

$$\mathrm{CL}_{s+b} = P_{s+b}(q \leq q_0). \tag{10.2}$$

10.2 Quantification of an Excess

In practice, however, the simple procedure outlined above must be modified in order to achieve a measure of statistical propriety. Its principle drawback is that a downward fluctuation of the background can be interpreted as the exclusion of all values of signal, including zero. In other words we may end up claiming to have excluded hypotheses to which we in fact do not have sensitivity. The most straightforward way to rectify this is to normalize by the confidence level in the background-only hypothesis. This has the effect of reducing our confidence in the signal exclusion in cases where we would tend to exclude the background only hypothesis as well, and is called the modified frequentist approach [64]

$$\mathrm{CL}_s = \mathrm{CL}_{s+b} \ / \ \mathrm{CL}_b. \tag{10.3}$$

The trade-off which one makes here in order to protect oneself against background fluctuations is an increase in over-coverage, where coverage is the probability that the interval of interest contains the true parameter value. For instance, if we make an observation in the heart of the background-only expectation, with $\mathrm{CL}_b = 0.5$, the CL_s will be twice the value of CL_{s+b}.

In a real experiment, in addition to the statistical uncertainties on event counts which are by construction included in the treatment above, we must also introduce systematic uncertainties. These systematics describe quantities such as our lack of knowledge of the theoretical cross section and the experimental momentum resolution, and necessitate further modifications to the test statistic. The systematic uncertainties are included by means of a set of parameters collectively denoted θ whose values do not necessarily interest us by themselves, but which must be estimated in order to arrive at the quantity in which we are interested. They are thus called nuisance

parameters, and because they affect the expected signal and background yields, these yields are now written as functions of the nuisance parameters: $s(\theta)$ and $b(\theta)$.

In order to allow their inclusion in the overall likelihood function in a factorized form, the uncertainties are assumed to be either fully correlated or fully uncorrelated with each other. In practice this is not a particularly onerous restriction as partially correlated uncertainties can generally be separated into correlated and uncorrelated pieces.

Starting from a default value $\tilde{\theta}$ for each nuisance parameter, we also require a distribution function which represents our best knowledge of its true value. These priors are denoted $\rho(\theta|\tilde{\theta})$, and the simplest method for their incorporation is to use the test statistic as above, but then to modify $s(\theta)$ and $b(\theta)$, according to their distributions, before the generation of each toy experiment during pseudo-experiment generation. While this approach was used at past experiments, its calculation is quite computationally intensive, and we thus introduce an alternative approach which lends itself to practical approximation.

We begin by rewriting the likelihood ratio in 10.2 with $\mathcal{L}$ in place of the explicit Poisson notation (which also allows the inclusion of analytic probability distribution functions), and in terms of the signal strength modifier $\mu = \sigma/\sigma_{SM}$. For practical reasons, we also replace the pure ratio with twice its negative logarithm

$$q_\mu = -2\log\frac{\mathcal{L}(\text{data}|\mu s + b)}{\mathcal{L}(\text{data}|b)}. \tag{10.4}$$

The systematic uncertainties are then included by denoting the set of parameters at the global likelihood maximum $\hat{\mu}$ and $\hat{\theta}$ and introducing the constraint that $0 \leq \hat{\mu} \leq \mu$, which gives us the final test statistic [65]

$$\tilde{q}_\mu = -2\log\frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}, \tag{10.5}$$

where we have also used μ as shorthand for $\mu s(\theta) + b(\theta)$ in the dependency of the likelihoods. This is called the profile likelihood method.

In this analysis, we are of course in the situation of quantifying the significance of the obvious excess visible in the previous section rather than providing exclusion limits. While the latter have been calculated in the mass range from 100 to 1000 GeV, the presence of a Standard Model-like Higgs boson at 126 GeV makes them difficult to interpret. In practice they could be viewed as constraints on a non-SM resonance which happened to have width and cross section times branching ratio similar to the Standard Model Higgs.

The characterization of an excess begins with the calculation of the probability under the background-only hypothesis for an upward fluctuation as great or greater than that observered, called the p-value. This is done by throwing background-only pseudo experiments to arrive at the probability distribution for the test statistic in 10.5. The p-value is then extracted with $P(q_0 \geq q_0^{\text{data}})$, and this is converted to a significance Z which is defined implicitly as

$$p = \frac{1}{\sqrt{2\pi}} \int_Z^\infty e^{-\frac{x^2}{2}} dx. \tag{10.6}$$

One of the more attractive properties of the profile likelihood test statistic is that in the asymptotic regime it can be approximated as a half χ^2 with one degree of freedom, thus allowing the use of

$$Z = \sqrt{q_0^{\text{data}}} \tag{10.7}$$

to an excellent approximation. In practice this property also allows the p-value to be approximated as

$$p = \frac{1}{2}\left[1 - \text{erf}\left(\frac{q_0^{\text{data}}}{2}\right)\right]. \tag{10.8}$$

For the sake of expediency, we do not include the diluting effect of having searched in a number of different mass regions, sometimes called the look-elsewhere effect.

% uncertainties

	ggH	vbfH	VH	ttH	qqZZ	ggZZ
higher order terms	7.5	0.2	0.5	6.6	2.9	24
Higgs acceptance	2	2	2	2	n/a	n/a
PDF/α_s	7.2	2.7	3.5	7.8	3.4	7.2
branching ratio	2	2	2	2	n/a	n/a
luminosity	2.2/2.6	2.2/2.6	2.2/2.6	2.2/2.6	2.2/2.6	2.2/2.6
online selection	1.5	1.5	1.5	1.5	1.5	1.5
offline selection	3-11	3-11	3-11	3-11	3-11	3-11
lepton scale	0.1-0.3	0.1-0.3	0.1-0.3	0.1-0.3	0.1-0.3	0.1-0.3
lepton resolution	20	20	20	20	20	20

Table 10.1: Sources of systematic uncertainty which are included in the statistical analysis for the Monte Carlo based processes. The two values given for the luminosity uncertainty correspond to 7 and 8 TeV, and where ranges are given the uncertainties differ for the 4e, 4μ, and $2e2\mu$ channels. To the fake backgrounds we apply a 20, 40, and 25% normalization uncertainty in the 4e, 4μ, and $2e2\mu$ channels, as well as deriving a set of alternative shapes (see Section 8.2.2).

10.3 Systematic Uncertainties

A number of systematic uncertainties are included using the methods in the previous section. Those which come from the theoretical calculations which provide the Standard Model expectation were described in Chapter 2. They are also summarized in Table 10.1. Experimental uncertainties on the Monte Carlo-based processes are calculated using the tag-and-probe method (see Section 6.6) and are split into components due to the online and offline selections. The lepton scale and resolution uncertainties are evaluated using the tag and probe method (Chapter 6) as the maximum post-correction deviation between data and Monte Carlo, and are propagated as shape uncertainties on the signal and background models. The scale uncertainty is found to be 0.3, 0.1, and 0.1% in the 4e, 4μ, and $2e2\mu$ channels, while the resolution uncertainty is 20% for all three channels. The fake background normalization and shape uncertainties are derived using a number of alternative methods and samples and were described in Section 8.2.2. This fake normalization uncertainty amounts to 20, 40, and 25% in the 4e, 4μ, and $2e2\mu$ channels.

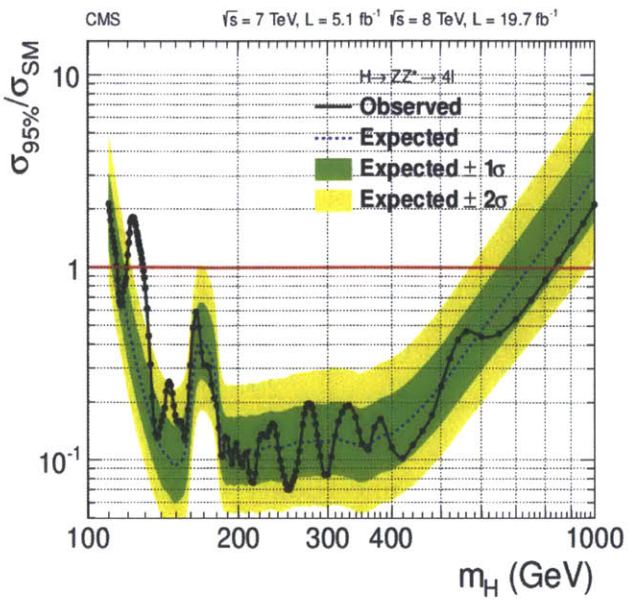

Figure 10-1: 95% confidence level upper limits on the ratio of the Higgs boson cross section to Standard Model expectation. The median background-only expectation (blue dashed line) is surrounded by its one-sigma (ugly green) and two-sigma (fluorescent yellow) uncertainties. The observed limit is shown in black. Taken from [62].

10.4 Results

The only significant excess which we observe is that between 125 and 126 GeV (see Figure 10-1), and the significance of the p-value around this excess is shown as a function of hypothesized Higgs boson mass in Figure 10-2. The maximum observed significance is 7.4 standard deviations. The inputs to this calculation are the four lepton mass and the BDT output described in Chapter 9. We also compare the number of expected and observed events via the signal strength modifier $\mu = \sigma/\sigma_{SM}$. The best-fit value of μ is found to be $1.0^{+0.3}_{-0.2}(\text{stat})^{+0.1}_{-0.1}(\text{syst})$ for Higgs boson mass hypotheses between 125 and 126 GeV. We also calculate μ for each channel individually and report the results in Table 10.2.

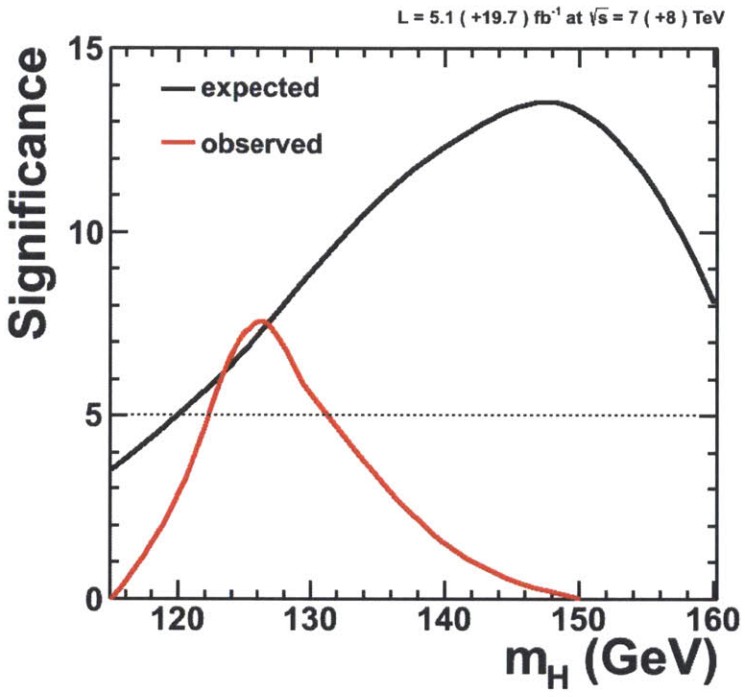

Figure 10-2: Significance of the observed fluctuation above background near the excess at 126 GeV.

channel	σ/σ_{SM}
4e	$1.5^{+0.9}_{-0.6}$
4μ	$1.1^{+0.5}_{-0.4}$
2e2μ	$0.85^{+0.4}_{-0.3}$

Table 10.2: Best fit values of $\mu = \sigma/\sigma_{SM}$ for each of the three decay channels with their total uncertainty. The overall value is $1.0^{+0.3}_{-0.2}(\text{stat})^{+0.1}_{-0.1}(\text{syst})$.

Chapter 11

Differential Cross Sections

We have seen in the previous chapter that we observe a signal strength parameter μ which is consistent with the Standard Model. We now turn to the task of checking the properties of this new particle against that same expectation to determine if it is indeed the source of electroweak symmetry breaking in the Standard Model.

In Section 2.6.1 we motivated the measurement of the p_{T} spectrum of the new particle as being of primary importance in studying the details of its production mechanisms. While from Figure 8-7 we can discern that, if the background p_{T} distribution conforms to expectation, the signal distribution also is not too far from the prediction. However, we have a wealth of additional event information beyond the four lepton p_{T}, and we can put this to good use in disentangling the contributions of signal and background.

11.1 Extraction of the signal and background distributions

We thus find ourselves with one variable, for instance p_{T}, which is of interest to us, and a number of discriminating variables (mainly $m_{4\ell}$) which have some separating power between signal and background. In order to extract the distribution of interest we will perform a likelihood fit in order to calculate a new set of event weights. The

log likelihood is written as a sum over the events e

$$\mathcal{L} = \sum_{e=1}^{N} \ln \left\{ \sum_{i=1}^{N_s} N_i f_i(y_e) \right\} - \sum_{i=1}^{N_s} N_i, \tag{11.1}$$

for N total events, N_s species, and N_i expected events for species i. The set of discriminating variables is represented by y, and $f_i(y_e)$ is the PDF for the i^{th} species evaluated at the values taken on by the y discriminating variables in event e.

Perhaps the most obvious way to attempt a separation of the signal and background distributions would be to use the signal probability as an event weight $\mathcal{P}$

$$\mathcal{P}(y_e) = \frac{N f(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}. \tag{11.2}$$

This procedure does in fact yield an unbiased estimator of the true distribution M_n of species n if the variable of interest is among the discriminating variables. However, in such an instance we would require a priori knowledge of the very distribution which we are attempting to measure. Furthermore, the use of the PDFs in the definition of the weights introduces a circular dependency such that the overall fit quality cannot be evaluated from the final results. In other words, the method is subject to uncontrollable systematic uncertainties.

We thus move to the more interesting case in which the variable of interest is not among the discriminating variables. The signal-probability weights would in this case yield a biased estimate of the true distribution M_n of species n

$$\langle N_n M_n \rangle = N_n \sum_{j=1}^{N_s} M_j N_j \int dy \frac{f_n(y) f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)}. \tag{11.3}$$

We note, however, that the problematic correction term on the right hand side is related to the likelihood by a series of derivatives

$$\frac{-\partial^2 \mathcal{L}}{\partial N_n \partial N_j} = \sum_{e=1}^{N} \frac{f_n(y_e) f_j(y_e)}{(\sum_{k=1}^{N_s} N_k f_k(y_e))^2}. \tag{11.4}$$

Defining these to be the elements of a matrix V, we can use their expectation values

$$\langle V_{nj}^{-1} \rangle \equiv \int dy \frac{f_n(y) f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \tag{11.5}$$

to rewrite equation 11.3

$$\langle M_n \rangle = \sum_{j=1}^{N_s} M_j N_j \langle V_{nj}^{-1} \rangle. \tag{11.6}$$

We may easily invert this matrix equation to recover the desired true distribution

$$N_n M_n = \sum_{j=1}^{N_s} \langle V_{nj} \rangle \langle M_j \rangle. \tag{11.7}$$

The correct event weight is thus the covariance-weighted quantity

$$\mathcal{P}_n = \frac{\sum_{j=1}^{N_s} V_{nj} f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}. \tag{11.8}$$

This is called the s-plot technique [66] and it tells us how to calculate a signal weight for each event such as to reconstruct the original signal distribution in the variable of interest. Similarly, a background weight is calculated for each event whose application yields the original background distribution. By construction these weights take on both positive and negative values, with a more positive signal weight meaning a more signal-like event and a more negative signal weight meaning a more background-like event.

The only restriction on the technique's applicability is that the discriminating variables must be entirely uncorrelated from the variable of interest. The maximal set of discriminating variables which we can use when the variable of interest is the four lepton p_{T} thus consists of $m_{4\ell}$ and the five production and decay angles listed in Section 9.1.

While the Higgs boson rapidity is also of some interest, in particular when fitting for parton distribution functions, it is unfortunately correlated with $m_{4\ell}$. While the correlation is small, it is enough to thwart accurate closure testing. We thus leave the rapidity measurement to later investigation.

11.2 Systematics

We account for a number of systematic uncertainties. All are propagated in a bin-by-bin manner since we are interested in the final shape. The uncertainty on the scale and resolution of the lepton p_{T} is treated slightly differently, so we discuss it first. We begin with the same analytic parametrizations as in the p-value analysis, and vary the mean and width up and down in order to obtain alternative shapes. The entire analysis is then performed with these alternative shapes. For each bin in the final distribution of the variable of interest, the difference between the distributions obtained with these "up" and "down" shapes and the central value is the uncertainty attributed to lepton p_{T} scale and resolution. This is the dominant systematic uncertainty.

For the remaining uncertainties we use event weights calculated in a separate step to obtain the "up" and "down" shapes for each quantity.

Online and offline lepton efficiency uncertainties, which were calculated with the tag-and-probe method, are treated in this way. Uncertainties on the renormalization and factorization scales in the Monte Carlo generators are treated similarly, using the standard prescription of varying each up and down by a factor of two. The lack of perfect knowledge of the proton parton distribution functions (PDFs) is estimated by generating alternative Monte Carlo samples by varying the PDFs within their uncertainties. The differences between these alternatives are summed in quadrature. The inaccuracy inherent to truncation of the series expansion at next-to-leading order in the POWHEG generator is found by comparing to the HRes generator, which is at next-to-next-to-leading order. Finally, the uncertainty on the scale used in the resummation of multiple soft gluon emissions in the latter generator is found by varying it up and down by a factor of two.

A number of Monte Carlo closure tests were performed in order to ensure that the s-plot technique was working as expected. Please note that in this context when we refer to "Monte Carlo" that where appropriate this is understood to be a shorthand for "expectation" and thus includes the fake background as estimated from data. Examples of these closure tests are shown in Figure 11-1. Here the central value of

the band is the Monte Carlo truth distribution for signal and background, while its size is given by the propagation of the systematic uncertainties listed above. The crosses are the shape obtained by applying the s-plot procedure to all of the Monte Carlo (signal and background) together. The uncertainties on the crosses are purely the statistical uncertainty from the number of events in the Monte Carlo sample. It is seen that the test closes to within the appropriate uncertainties.

We separate the uncertainties in this manner such as to enable the following interpretation: the size of the band shows how good the measurement could be given current Monte Carlo samples and detector performance (independent of the present luminosity), while the uncertainties on the crosses tell us the precision of the measurement given the current size of our dataset. In other words, the analysis is systematically limited if the error bars of the crosses are smaller than the size of the band. We maintain this convention when showing results with data, except that the colored crosses are replaced with black crosses.

11.3 Unfolding

We are in general interested in true, or generator-level, rather than reconstructed distributions, so we also unfold the shapes resulting from the s-plot procedure in order to remove detector effects from the final result. We use the RooUnfold package [67] to achieve this because it is well-validated and enjoys widespread use within CMS. We test several of the methods included in the package, and incorporate the relevant statistical and systematic uncertainties into the bin-by-bin uncertainties in the final distributions.

We perform a number of Monte Carlo closure tests here as well. We first check that the unfolding procedure alone successfully recreates the original generator-level distributions. The results of this test are found in Figure 11-2, where it is seen that the procedure performs as expected. Since the p_{T} evinces minimal differences between generator and reconstructed level, the closure test was also performed on the mass of the Z boson nearest to the Z pole (Z_1). The angular variables, like p_{T}, do not change

significantly between generator and reconstructed level once selection has been taken into account.

We also perform closure tests to ensure that the concatenation of s-plot and unfolding together behave as expected. This presents some difficulty because the s-plot procedure is applicable only for variables of interest which are uncorrelated with the discriminating variables. The only variables which satisfy this criterion (for instance, p_{T} and the angular variables) unfortunately are not affected significantly by detector resolution. On the other hand variables which satisfy the latter (e.g. the Z_1 mass) are strongly correlated with $m_{4\ell}$ so cannot be used to check the s-plot procedure.

We thus create a toy variable which will exhibit both of the characteristics that we need. We choose this variable to be distributed as a Breit-Wigner for both signal and background, but with different shape parameters for each. So for each event in the Monte Carlo samples, we pull a random value from a Breit-Wigner distribution, and smear and scale it to mimic detector effects. We then apply the s-plot procedure to the resulting toy variable to separate signal from background. Finally, we unfold the signal and background distributions which were obtained with the s-plot technique in order to recover the original unsmeared toy Breit-Wigner distributions from which we had pulled the random values. The results of this test are shown in Figure 11-3, where it can be seen that the chaining of s-plot and unfolding also performs as anticipated.

11.4 Results

Finally, we present the distributions of kinematic variables for signal and background which have been extracted with the maximal set of uncorrelated kinematic variables and unfolded back to generator level. We report results which are normalized to the cross section integrated over the acceptance region. As seen in Figures 11-4, 11-5, and 11-6, the data is consistent with Standard Model expectation. In Figure 11-4 we also show the p_{T} distributions for the various Higgs production mechanisms as a reminder of the differences between which we are trying to discriminate.

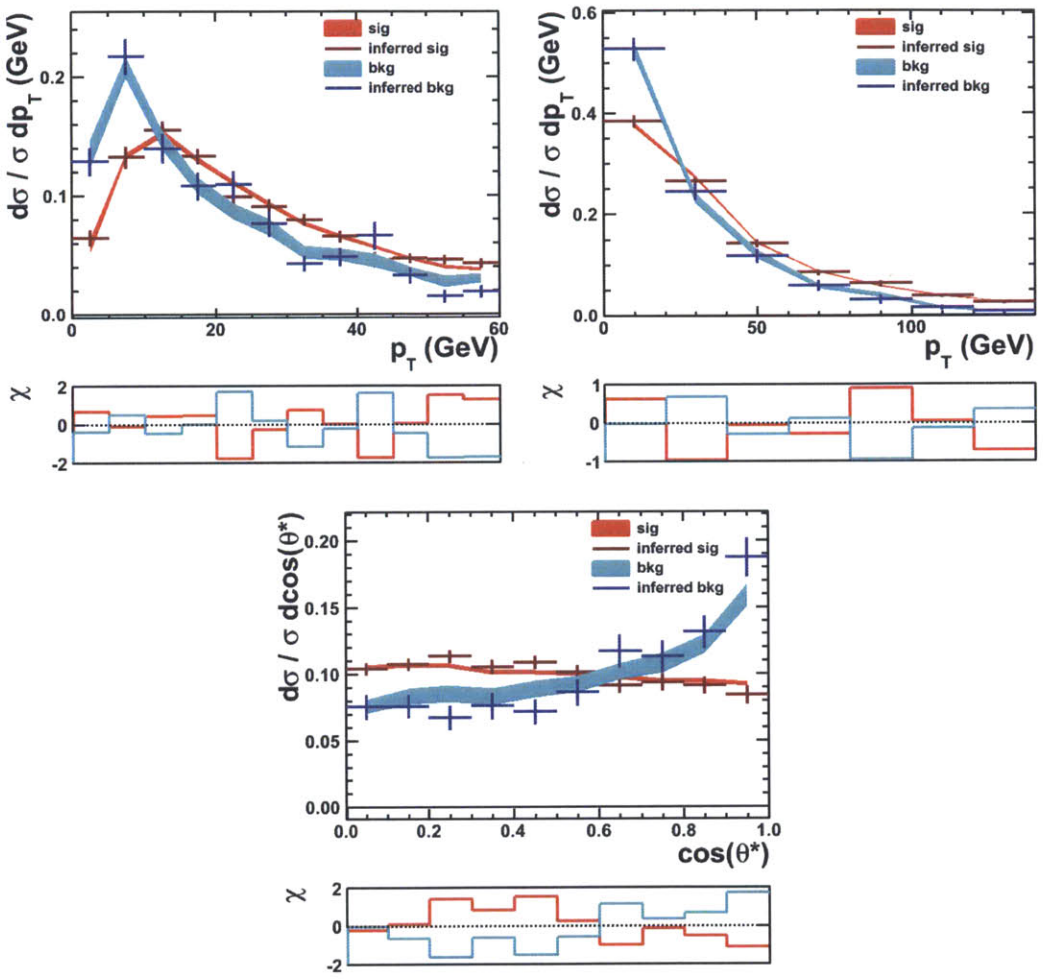

Figure 11-1: Monte Carlo closure test validating the s-plot procedure for p_{T} in narrow (top left) and wide (top right) ranges, and $\cos\theta^*$ (bottom). The other angular variables give similar closure results.

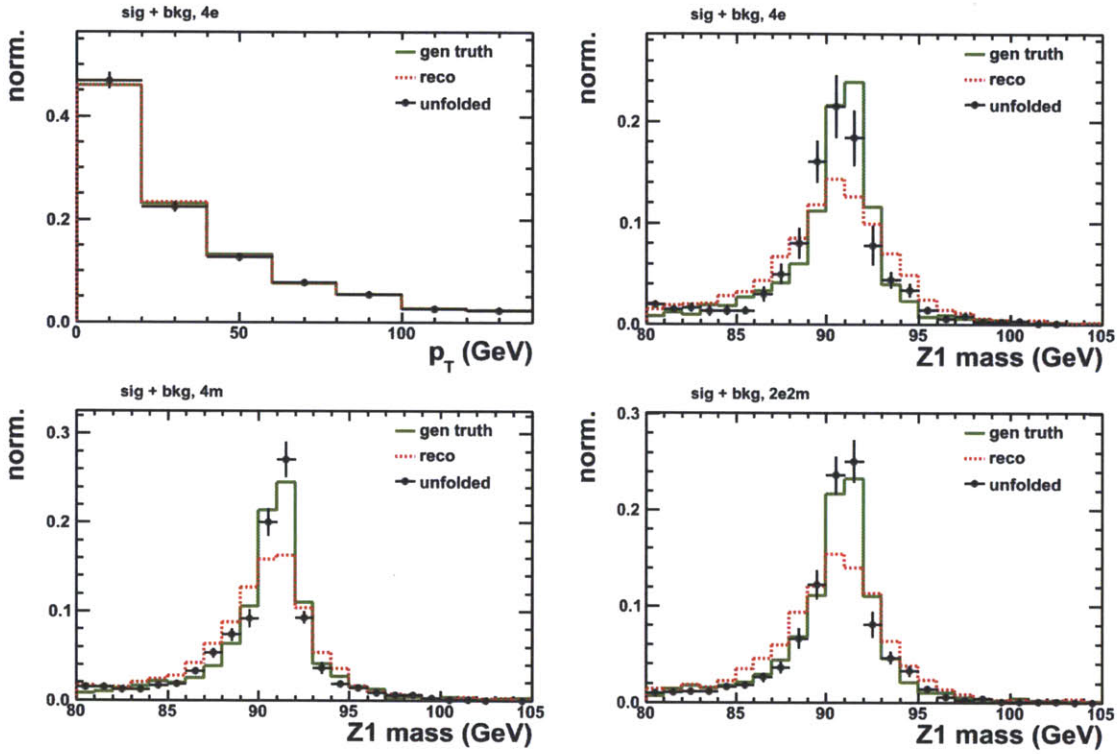

Figure 11-2: Monte Carlo closure test validating the unfolding procedure alone. It is seen that p_{T} (top left) has minimal change between generator and reconstructed shapes, so we also show successful closure on the mass of the Z boson nearest to the Z pole (Z_1). The sub-channels (4e, 4μ, and $2e2\mu$) are shown separately because of the difference in resolution between electrons and muons.

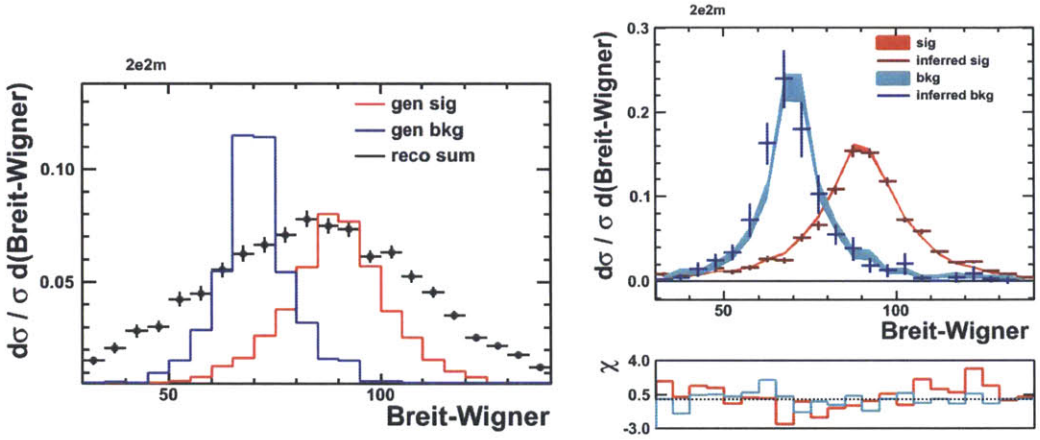

Figure 11-3: Monte Carlo closure test of the chaining of the s-plot and unfolding procedures on a toy variable with Breit-Wigner distribution for both signal and background. On the left in red and blue we see the original generator distributions of the variable for signal and background. The black points on the left are the result of smearing these and then adding the signal and background together. Application of the s-plot procedure to these black points to separate signal from background, followed by the unfolding of detector effects, gives us the colored crosses in the right-hand plot. It is seen that these crosses accurately recreate the bands in the right-hand plot, which are the original generator distributions.

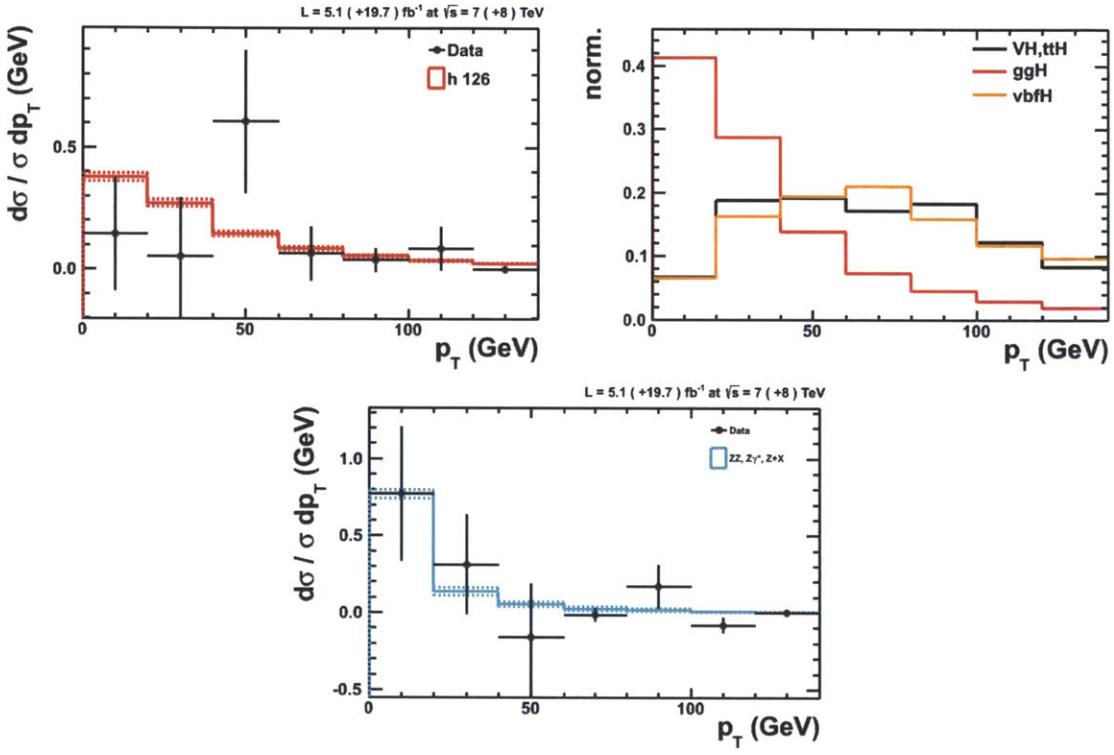

Figure 11-4: Normalized differential cross sections as a function of p_{T} for signal (top left) and background (bottom) extracted using $m_{4\ell}$ and the five decay angles described in Section 9.1 ($\cos\theta_1$, $\cos\theta_2$, Φ, Φ_1, $\cos\theta^\star$) and unfolded back to generator level. The colored band is the expectation, with its size indicating systematic uncertainties, while the points give the observed data distribution and its statistical uncertainty. This breakdown of uncertainties is described in more detail in the text. At top right we show the p_{T} distributions for the three Higgs production mechanisms. In the Standard Model the red line in the top left consists of ggH, vbfH, and VttH in the ratio 20:1.5:1 (see Table 8.1).

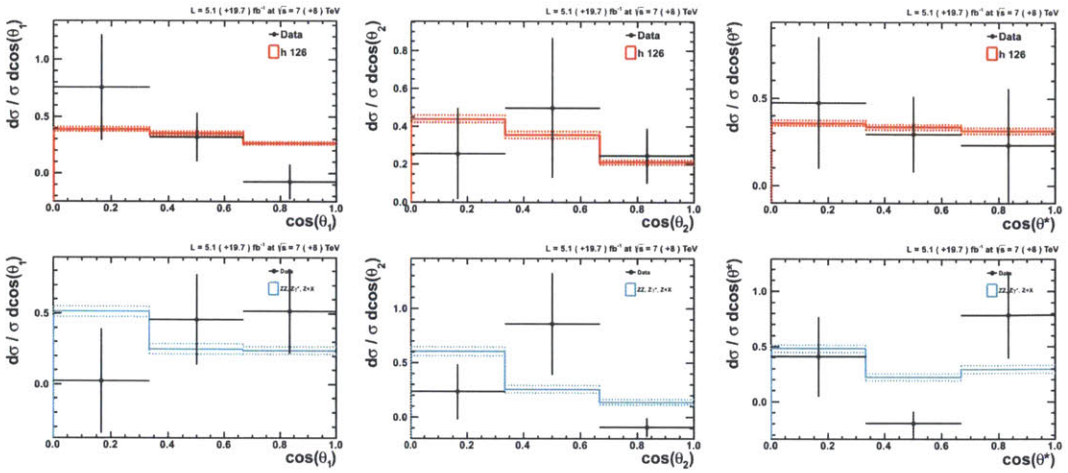

Figure 11-5: Normalized differential cross sections for the first three of the five decay angles described in Section 9.1 (see also Figure 11-6).

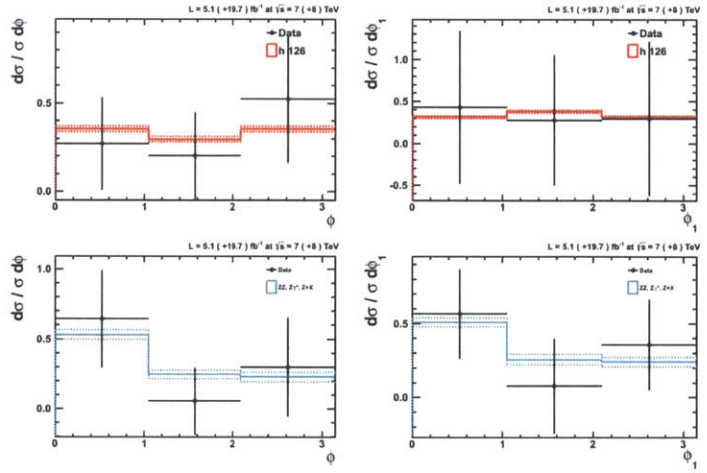

Figure 11-6: Normalized differential cross sections for the other two of the five decay angles described in Section 9.1 (see also Figure 11-5).

Chapter 12

Spin and Parity Tests

The spin and parity of the new particle also provide a means of testing its compatibility with expectation. The Standard Model Higgs boson is predicted to have spin zero and positive parity and charge conjugation, denoted 0^{++}. However, it is possible to create models in which other spin or parity hypotheses could manifest themselves. The observation of a signal in the two photon decay channel eliminates the spin-one possibility and tells us that the charge conjugation is positive [68]. The negative parity, or pseudoscalar, hypothesis provides the remaining alternative with the greatest measure of both plausibility and sensitivity, so we focus on this possibility. The CMS publication [62], however, runs through a number of other models derived from more exotic possibilities.

We use JHUgen [69] to generate simulated events for the scalar and pseudoscalar hypotheses, and pass them through the same decay and showering procedure as described for the previous Monte Carlo samples in Chapter 8.

We must then create some means of discriminating between the two signal hypotheses. The first step is to remove as much background as possible, and to this end we start with the BDT described in Chapter 9. Similarly, we use the seven variables which depend on the Higgs boson decay (five angles and two masses, see Section 9.1) to train a new BDT on scalar and pseudoscalar samples. Its input variables are shown in Figure 12-1, and the output discriminant may be found in Figure 12-2.

In order to use this information in a statistical analysis as in Chapter 10, the

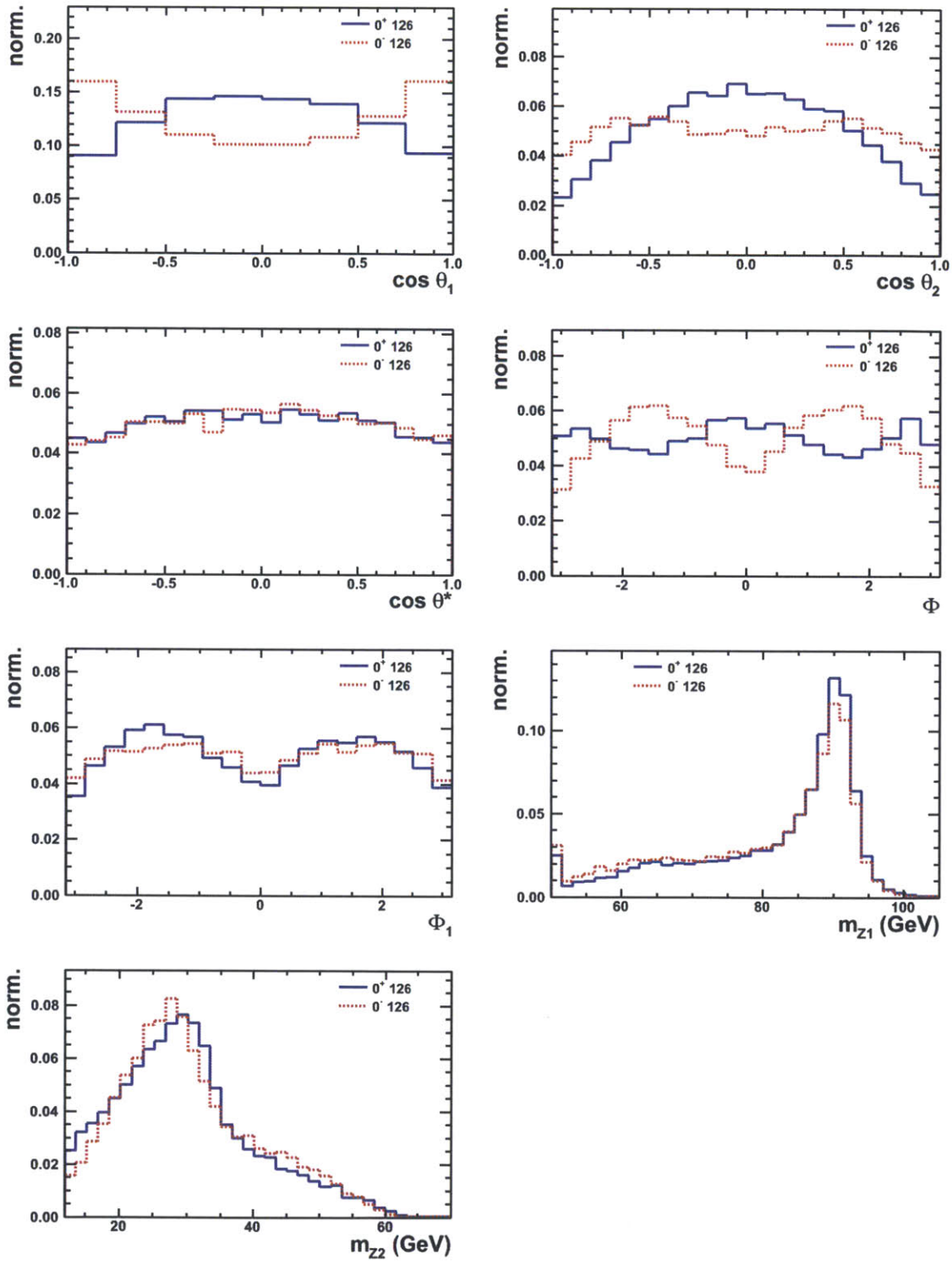

Figure 12-1: Input variables for the BDT trained to discriminate between scalar and pseudoscalar hypotheses.

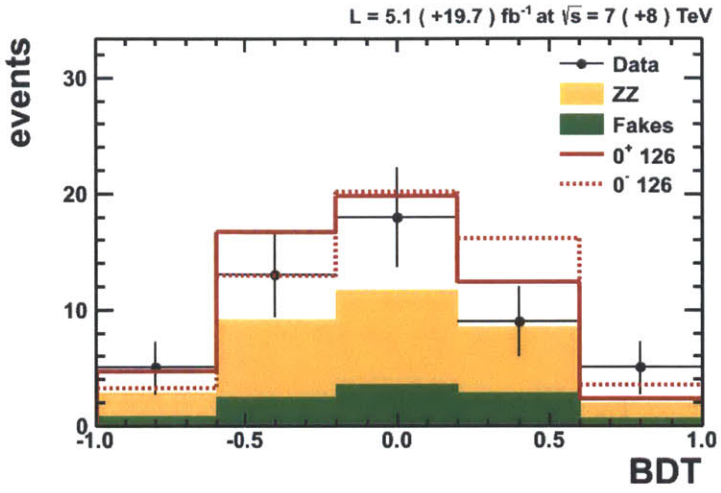

Figure 12-2: Output of the BDT trained to discriminate between scalar and pseudoscalar hypotheses.

simplest method is to add the new BDT as a third dimension such that we fit with $m_{4\ell}$, signal-background BDT, and scalar-pseudoscalar BDT. This approach, however, suffers from difficulties in the creation of the models of expectation. We do not have sufficient Monte Carlo events to populate three-dimensional templates. The other possibility is to assume no correlation between some subset of the three variables which we wish to use as axes. However, the only source of information which we have to measure the correlations of these three variables is the Monte Carlo samples, and as mentioned they do not have sufficient size to make this determination.

We therefore sidestep the issue by including $m_{4\ell}$ in the signal-background BDT. This new BDT of course has drastically increased discrimination by itself because $m_{4\ell}$ is the most discriminating variable. It is shown in Figure 12-3.

In order to quantify the relative compatibility of the two hypotheses, we proceed with a statistical procedure similar to that described in Chapter 10. While floating the signal strength as a nuisance parameter, we perform a two dimensional fit in the signal-background BDT (with $m_{4\ell}$) and the scalar-pseudoscalar BDT. Our test statistic is then the ratio of the resulting likelihoods for the signal plus background

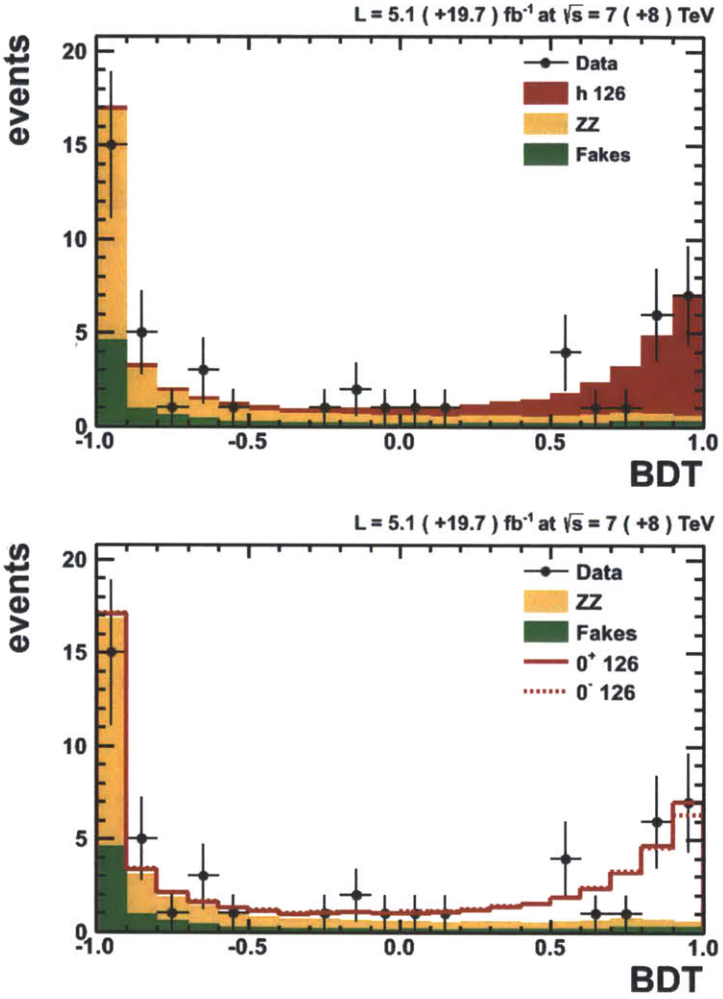

Figure 12-3: Output of the signal-background BDT into which $m_{4\ell}$ has been inserted in order to reduce the dimensionality of the fit in the statistical analysis. The signal in the top plot is from the POWHEG NLO generator which is used for the signal strength and differential cross section analyses, while the bottom plot shows the scalar and pseudoscalar Higgs signals from the JHUgen LO generator which are use for the spin-parity studies.

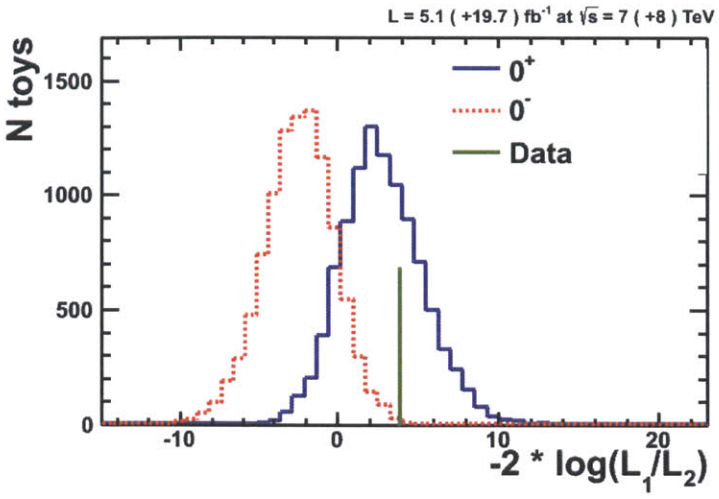

Figure 12-4: Likelihood ratio distributions for scalar (solid blue) and pseudoscalar (red) hypotheses, with the observed value drawn as a green line. The observed value disfavors the pseudoscalar hypothesis at 2.9 sigma.

hypothesis for the pseudoscalar and scalar hypotheses

$$q = \frac{\mathcal{L}(0^-)}{\mathcal{L}(0^+)}. \tag{12.1}$$

As in Chapter 10, we then convert these distributions into p-values and significance. The distribution of expected likelihood ratios for the two hypotheses is shown in Figure 12-4 with the observed value superimposed. The observed value is in the bulk of the expected scalar distribution, and disfavors the pseudoscalar hypothesis at 2.9 sigma. This is to be compared with the expected value (the tail of the pseudoscalar distribution which lies beyond the median of the scalar distribution) of 2.3 sigma.

Chapter 13

Projections

We have in previous chapters extracted much of the information which is available in the current dataset. We have discovered a particle with the mass expected for the Standard Model Higgs boson from both indirect analyses and previous direct searches. The product of its cross section with the ZZ branching ratio is within about twenty percent of expectation. It appears to be a scalar, as expected in the Standard Model, and we can tentatively assert that its p_{T} spectrum does not deviate drastically from the Standard Model expectation. The same dataset has also been analyzed for decays of the Higgs boson in other channels, and in all cases agreement is found with the Standard Model [70, 71].

The next questions to be answered, now that we know there exists a particle which is at least an excellent simulacrum of the Standard Model Higgs boson, is how deep the similarity runs and whether there exist other, new physical processes hidden in the corners. The principle means of assessing this will be through increased data accumulation in the existing analyses. In particular, in order to investigate whether the particle we observe is produced and couples precisely as expected, the measurements of the signal strength parameter $\mu = \sigma/\sigma_{SM}$ and the differential p_{T} distribution are of great importance. As such, we present here estimates of the uncertainties which can be expected with a number of different future integrated luminosities.

In Figure 13-1 we show projections of the differential cross section to 300, 1000, and 3000 fb^{-1}. The faux data points are drawn from signal and background Monte

Carlo samples according to the number of expected events at the given luminosity. The cross sections are scaled to the 14 TeV values, which for signal and background entails multiplying by 2-2.5. We compare the distributions for both the NLO POWHEG and the NNLL HRes generators (see Section 2.6.1).

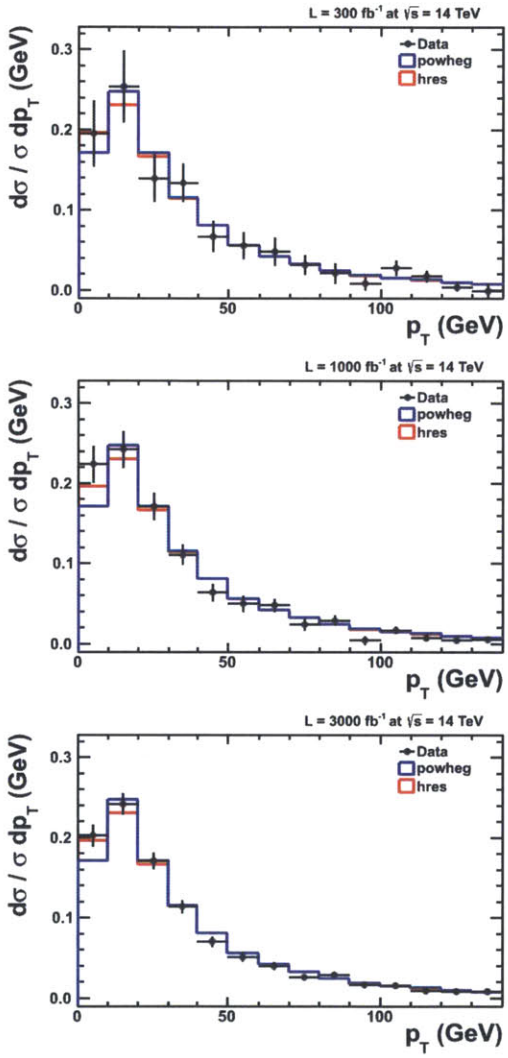

Figure 13-1: The expected performance of the differential cross section analysis at luminosities of 300 (top), 1000 (middle), and 3000 fb^{-1} (bottom). Faux data points are drawn from new signal and background Monte Carlo samples. Expectation is shown for both the NLO POWHEG and NNLL HRes generators (see Section 2.6.1).

In Table 13, meanwhile, we show the expected uncertainty on the signal strength parameter μ for 300, 1000, and 3000 fb^{-1}. As for the differential cross section, the expected number of events is scaled to the appropriate luminosity and to the 14 TeV

luminosity (fb^{-1})	total uncertainty	stat uncertainty
300	0.12	0.04
1000	0.10	0.02
3000	0.10	0.01

Table 13.1: Expected total and statistical uncertainty on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for 300, 1000, and 3000 fb^{-1} at 14 TeV.

cross sections. Because the measurement becomes systematically limited at a few hundred inverse femtobarns, we show the statistical uncertainty separately. As in past experiments, we expect that with time the systematic uncertainties will also be greatly reduced. This will be due to improved detector performance, more stringent constraints on theoretical inputs (for instance in the PDF fits), and improved analysis techniques.

It can be seen that with future LHC runs, the sensitivity to Higgs boson couplings will increase greatly, giving increased potential for the discovery of new physics beyond the Standard Model.

Chapter 14

Conclusion

We have thus presented the conclusive discovery of a new particle while searching for the Standard Model Higgs boson, which is of course a roundabout way of saying that we're confident in, and indeed have no reason to disfavor, the conclusion that we have discovered the Higgs boson.

We have described an analysis of 25 fb^{-1} of proton-proton collisions at 7 and 8 TeV from run I of the LHC in the context of decays of a Higgs boson to two Z bosons to four charged leptons. We find a single new particle with a mass of 125.6 $\pm$0.4(stat) $\pm$0.2(syst) GeV. In the context of the Standard Model this new particle has a signal strength parameter $\mu = \sigma/\sigma_{SM}$ equal to $1.0^{+0.3}_{-0.2}(\text{stat})^{+0.1}_{-0.1}(\text{syst})$. We have also measured its differential cross section in transverse momentum and in a number of decay angles and find it to conform to SM expectation. In addition, we test the new particle's spin and parity and find that the SM prediction of $\text{J}^{\text{PC}} = 0^{++}$ is preferred over a negative CP state.

These results are paralleled by those in the two photon, WW, and $\tau\tau$ decay channels [70], where all coupling and properties measurements so far agree with expectation. Measurements from the ATLAS detector in all of these channels further bolster the position of these analyses [71].

This discovery constitutes a confirmation of the theoretical structure of the Standard Model and culminates the truly extraordinary quantity of effort which has been invested in the decades-long quest to establish its authenticity. We now know that the

Standard Model provides the definitive account of the workings of the universe at the smallest scales which we have managed to probe. As in relativity's supplantation of Newtonian dynamics we may imagine, and indeed hope against hope, that new theories become necessary upon higher-precision investigation. But the Standard Model's ultimate effectiveness and equivalent veracity in its current domain of applicability cannot be questioned.

This is of course not to say that great questions do not remain. It is eminently plausible that Supersymmetry will reveal its distressingly convoluted but indubitably well-motivated countenance during the LHC's run II in 2015 and beyond. There exists an immense confidence that our surroundings are composed mainly of a previously unmeasured type of matter. There is no reason to consider it impossible that this dark matter can be produced in higher-energy proton collisions, or that its direct detection in the withering profusion of current and near-future experiments will not point the way to unexplored regions of the universe's constituent structures.

In conclusion, then, we find ourselves at a stage in the interrogation of our surroundings at which we comprehend vastly more than was possible in the very recent past. We naturally pause momentarily to admire our most recent fabrication and its apparent completeness before acknowledging with full-throated anticipation that from all prior experience we can expect to soon suffer the bombardment of fantastic new phenomena unexplainable by anything which we currently imagine.

Bibliography

[1] Enrico Fermi. Sulla quantizzazione del gas perfetto monoatomico. *Rendiconti Lincei*, 3:145 149, 1926. arXiv:cond-mat/9912229 (translated by Alberto Zannoni).

[2] P. A. M. Dirac. On the theory of quantum mechanics. *Proceedings of the Royal Society*, Series A 112 (762):661–677, 1926.

[3] Satyendra Nath Bose. Plancks Gesetz und Lichtquantenhypothese. *Zeitschrift fr Physik*, 26:178–181, 1924.

[4] S. L. Glashow. Partial-symmetries of weak interactions. *Nucl. Phys.*, 22:579, 1961.

[5] S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264, 1967.

[6] A. Salam. Weak and electromagnetic interactions. *Proceedings of the eighth Nobel symposium*, page 367, 1968.

[7] Frank Close. *The Infinity Puzzle: Quantum Field Theory and the Hunt for an Orderly Universe.* Oxford University Press, 2011.

[8] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. Global conservation laws and massless particles. *Physical Review Letters*, 13 (20):585 87, 1964.

[9] P. W. Anderson. Plasmons, gauge invariance, and mass. *Physical Review*, 130 (1):439442, 1962.

[10] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters*, 13 (9):321323, 1964.

[11] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Physical Review Letters*, 13 (16):508509, 1964.

[12] G. 't Hooft and M. Veltman. Regularization and renormalization of gauge fields. *Nuclear Physics B*, 44 (1):189219, 1972.

[13] John R. Ellis, Mary K Gaillard, and Dimitri V Nanopoulos. A phenomenological profile of the higgs boson. *Nucl.Phys. B*, 106:292, 1976.

[14] The CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Phys. Lett. B*, 716:30, 2012. arXiv:1207.7235.

[15] The ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Phys. Lett. B*, 716:1–29, 2012. arXiv:1207.7214.

[16] Chris Quigg. Electroweak symmetry breaking and the higgs sector. *Acta Phys Polon*, B30:2145–2192, 1999.

[17] Chris Quigg. The electroweak theory. In Jonathan L. Rosner, editor, *Flavor Physics for the Millennium: TASI 2000*, pages 3–67, Singapore, 2001.

[18] David Griffiths. *Introduction to Elementary Particles.* John Wiley & Sons, 1987.

[19] R. V. Harlander. Virtual corrections to ggh to two loops in the heavy top limit. *Phys. Lett. B*, 492:74–80, 2000. hep-ph/0007289.

[20] D. Catani, M. de Florian, M. Grazzini, and P. Nason. Soft-gluon resummation for higgs boson production at hadron colliders. *JHEP*, 07:028, 2003. hep-ph/0306211.

[21] G. Actis, C. Passarino, C. Sturm, and S. Uccirati. Nlo electroweak corrections to higgs boson production at hadron colliders. *Phys. Lett. B*, 670:12–17, 2008. arXiv:0809.1301.

[22] S. Marzani, R. D. Ball, V. Del Duca, S. Forte, and A. Vicin. Higgs production via gluon-gluon fusion with finite top mass beyond next-to-leading order. *Nucl. Phys. B*, 800:127–145, 2008. arXiv:0801.2544.

[23] The LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections.* CERN-2012-002 (CERN), 2010-2012.

[24] A. Djouadi, J. Kalinowski, , and M. Spira. Hdecay: A program for higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998.

[25] A. Djouadi, J. Kalinowski, M. Mhllcitner, , and M. Spira. An update of the program hdecay. *The Les Houches 2009 workshop on TeV colliders*, 2010. arXiv:1003.1643.

[26] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber. Radiative corrections to the semileptonic and hadronic higgs-boson decays h $\rightarrow$ ww/zz $\rightarrow$ 4 fermions. *JHEP*, 0702:080, 2007. http://omnibus.uni-freiburg.dc/ sd565/programs/prophecy4f/prophecy4f.html.

[27] M L Strong. Testing cp properties of higgs bosons. 1995. arXiv:hep-ph/9504345v1.

[28] M. Kraemer, J. Kuehn, M. L. Stong, and P. M. Zerwas. Prospects of measuring the parity of higgs particles. *Z Phys C*, 64:21–30, 1994. arXiv:hep-ph/9404280.

[29] Tom Melia, Paolo Nason, Raoul Rntsch, and Giulia Zanderighi. W+w-, wz and zz production in the powheg box. *JHEP*, 1111:078, 2011. arXiv:1107.5051.

[30] P. Mertsch T. Binoth, N. Kauer. Gluon-induced qcd corrections to pp→zz→4ℓ. arXiv:0807.0024.

[31] E.W.N. Glover and J.J. Van der Bij. Z-boson pair production via gluon fusion. *Nuclear Physics B*, 321:561–590, 1989.

[32] S. Alioli, P. Nason, C. Oleari, and E. Re. Higgs boson production in gluon fusion. *JHEP*, 0904:002, 2009. arXiv:0812.0578.

[33] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *JHEP*, 0605:026, 2006. arXiv:hep-ph/0603175v2.

[34] B Andersson et al. *Phys. Rept.*, 97:31, 1983.

[35] Thais L. Lungov and Carlos O. Escobar. Using rapidity gaps to distinguish between higgs boson production by w and gluon fusion. *Phys. Rev. D*, 53:48574865, 1996.

[36] Daniel de Florian, Giancarlo Ferrera, Massimiliano Grazzini, and Damiano Tommasini. Higgs boson production at the lhc: transverse momentum resummation effects in the h→ $\gamma\gamma$, h→ $ww \to l\nu l\nu$ and h→ $zz \to 4l$ decay modes. *JHEP*, 06, 2012.

[37] Christian Arnesen, Ira Z. Rothstein, and Jure Zupan. Smoking guns for on-shell new physics at the lhc. *Phys Rev Lett*, 103:151801, 2009. arXiv:0809.1429.

[38] J. Beringer et al. (PDG). *PR D*, 86:010001, 2012.

[39] P Teixeira-Dias. Higgs boson searches at lep. *Journal of Physics: Conference Series*, 110:042030, 2008.

[40] Peter B. Renton. Electroweak fits and constraints on the higgs mass. 2004. arXiv:hep-ph/0410177.

[41] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Ludwig, K. Moenig, M. Schott, and J. Stelzer. Updated status of the global electroweak fit and constraints on new physics. *Eur. Phys. J.*, C72:2003, 2012. arXiv:1107.0975.

[42] Kyle J. Knoepfel. Standard model higgs boson searches at the tevatron. 2013. arXiv:1305.1530.

[43] Lyndon Evans and Philip Bryant (editors). Lhc machine. *JINST*, 3:S08001, 2008.

[44] C.E. Hill, A. Lombardi, R. Scrivens, M. Vretenar, A. Feschenko, and A. Liou. Tests of the cern proton linac performance for lhc-type beams. *eConf C000821*, TUD17, 2000. arXiv:physics/0008102.

[45] U. Bigliani et al. The rf acc. syst. for the cern ps booster. *IEEE Trans. Nucl. Sci. NS-18*, 3:233, 1971.

[46] R. Garoby et al. Rf system for high beam intensity acceleration in the cern ps. *13th IEEE Particle Accelerator Conference*, 1989. CERN/PS 89-28 (RF).

[47] Collier P. et. al. The sps as lhc injector, conceptual design. 1997. CERN-SL-97-007-DI.

[48] D Acosta (editor). Cms physics tdr: Volume i (ptdr1), detector performace and software. 2006. CERN-LHCC-2006-001.

[49] CMS Collaboration. Energy calibration and resolution of the cms electromagnetic calorimeter in pp collisions at sqrt(s) = 7 tev. *JINST*, 8:P09009, 2013. arXiv:1306.2016v2.

[50] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *JINST*, 6:P11002, 2011.

[51] The CMS Collaboration. Energy calibration and resolution of the cms electromagnetic calorimeter in pp collisions at sqrt(s) = 7 tev. *JINST*, 8:09009, 2013. arXiv:1306.2016.

[52] L Agostino et al. Commissioning of the cms high level trigger. *JINST*, 4:P10005, 2009.

[53] Cosma Shalizi. Classification and regression trees. 2009. http://www.stat.cmu.edu/ cshalizi/350/lectures/22/lecture-22.pdf.

[54] R.E. Schapire. The strength of weak learnability. *Machine Learning*, 5:197–227, 1990.

[55] Y. Freud. Boosting a weak learning algorithm by majority. *Inform. and Comput.*, 121:256–285, 1995.

[56] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. Tmva: Toolkit for multivariate data analysis. *PoS*, 2007:040, ACAT.

[57] J Bendavid. A regression tree implementation. *Come on Josh, write a paper.*

[58] R. Fruhwirth. Application of kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A:262, 1987.

[59] Hans Bethe and Walter Heitler. On the stopping of fast particles and on the creation of positive electrons. *Proceedings of the Royal Society of London*, Series A Volume 146 Issue 856:83–112, 1934.

[60] CMS Collaboration. Measurement of the inclusive w and z production cross sections in pp collisions at sqrt s = 7 tev. *JHEP*, 10:132, 2011.

[61] J. Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.

[62] The CMS Collaboration. Measurement of the properties of a higgs boson in the four-lepton final state. *Phys. Rev. D*, 2014. arXiv:1312.5353.

[63] T. Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999. arXiv:hep-ex/9902006.

[64] ATLAS and LHC Higgs Combination Group CMS Collaborations. Procedure for the lhc higgs boson search combination in summer 2011. *ATL-PHYS-PUB/CMS NOTE*, 2011/005, 2011.

[65] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur.Phys.J.*, C71:1554, 2011.

[66] M. Pivk and F. R. Le Diberder. Splot: A statistical tool to unfold data distributions. *Nucl.Instrum.Meth.*, A555:356–369, 2005. arXiv:physics/0402083.

[67] Tim e Adye. Unfolding algorithms and tests using roounfold. *Proceedings of the phystat 2011 workshop, CERN, geneva, switzerland*, page 313, 2011. arXiv:1105.1160.

[68] M. Grazzini. Higgs spin/cp properties. *Aspen*, 2013. http://indico.cern.ch/getFile.py/access?contribId=51&sessionId=1&resId=0&materialId=slides

[69] S Bolognesi et al. On the spin and parity of a single-produced resonance at the lhc. *Phys.Rev.D*, 86:095031, 2012.

[70] The CMS Collaboration. Measurement of higgs boson production and properties in the ww decay channel with leptonic final states. *submitted to JHEP*, 2013. arXiv:1312.1129v1.

[71] The ATLAS Collaboration. Evidence for the spin-0 nature of the higgs boson using atlas data. *Phys. Lett. B*, 726:120, 2013. arXiv:1307.1432v1.