Discovery of a Higgs Boson with the ATLAS Detector

by

Aaron James Armbruster

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in the University of Michigan
2013

Doctoral Committee:

Professor Jianming Qian, Chair
Professor Homer A. Neal
Associate Professor Aaron T. Pierce
Professor Virginia R. Young
Assistant Professor Junjie Zhu

ACKNOWLEDGEMENTS

"Simplicity, patience, compassion.
These three are your greatest treasures.
Simple in actions and thoughts, you return to the source of being.
Patient with both friends and enemies,
you accord with the way things are.
Compassionate toward yourself,
you reconcile all beings in the world."

- Lao Tzu, *Tao Te Ching*

First and foremost I would like to thank my family, without the love and support of whom I would not have gotten to where I am today. My advisor, Jianming Qian, who has been an endless source of knowledge, has given me the guidance needed to stand out in a crowd of 3000. Fellow students and postdocs at Michigan, past and present: Hao Liu, John Purdham, Jonas Strandberg, Devin Harper, Shannon Walch, Jonathan Long, Jianbei Liu, Xuefei Li, Magda Chełstowska, and many others, you have been a source of both frendship and intellectual inspiration. Rudi Thun, who was the co-founder of the Michigan Higgs group along with Jianming, was always a voice of reason that never failed to counter capriciousness when necessary. Finally, to the many friends and colleagues at CERN and within the Higgs group, in particular within both HSG3 and HSG7, this would been a much more difficult journey without you walking beside me.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

ABSTRACT

Discovery of a Higgs Boson with the ATLAS Detector

by

Aaron James Armbruster

Chair: Jianming Qian

Details of the discovery of a Higgs boson and measurements of its properties with the ATLAS detector are presented, using up to 4.8 and 21 fb^{-1} of data at $\sqrt{s} = 7$ and 8 TeV, respectively. The mass of the boson is measured to be $m_H = 125.5 \pm 0.2$ (stat) $^{+0.5}_{-0.6}$ (sys) GeV using a combination of the $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ channels. At 125.5 GeV, a combination of the $H \to WW^{(*)} \to \ell\nu\ell\nu$, $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to \ell\ell\ell\ell$, $H \to b\bar{b}$, and $H \to \tau^+\tau^-$ channels yields a measured signal strength with respect to the Standard Model of $\hat{\mu} = 1.3 \pm 0.2$. In this combination the couplings of the boson are explored under various assumptions motivated by Beyond Standard Model scenarios, with no significant deviations from the Standard Model observed. Strong evidence for the vector boson fusion production process is found with an observed (expected) significance of 3.1 (2.5) standard deviations. Among the individual channels, the search in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay mode is outlined in detail. Combining the full 7 and 8 TeV datasets, the analysis yields strong evidence for this mode with an observed (expected) significance of 3.8 (3.7) standard deviations at $m_H = 125$ GeV. At this mass, the signal strength in the channel is measured to be $\hat{\mu} = 1.01 \pm 0.31$. The statistical methods used in these measurements, and particle searches in general, are thoroughly detailed. Further, novel techniques are presented, including polynomial interpolation

of nuisance parameters, procedures for asymptotically approximating quantiles of expected limits, uncapping of profile likelihood test statistics, and the applications of B-Splines for interpolating signal distributions in low resolution search channels.

CHAPTER I

Introduction

"I believe our future depends powerfully on how well we understand this Cosmos, in which we float like a mote of dust in the morning sky."

- Carl Sagan

The first scientific experiment recorded by humans was performed by the Greek scholar Eratosthenes in 240 BC. In his hometown of Syene in Egypt, he knew that at noon on the summer solstice the sun is directly overhead, such that vertical objects cast no shadow on the ground. At the same time, an appreciable shadow is cast in Alexandria. Knowing this, he determined that the Earth was round and accurately measured its circumfrance. This was the first spark of scientific methodology that led to the fire of modern science. The last century in particular has seen an explosion of knowledge about the nature of the universe at small scales. From the discovery of the electron by J. J. Thompson in 1897 to the development of the Standard Model (SM) of particle physics, the Universe continues to reveal to us its immense grandness and elegance.

The SM has been remarkably successful at predicting and describing phenomena in the energy range currently accessible to colliders, and holds against the precision electroweak tests performed at LEP, SLC, and the Tevatron. It is a $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge theory, where the $SU(2)_L \times U(1)_Y$ portion describes the electroweak force with two charged ($W^{\pm}$) and two neutral (γ/Z) vector bosons. The $SU(3)_C$ portion contains eight spin-1 vector

fields that give rise to the eight massless gluons (g). The gauge bosons mediate interactions between three families of fermions and themselves. The $SU(2)_L$ fields couple to left-handed fermion doublets, while the $U(1)_Y$ field couples to fermions with hypercharge Y. The physical electroweak bosons are the result of the $SU(2)_L \times U(1)_Y$ symmetry breaking, in the form of two neutral currents (the massive Z and massless γ) and two massive charged currents ($W^{\pm}$).

The mechanism by which particles aquire mass is in the process of being experimentally verified, which is the subject of this thesis: The Higgs boson arises from a complex scalar field added to the $SU(2)_L \times U(1)_Y$ theory as a $SU(2)_L$ doublet with a non-zero vacuum expectation value. This non-zero vacuum expectation value generates mass terms for the gauge bosons and itself, previously required to be zero by the $SU(2)_L$ symmetry. It has been sought after since its proposal in 1964 by Robert Brout, François Englert, Peter Higgs, Gerald Guralnik, C. Richard Hagen, and Tom Kibble [51, 42, 49] and its incorporation into the SM by Steven Weinberg in 1967 [73]. It is also attractive because it softens $f\bar{f} \to W^+W^-$ scattering. Without this contribution the cross section would diverge at high energies and violate unitarity [57].

In July 2012, both ATLAS and CMS at the Large Hadron Collider (LHC) announced the discovery of a boson with properties consistent with the SM Higgs [22, 32], with Tevatron claiming evidence consistent with the LHC results [47]. The results of the full 7 and 8 TeV runs (Run I) show that it is likely a CP even scalar particle, and therefore qualifies as a Higgs boson [12, 17, 15, 33, 18, 11]. Studies of the couplings with the LHC Run I data also confirm compatibility with the SM Higgs boson with mass near 125 GeV [9, 3]. However, more data is necessary to ascertain if it is the Higgs boson of the SM or perhaps something more exotic.

CHAPTER II

Standard Model of Particle Physics

"Any intelligent fool can make things bigger, more complex, and more violent. It takes a touch of genius - and a lot of courage - to move in the opposite direction."

- Albert Einstein

The next sections motivate and lay the groundwork for the addition of the Higgs boson to the SM. This requires an understanding of gauge theories, which are intimately related to the notion of symmetry. The implications of the invariance of the laws of physics under transformations are first discussed in Section 2.1. Following the work of Dirac in Section 2.2 that describes the Lagrangian of free fermions, the unification of Quantum Mechanics (QM) with Maxwell's laws in Section 2.3 is made possible by adding to the Lagrangian a gauge field and requiring invariance under $U(1)_Y$ transformations. This unification is extended in Section 2.4 to combine the electromagnetic and weak forces into the electroweak force, wherein the motivation is provided for the Higgs mechanism covered in Section 2.5. While not related to the Higgs mechanism, an understanding of Quantum Chromodynamics (QCD) is essential and is discussed in Section 2.6. In the last section (2.7), the production and decays of the Higgs relevant to the LHC are discussed.

2.1 Noether's Theorem

Underlying the theory of particles, and physics in general, is symmetry. Noether's theorem, first proven in 1918 by Emmy Noether, states that for every transformation of a field $\phi(x)$ that leaves the action invariant there exists a conserved current. The invariance of the action is equivalent to the Lagrangian density $\mathcal{L}(x)$, which is a function of both $\phi(x)$ and its derivative $\partial^\mu\phi(x)$, being invariant up to a surface term. The action is defined by the integral of the Lagrangian density (henceforth referred to simply as the Lagrangian) over all space-time.

$$S = \int \mathcal{L}(x) d^4x \tag{2.1}$$

Consider an infinitesimal transformation of the field $\phi(x) \to \phi(x) + \delta\phi(x)$ that conserves the action:

$$\begin{aligned} 0 &= \delta S \\ &= \int d^4x \left\{ \tfrac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \tfrac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta(\partial_\mu\phi) \right\} \\ &= \int d^4x \left\{ \tfrac{\partial\mathcal{L}}{\partial\phi}\delta\phi - \partial_\mu\left(\tfrac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\delta\phi + \partial_\mu\left(\tfrac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) \right\} \end{aligned} \tag{2.2}$$

The third term represents an overall surface term. Assuming no boundary terms on the action, this term is zero. This indicates that invariance of the action is equivalent to the Euler-Lagrange equation of motion:

$$\tfrac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\left(\tfrac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right) = 0 \tag{2.3}$$

Turning this around, one can require that a shift in the Lagrangian induced by a local transformation of the field must be zero up to a local four divergence:

$$\mathcal{L}(x) \to \mathcal{L}'(x) = \mathcal{L}(x) + \partial_\mu\mathcal{J}^\mu(x) \Rightarrow \delta\mathcal{L}(x) = \partial_\mu\mathcal{J}^\mu(x) \tag{2.4}$$

The shift in the Lagrangian can also be written with respect to the field:

$$\begin{aligned}\delta\mathcal{L}(x) &= \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta(\partial_\mu\phi) \\ &= \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\delta\phi + \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) \\ &= \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right)\end{aligned} \tag{2.5}$$

Equating the two forms of $\delta\mathcal{L}(x)$, one obtains a conserved current $j^\mu(x)$:

$$\begin{aligned}j^\mu(x) &\equiv \frac{\partial\mathcal{L}(x)}{\partial(\partial_\mu\phi)}\delta\phi + \mathcal{J}^\mu(x) \\ \partial_\mu j^\mu(x) &= \partial_\mu\left(\frac{\partial\mathcal{L}(x)}{\partial(\partial_\mu\phi)}\delta\phi\right) + \partial_\mu\mathcal{J}^\mu(x) = 0\end{aligned} \tag{2.6}$$

2.2 Dirac Lagrangian

The Dirac Lagrangian describes the behavior of relativistic spin-1/2 fields $\psi(x)$ in free space:

$$\mathcal{L}_{\text{Dirac}}(x) = \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x) \tag{2.7}$$

In 4-dimensions he terms γ^μ are 4×4 matrices satisfying the anti-commutation relationship:

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} \tag{2.8}$$

where $g^{\mu\nu}$ is the metric tensor. In four dimensional Minkowski space, these matrices can be represented in terms of the Pauli sigma matrices σ^i, $i = 1, 2, 3$:

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix} \tag{2.9}$$

σ^μ and $\bar{\sigma}^\mu$ are the four-vector notations for the Pauli sigma matrices, such that $\sigma^\mu \equiv (\mathbf{I}, \sigma)$ and $\bar{\sigma}^\mu \equiv (\mathbf{I}, -\sigma)$, with $\sigma \equiv (\sigma^1, \sigma^2, \sigma^3)$. The term $\bar{\psi}(x) \equiv \psi^\dagger\gamma^0$ requires the additional γ^0 for the $\bar{\psi}\psi$ term to be Lorentz invariant. The term $\psi(x)$ can be further decomposed into left

and right handed Weyl spinors:

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \tag{2.10}$$

Applying the Euler-Lagrange formula to $\mathcal{L}_{\text{Dirac}}(x)$, one obtains the Dirac equation of motion for $\psi(x)$:

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = 0 \tag{2.11}$$

This represents the equation of motion for a free (non-interacting) fermion. The next section shows how interactions with the electromagnetic field can be incorporated into the Lagrangian.

2.3 Quantum Electrodynamics

Quantum Electrodynamics (QED) represents a unification of QM with Maxwell's laws that describe how charged fermions interact with the electromagnetic field at low energies, i.e. far from the $W^\pm/Z^0$ masses:

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \tfrac{\rho}{\varepsilon_0} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\tfrac{1}{c}\tfrac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{B} &= \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \tfrac{\partial \mathbf{E}}{\partial t} \end{aligned} \tag{2.12}$$

$\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic field, ρ is the electric charge density of space, μ_0 is the permeability of free space, and ε_0 is the permittivity of free space.

Switching to natural units and introducing the electromagnetic four-potential $A^\mu = (\phi, \mathbf{A})$, where $\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla\phi$ and $\mathbf{B} = \nabla \times \mathbf{A}$, and a corresponding field strength tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, Maxwell's equations are superseded by the Lagrangian for QED:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{2.13}$$

$D_\mu \equiv \partial_\mu - ieA_\mu$ is the covariant derivative required for gauge invariance, where $e = |e|$ is taken to be positive. This term ultimately yields the interaction terms in the Lagrangian between the fermion and electromagnetic field. Applying the Euler-Lagrange equation with respect to A^μ, one obtains two of Maxwell's four equations:

$$\begin{aligned}
\mathcal{L}_{QED} &= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \tfrac{1}{4}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu - \partial_\nu A_\mu \partial^\mu A^\nu + \partial_\nu A_\mu \partial^\nu A^\mu) \\
&= \bar{\psi}(i\gamma^\mu \partial_\mu - e\gamma^\mu A_\mu - m)\psi - \tfrac{1}{2}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu)
\end{aligned} \tag{2.14}$$

$$\begin{aligned}
\tfrac{\partial \mathcal{L}}{\partial A_\mu} - \partial_\nu(\tfrac{\partial \mathcal{L}}{\partial(\partial_\nu A_\mu)}) &= 0 \\
\Rightarrow \partial_\nu(\partial^\mu A^\nu - \partial^\nu A^\mu) &= -e\bar{\psi}\gamma^\mu\psi
\end{aligned} \tag{2.15}$$

Defining the four current $j^\mu \equiv -e\bar{\psi}\gamma^\mu\psi$, the last equation can be written in a more elegant form:

$$\partial_\nu F^{\mu\nu} = j^\mu \tag{2.16}$$

The conservation of j^μ follows directly:

$$\begin{aligned}
\partial_\mu j^\mu &= \partial_\mu \partial_\nu F^{\mu\nu} \\
&= \partial_\mu \partial_\nu (\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&= 0
\end{aligned} \tag{2.17}$$

Eq. 2.16 yields four differential equations, one for each index of μ. Writing out the $\mu = 0$ and $\mu = 1, 2, 3$ components separately and writing $j^\mu = (\rho, \mathbf{J})$ in terms of its temporal and spatial components ρ and $\mathbf{J}$, respectively, one obtains two of Maxwell's four equations:

$$\begin{aligned}
-\nabla \cdot \tfrac{\partial \mathbf{A}}{\partial t} + \nabla^2 \phi &= j^0 \;\Leftrightarrow\; \nabla \cdot \mathbf{E} = \tfrac{\rho}{\varepsilon_0} \\
\nabla \times (\nabla \times \mathbf{A}) - \tfrac{\partial^2 \mathbf{A}}{\partial t^2} + \tfrac{\partial}{\partial t}\nabla\phi &= j^i \;\Leftrightarrow\; \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \tfrac{\partial \mathbf{E}}{\partial t}
\end{aligned} \tag{2.18}$$

Noting that $\mathbf{E}$ and $\mathbf{B}$ are functions of $\mathbf{A}$ and ϕ, the last two equations follow directly

from the identities of second derivatives:

$$\begin{aligned} \nabla \times \mathbf{E} &= -\tfrac{\partial}{\partial t}\nabla \times \mathbf{A} - \nabla \times \nabla\phi &&= -\tfrac{\partial \mathbf{B}}{\partial t} \\ \nabla \cdot \mathbf{B} &= \nabla \cdot (\nabla \times \mathbf{A}) &&= 0 \end{aligned} \tag{2.19}$$

It is also useful to obtain the equation of motion of fermions by applying the Euler-Lagrange equation with respect to ψ. With respect to Eq. 2.11, the only additional term is the interaction term with the field A_μ:

$$(i\gamma^\mu \partial_\mu - m)\psi = -e\gamma^\mu A_\mu \psi \tag{2.20}$$

$\mathcal{L}_{\text{QED}}$ is invariant under a gauge transformation of the field ψ when the gauge field A_μ is simultaneously transformed:

$$\begin{aligned} \psi(x) &\rightarrow \psi'(x) &&= \psi(x)e^{i\theta(x)} \\ A_\mu(x) &\rightarrow A'_\mu(x) &&= A_\mu(x) + \tfrac{1}{e}\partial_\mu\theta(x) \end{aligned} \tag{2.21}$$

The terms $F^{\mu\nu}$ (and therefore $F^{\mu\nu}F_{\mu\nu}$) and $m\bar{\psi}\psi$ are trivially invariant. The interaction and kinetic terms each pick up a term differing by a sign and cancel:

$$\begin{aligned} \bar{\psi}\gamma^\mu D_\mu \psi \rightarrow \bar{\psi}'\gamma^\mu D'_\mu \psi' &= \bar{\psi}e^{-i\theta(x)}\gamma^\mu(\partial_\mu - ieA_\mu(x) - i\partial_\mu\theta(x))\psi e^{i\theta(x)} \\ &= \bar{\psi}\gamma^\mu D_\mu \psi + \bar{\psi}i\gamma^\mu\partial_\mu\theta(x)\psi - \bar{\psi}i\gamma^\mu\psi\partial_\mu\theta(x) \\ &= \bar{\psi}\gamma^\mu D_\mu \psi \end{aligned} \tag{2.22}$$

2.4 Electroweak Unification

QED works well in describing the interactions between charged fermions and the electromagnetic field, but is not intended or able to describe phenomena arising from the weak interactions, such as beta decays. For this, we need to extend the model to a larger theory.

The spin-1 gauge fields W^a_μ $(a = 1, 2, 3)$ transform under the adjoint of the $\text{SU}(2)_L$ group

with associated generators $T^a = \frac{1}{2}\sigma^a$, where σ^a are the Pauli matrices. The gauge field B_μ is associated with $\mathrm{U}(1)_Y$. Fermions enter $\mathrm{SU}(2)_L$ as left-handed doublets Ψ_L, while the right handed fermions ψ_R are $\mathrm{SU}(2)_L$ singlets that are charged under $\mathrm{U}(1)_Y$. Ψ_L is a doublet representation of $\mathrm{SU}(2)_L$. ψ_R is a singlet under $\mathrm{SU}(2)_L$, but transforms under $\mathrm{U}(1)_Y$:

$$\mathcal{L}_{EW} = \bar{\Psi}_L i\gamma^\mu D^L_\mu \Psi_L + \bar{\psi}_R i\gamma^\mu D^R_\mu \psi_R - \tfrac{1}{4} W^{\mu\nu}_a W^a_{\mu\nu} - \tfrac{1}{4} B^{\mu\nu} B_{\mu\nu} \tag{2.23}$$

$$\Psi_L = \begin{pmatrix} \psi'_L \\ \psi_L \end{pmatrix} \tag{2.24}$$

Ψ_L and ψ_R can be either leptons or quarks. There are three generations of both, which will remain implicit in the notation:

$$\begin{array}{ccccccc} & \text{Leptons :} & & & & \text{Quarks :} & \\ \begin{pmatrix} \psi'_L \\ \psi_L \end{pmatrix} & = & \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} & ; & \begin{pmatrix} \psi'_L \\ \psi_L \end{pmatrix} & = & \begin{pmatrix} u_L \\ d_L \end{pmatrix} \\ \psi_R & = & e_R & ; & \psi_R & = & u_R, d_R \end{array} \tag{2.25}$$

The covariant derivative contains the four gauge fields that will form the interaction terms with the fermions. This is different for the left-handed doublets and right-handed singlets:

$$\begin{aligned} D^L_\mu &= \partial_\mu - i g_1 \tfrac{Y}{2} B_\mu - i g_2 T^a W^a_\mu \\ D^R_\mu &= \partial_\mu - i g_1 \tfrac{Y}{2} B_\mu \end{aligned} \tag{2.26}$$

The hypercharge of the fermions is $Y = 2Q - 2I^3$, where Q is the electromagnetic charge and I^3 is the third component of the weak isospin. The eigenvalues of each for the different fermion flavors are listed in Table 2.1.

g_1 and g_2 are the coupling constants of the fermions to the gauge fields. The form of the field strength tensor $B_{\mu\nu}$ of B_μ is identical to that of the photon in QED. $W^a_{\mu\nu}$ is the field

	ν_L	e_L	e_R	u_L	d_L	u_R	d_R
Q	0	-1	-1	+2/3	-1/3	+2/3	-1/3
I^3	+1/2	-1/2	0	+1/2	-1/2	0	0
Y	-1	-1	-2	+1/3	+1/3	+4/3	-2/3

Table 2.1: Eigenvalues of the electromagnetic charge Q, the third component of the weak isospin I^3, and the weak hypercharge $Y = 2Q - 2I^3$ for the fermions of the SM.

strength tensor of W^a_μ. Because the generators T^a of $\mathrm{SU}(2)_L$ do not commute, $W^a_{\mu\nu}$ picks up an additional term:

$$\begin{aligned} W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2 \epsilon^{abc} W^b_\mu W^c_\nu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{2.27}$$

ϵ^{abc} is the antisymmetric tensor that arises due to the commutation relation that is the $\mathrm{SU}(2)_L$ algebra. There is no such term in $B_{\mu\nu}$ since the lone $\mathrm{U}(1)_Y$ generator Y trivially commutes:

$$\begin{aligned} [T^a, T^b] &= i\epsilon^{abc} T^c \\ [Y, Y] &= 0 \end{aligned} \tag{2.28}$$

The extra term in $W^a_{\mu\nu}$ from the non-abelian nature of $\mathrm{SU}(2)_L$ gives rise to self-interactions among the gauge bosons in this group that aren't present in $\mathrm{U}(1)_Y$ [63].

An $\mathrm{SU}(2)_L$ transformation of the Lagrangian yields the transformed fermion doublet and gauge fields required for invariance of the Lagrangian:

$$\begin{aligned} \Psi'_L &= \Psi_L e^{i\beta_a(x)T^a} \\ \vec{W}'_\mu &= \vec{W}_\mu - \tfrac{1}{g_2}\partial_\mu \vec{\beta}(x) - \vec{\beta}(x) \times \vec{W}_\mu \end{aligned} \tag{2.29}$$

Mass terms, however, will break this symmetry. Expanding the fermion field in terms of the left and right handed components shows why this is:

$$\begin{aligned} m\bar{\Psi}\Psi &= m\bar{\Psi}(\tfrac{1}{2}(1-\gamma^5) + \tfrac{1}{2}(1+\gamma^5))\Psi \\ &= m(\bar{\Psi}_R \Psi_L + \bar{\Psi}_L \Psi_R) \end{aligned} \tag{2.30}$$

Since only Ψ_L transforms under $\mathrm{SU}(2)_L$, these mass terms are not invariant under the transformation in Equation 2.29; neither are mass terms $M_a^2 W_\mu^a W^{a\mu}$ for the bosons. Fermions and bosons will have to acquire mass another way (Section 2.5).

In general, the physical bosons that couple to the fermions can be mixtures of the gauge bosons. The mixing of the W_μ^3 and the B_μ can be represented by a rotation by the Weinberg angle θ_W, where $\sin\theta_W \equiv s_W \equiv \frac{g_1}{\sqrt{g_1^2+g_2^2}}$ and $\cos\theta_W \equiv c_W \equiv \frac{g_2}{\sqrt{g_1^2+g_2^2}}$:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} c_W & s_W \\ -s_W & c_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{2.31}$$

The motivation for this choice of rotation will be more clear when discussing the Higgs mechanism in Section 2.5. Following from the rotation, B_μ and W_μ^3 can be written in terms of A_μ, Z_μ, and θ_W:

$$\begin{aligned} B_\mu &= c_W A_\mu - s_W Z_\mu \\ W_\mu^3 &= s_W A_\mu + c_W Z_\mu \end{aligned} \tag{2.32}$$

The covariant derivative can be written in terms of the mass eigenstates A_μ, Z_μ, and $W_\mu^\pm$

corresponding to the physical bosons γ, Z, and $W^{\pm}$, respectively:

$$\begin{aligned} iD_\mu^L &= i\left(\partial_\mu - ig_1\tfrac{Y}{2}B_\mu - ig_2T^aW_\mu^a\right) \\ &= \begin{pmatrix} a_\mu^{11} & a_\mu^{12} \\ a_\mu^{21} & a_\mu^{22} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} a_\mu^{11} &\equiv i\partial_\mu + \tfrac{1}{2}g_1YB_\mu + \tfrac{1}{2}g_2W_\mu^3 \\ &= i\partial_\mu + \tfrac{1}{2}(g_1c_WY + g_2s_W)A_\mu - \tfrac{1}{2}(g_1s_WY - g_2c_W)Z_\mu \\ a_\mu^{12} &\equiv \tfrac{1}{2}g_2(W_\mu^1 - iW_\mu^2) \\ &= \tfrac{1}{\sqrt{2}}g_2W_\mu^+ \\ a_\mu^{21} &\equiv \tfrac{1}{2}g_2(W_\mu^1 + iW_\mu^2) \\ &= \tfrac{1}{\sqrt{2}}g_2W_\mu^- \\ a_\mu^{22} &\equiv i\partial_\mu + \tfrac{1}{2}g_1YB_\mu - \tfrac{1}{2}g_2W_\mu^3 \\ &= i\partial_\mu + \tfrac{1}{2}(g_1c_WY - g_2s_W)A_\mu - \tfrac{1}{2}(g_1s_WY + g_2c_W)Z_\mu \end{aligned} \tag{2.33}$$

Multiplying the interaction terms with the fermion doublets and rewriting the coefficients in terms of θ_W reveals the electroweak fermion-boson interactions:

$$\begin{aligned} \bar{\Psi}_L i\gamma^\mu D_\mu^L \Psi_L &= \begin{pmatrix} \bar{\psi}'_L & \bar{\psi}_L \end{pmatrix} \gamma^\mu \begin{pmatrix} a_\mu^{11} & a_\mu^{12} \\ a_\mu^{21} & a_\mu^{22} \end{pmatrix} \begin{pmatrix} \psi'_L \\ \psi_L \end{pmatrix} \\ &= \bar{\psi}'_L\gamma^\mu a_\mu^{11}\psi'_L + \bar{\psi}'_L\gamma^\mu a_\mu^{12}\psi_L \\ &\quad + \bar{\psi}_L\gamma^\mu a_\mu^{21}\psi'_L + \bar{\psi}_L\gamma^\mu a_\mu^{22}\psi_L \\ &= i\bar{\psi}'_L\gamma^\mu\partial_\mu\psi'_L + i\bar{\psi}_L\gamma^\mu\partial_\mu\psi_L \\ &\quad + \tfrac{1}{2}e(Y+1)\bar{\psi}'_L\gamma^\mu\psi'_LA_\mu + \tfrac{1}{2}e(Y-1)\bar{\psi}_L\gamma^\mu\psi_LA_\mu \\ &\quad - \tfrac{1}{2}(g_1s_WY - g_2c_W)\bar{\psi}'_L\gamma^\mu\psi'_LZ_\mu - \tfrac{1}{2}(g_1s_WY + g_2c_W)\bar{\psi}_L\gamma^\mu\psi_LZ_\mu \\ &\quad + \tfrac{1}{\sqrt{2}}g_2\bar{\psi}'_L\gamma^\mu\psi_LW_\mu^+ + \tfrac{1}{\sqrt{2}}g_2\bar{\psi}_L\gamma^\mu\psi'_LW_\mu^- \end{aligned} \tag{2.34}$$

The identity $g_1c_W = g_2s_W \equiv e$, the electromagnetic charge, has been used in the A_μ

terms. The right handed singlet terms are similar:

$$\begin{aligned}\bar{\psi}_R i\gamma^\mu D^R_\mu \psi_R &= \bar{\psi}_R i\gamma_\mu(\partial_\mu - ig_1\tfrac{Y}{2}B_\mu)\psi_R \\ &= i\bar{\psi}_R\gamma_\mu\partial_\mu\psi_R + \tfrac{1}{2}eY\bar{\psi}_R\gamma_\mu\psi_R A_\mu - \tfrac{1}{2}g_1 s_W Y\bar{\psi}_R\gamma_\mu\psi_R Z_\mu\end{aligned} \tag{2.35}$$

Inserting the SM fermions in place of ψ'_L, ψ_L, and ψ_R and replacing the weak hypercharge with its corresponding eigenvalues, the seventeen interaction terms come about:

	Left − Handed Interactions :	
EM	−	$-e\bar{e}_L\gamma^\mu e_L A_\mu$
	$\frac{2}{3}e\bar{u}_L\gamma^\mu u_L A_\mu$	$\frac{1}{6}e\bar{d}_L\gamma^\mu d_L A_\mu$
Neutral Current	$\frac{1}{2}(g_1 s_W + g_2 c_W)\bar{\nu}'_L\gamma^\mu\nu'_L Z_\mu$	$\frac{1}{2}(g_1 s_W - g_2 c_W)\bar{e}_L\gamma^\mu e_L Z_\mu$
	$-\frac{1}{2}(\frac{1}{3}g_1 s_W - g_2 c_W)\bar{u}_L\gamma^\mu u_L Z_\mu$	$-\frac{1}{2}(\frac{1}{3}g_1 s_W + g_2 c_W)\bar{d}_L\gamma^\mu d_L Z_\mu$
Charged Current	$\frac{1}{\sqrt{2}}g_2\bar{\nu}_L\gamma^\mu e_L W^+_\mu$	$\frac{1}{\sqrt{2}}g_2\bar{e}_L\gamma^\mu\nu_L W^-_\mu$
	$\frac{1}{\sqrt{2}}g_2\bar{u}_L\gamma^\mu d_L W^+_\mu$	$\frac{1}{\sqrt{2}}g_2\bar{d}_L\gamma^\mu u_L W^-_\mu$
	Right − Handed Interactions :	
EM	−	$-e\bar{e}_R\gamma_\mu e_R A_\mu$
	$\frac{2}{3}e\bar{u}_R\gamma_\mu u_R A_\mu$	$-\frac{1}{3}e\bar{d}_R\gamma_\mu d_R A_\mu$
Neutral Current	−	$g_1 s_W\bar{e}_R\gamma_\mu e_R Z_\mu$
	$-\frac{2}{3}g_1 s_W\bar{u}_R\gamma_\mu u_R Z_\mu$	$\frac{1}{3}g_1 s_W\bar{d}_R\gamma_\mu d_R Z_\mu$

(2.36)

The addition of an $SU(2)_L$ gauge symmetry therefore inseparably unifies the electromagnetic and weak forces into the electroweak force. The electroweak Lagrangian is able to describe a plethora of physics, but leaves many unanswered questions. In particular, why are the weak bosons and fermions observed to be massive given that mass terms break $SU(2)_L$ symmetry?

2.5 Higgs Sector

The Higgs boson offers a potential solution to mass terms for weak bosons and fermions that otherwise violate the $SU(2)_L$ symmetry. This requires adding to the electroweak Lagrangian an extra spin-0 scalar field Φ as a complex doublet:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 - i\phi_2 \\ \phi_3 - i\phi_4 \end{pmatrix} \tag{2.37}$$

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) + \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 \tag{2.38}$$

Here the covariant derivative in the kinetic term is the same as the left-handed equation 2.26 with $Y_\Phi = +1$ as the eigenvalue of the weak hypercharge. The term $V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$ is the potential term of the Lagrangian. If $\mu^2 < 0$, $V(\Phi)$ has a minimum at $\Phi = 0$. If $\mu^2 > 0$, $V(\Phi)$ has a minimum at $\sqrt{|\Phi|^2} = \sqrt{\frac{\mu^2}{\lambda}}$ such that $V(\Phi)$ has a non-zero expectation value. Note that in this form the Lagrangian in Eq. 2.38 is trivially $SU(2)_L$ invariant.

To simplify the coming algebra, it's useful to work in the *unitary gauge* [38]. $\mathcal{L}_{\text{extra}}$ is $SU(2)_L$ invariant, so we can always make a transformation to this gauge that removes the upper component of Φ:

$$\begin{aligned} U(x)\Phi &= \tfrac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \phi(x) \end{pmatrix} \\ &= \tfrac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \end{aligned} \tag{2.39}$$

In the second line the remaining lower component has been expanded around a vacuum expectation value v by a small perturbation $h(x)$. Minimizing the potential term with respect to $h(x)$ shows that the minimum occurs at v, corresponding to the minimum $\sqrt{|\Phi|^2} = \frac{\mu^2}{\lambda}$ of $V(\Phi)$ mentioned above. Writing out the potential term reveals the self-interaction terms of

the field $h(x)$:

$$\begin{aligned}
\mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2 &= \tfrac{1}{2}\mu^2(v^2+h^2+2vh) - \tfrac{1}{4}\lambda(v^4+h^4+4vh^3+6v^2h^2+4v^3h) \\
&= v(\mu^2-\lambda v^2)h + (\tfrac{1}{2}\mu^2 - \tfrac{3}{2}\lambda v^2)h^2 - v\lambda h^3 - \tfrac{1}{4}\lambda h^4 + \text{const} \qquad (2.40)\\
&= \lambda v^2 h^2 - v\lambda h^3 - \tfrac{1}{4}\lambda h^4 + \text{const}
\end{aligned}$$

In the last line the substitution $\mu^2 = \lambda v^2$ was made, which removes the term linear in h. The remaining parts show the cubic and quartic self-interaction terms. Also revealed is the mass term $\lambda v^2 h^2$, yielding $M_H = \sqrt{2\lambda}v$. The kinetic term gives the effect on the gauge bosons:

$$\begin{aligned}
(D_\mu\Phi)^\dagger(D^\mu\Phi) &= \tfrac{1}{2}\left|\begin{pmatrix} \partial_\mu - \frac{i}{2}eA_\mu + \frac{i}{2}(g_1 s_W - g_2 c_W)Z_\mu & -i\frac{1}{\sqrt{2}}g_2 W_\mu^+ \\ -i\frac{1}{\sqrt{2}}g_2 W_\mu^- & \partial_\mu + \frac{i}{2}\sqrt{g_1^2+g_2^2}Z_\mu \end{pmatrix}\begin{pmatrix} 0 \\ v+h \end{pmatrix}\right|^2 \\
&= \tfrac{1}{2}(\partial_\mu h)^2 + \tfrac{g_1^2+g_2^2}{8}v^2 Z_\mu^2 + \tfrac{g_1^2+g_2^2}{8}h^2 Z_\mu^2 + \tfrac{g_1^2+g_2^2}{4}vhZ_\mu^2 \\
&\quad + \tfrac{g_2^2}{4}v^2 W_\mu^+ W^{-,\mu} + \tfrac{g_2^2}{4}h^2 W_\mu^+ W^{-,\mu} + \tfrac{g_2^2}{2}vhW_\mu^+ W^{-,\mu}
\end{aligned} \qquad (2.41)$$

The mass terms $\frac{g_2^2}{4}v^2 W_\mu^+ W^{-,\mu} = M_W^2 W_\mu^2$, $\frac{g_1^2+g_2^2}{8}v^2 Z_\mu^2 = \frac{1}{2}M_Z^2 Z_\mu^2$, and $0 = \frac{1}{2}M_A^2 A_\mu^2$ show the $W^\pm$ and Z bosons masses $\frac{1}{2}g_2 v$ and $\frac{\sqrt{g_1^2+g_2^2}}{2}v$, respectively, while leaving the photon massless as observed in nature. Cubic and quartic vertices between the bosons arise as well. It also predicts the ratio of W and Z masses to be $\frac{M_W}{M_Z} = c_W$, which is well established [62]. The vacuum expectation value $v \approx 246$ GeV can also be inferred from these relations. Written in terms of the field $h(x)$, the Lagrangian is not obviously $\mathrm{SU}(2)_L$ invariant (though it is!); this symmetry has been "spontaneously" broken by expanding around v.

The extra doublet can accommodate fermion masses when adding interaction terms between the fermions and Φ. For the down-type fermions ψ_L and ψ_R with corresponding

Yukawa couplings λ_d:

$$\begin{aligned}
\mathcal{L}^{\text{down}}_{\text{fermion}-\text{Higgs}} &= -\lambda_d \bar{\Psi}_L \Phi \psi_R + h.c. \\
&= -\tfrac{1}{\sqrt{2}}\lambda_d (\bar{\psi}'_L, \bar{\psi}_L) \begin{pmatrix} 0 \\ v+h \end{pmatrix} \psi_R + h.c. \\
&= -\tfrac{1}{\sqrt{2}}\lambda_d v \bar{\psi}_L \psi_R - \tfrac{1}{\sqrt{2}}\lambda_d \bar{\psi}_L \psi_R h + h.c. \\
&= -\tfrac{1}{\sqrt{2}}\lambda_d v \bar{\psi}\psi - \tfrac{1}{\sqrt{2}}\lambda_d \bar{\psi}\psi h
\end{aligned} \tag{2.42}$$

The expansion shows down-type fermion masses $m_d = \frac{1}{\sqrt{2}}\lambda_d v$ and the interactions of the Higgs with fermions. Generating mass is similar for the up-type fermions ψ'_L and ψ'_R. Adding to the Lagrangian terms where the Higgs field has been rotated reveals these final mass terms:

$$\begin{aligned}
\mathcal{L}^{\text{up}}_{\text{fermion}-\text{Higgs}} &= -\lambda_u \bar{\Psi}_L \Phi' \psi'_R + h.c., \\
\Phi' &= -i\sigma_2 \Phi = \begin{pmatrix} v+h \\ 0 \end{pmatrix} \\
\Rightarrow \quad \mathcal{L}^{\text{up}}_{\text{fermion}-\text{Higgs}} &= -\tfrac{1}{\sqrt{2}}\lambda_u (\bar{\psi}'_L, \bar{\psi}_L) \begin{pmatrix} v+h \\ 0 \end{pmatrix} \psi'_R + h.c. \\
&= -\tfrac{1}{\sqrt{2}}\lambda_u v \bar{\psi}'_L \psi'_R - \tfrac{1}{\sqrt{2}}\lambda_u \bar{\psi}'_L \psi'_R h + h.c. \\
&= -\tfrac{1}{\sqrt{2}}\lambda_u v \bar{\psi}'\psi' - \tfrac{1}{\sqrt{2}}\lambda_u \bar{\psi}'\psi' h
\end{aligned} \tag{2.43}$$

Just as with the down-type fermions, the up-type fermion-Higgs interaction terms arise and the masses $m_u = \frac{1}{\sqrt{2}}\lambda_u v$ are generated.

2.6 QCD

QCD is the theory of the strong force between quarks mediated by gluons. It is described by the $\text{SU}(3)_C$ group. There are eight spin-1 massless gluon fields G^a_μ (a=1,...,8). The quarks

Q are triplets under $\mathrm{SU}(3)_C$:

$$\mathcal{L}_{QCD} = \bar{Q}(i\gamma^\mu D_\mu - m)Q - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{2.44}$$

$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\nu G^c_\nu$ is the field strength tensor of G^a_μ. g_s is the strong coupling, and f^{abc} are the associated structure constants of the $\mathrm{SU}(3)_C$ algebra:

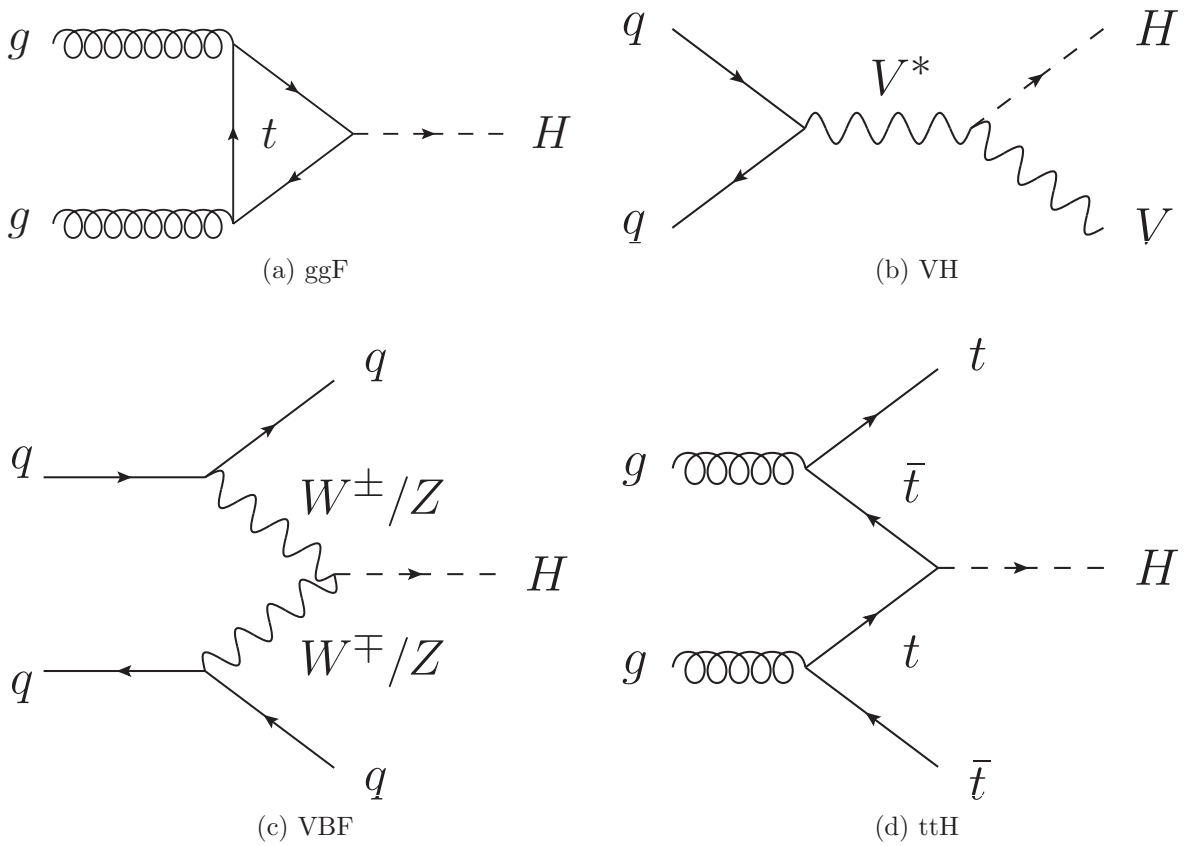
$$[t^a, t^b] = i f^{abc} t^c \tag{2.45}$$

The eight t^a are one-half times the 3x3 Gell-Mann matrices λ^a, the generators of $\mathrm{SU}(3)_C$. Similar to $\mathrm{SU}(2)_L$, the non-commuting $\mathrm{SU}(3)_C$ algebra gives rise to self-interaction among the gluons fields [63].

2.7 Higgs at the LHC

There are several Higgs boson production modes at the LHC. The dominant mode is gluon-gluon fusion ($gg \to H$, denoted ggF), in which a top- and b-quark loop mediates an effective coupling between the Higgs and two initial state gluons. Vector Boson Fusion (VBF) occurs through the process $q\bar{q} \to q'\bar{q}'V^{(*)}V^{(*)} \to q'\bar{q}'H$ ($V^{(*)} = W^{(*)}, Z^{(*)}$), where the vector bosons are radiated off of two initial state quarks and fuse to give a Higgs boson. There is a mode where a Higgs is produced in association with a vector boson (VH, $V = W, Z$), also known as Higgs-strahlung, which occurs through the process $q\bar{q} \to V^* \to VH$. Finally, there is the $t\bar{t}H$ mode through $gg/q\bar{q} \to t\bar{t}H$. The Feynman diagrams for each process are shown in Figure 2.1. The production cross sections versus m_H for these processes are also shown in Figure 2.2.

The Higgs boson has no appreciable lifetime at masses larger than a few GeV and will decay immediately into final state fermion or boson pairs. The decay into boson pairs can be divided into two-, three-, and four-body decays, as well as decays into photon pairs, gluon



(a) ggF (b) VH

(c) VBF (d) ttH

Figure 2.1: Feynman diagrams for the Higgs production modes accessible at the LHC

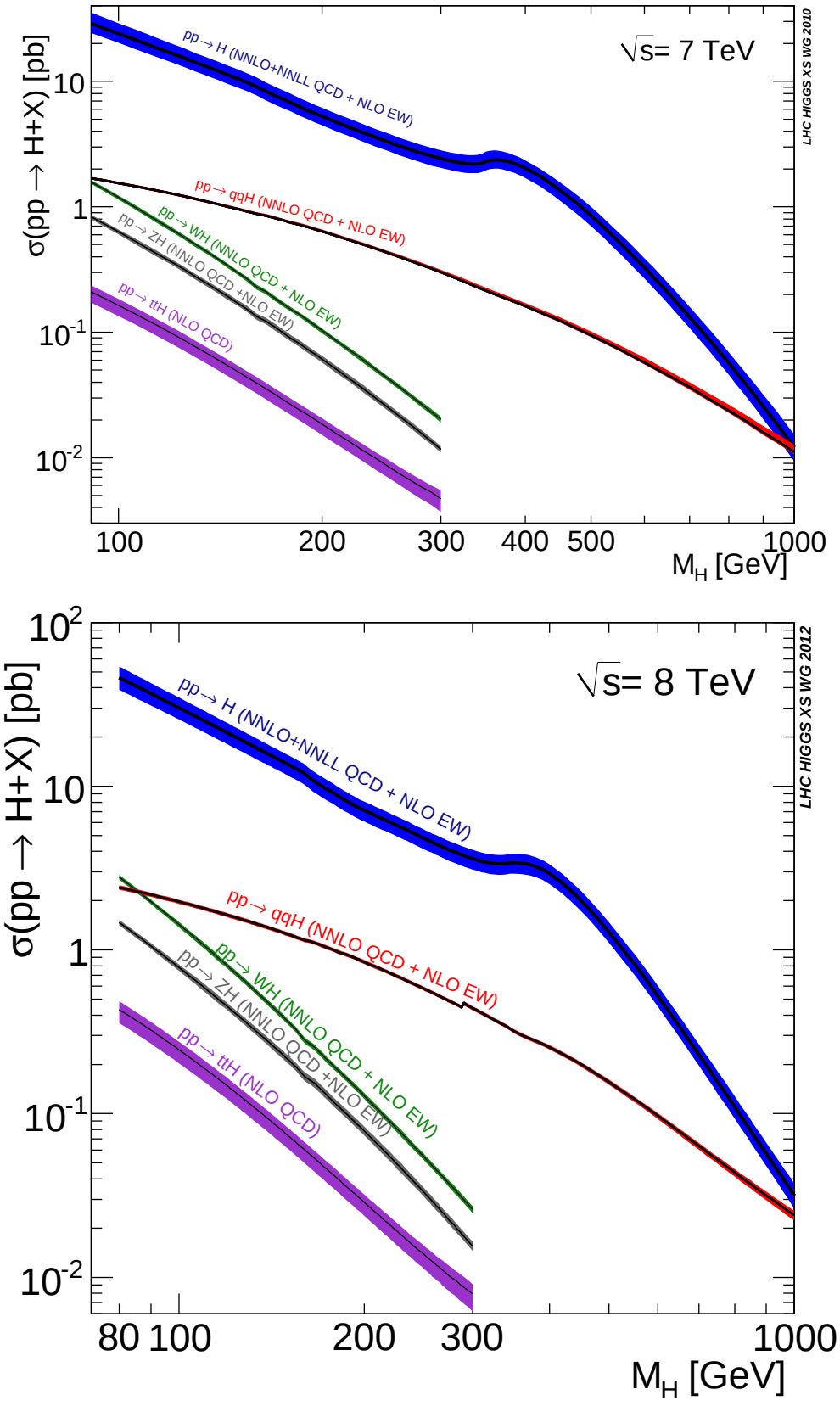


Figure 2.2: Production cross section for various production modes as a function of the Higgs mass m_H for both 7 and 8 TeV center-of-mass energies.

pairs, or Z+photon through loop vertices. Possible decay diagrams are shown in Figure 2.3. The leading-order partial width to fermions can be written as a function of the fermion mass [41, 67]:

$$\Gamma(H \to f\bar{f}) = \frac{1}{4\pi\sqrt{2}} G_\mu N_c m_H m_f^2 \beta_f^3 \tag{2.46}$$

G_μ is the Fermi constant. N_c is the color factor, which is 3 (1) for quarks (leptons). m_f is the mass of the fermion, and $\beta_f = \left(1 - \frac{4m_f}{m_H}\right)^{1/2}$ is the fermion velocity in the final state. Next-to-leading-order QCD corrections are substantial when decaying into light quark pairs, however, so this requires a slight modification in these cases [27, 52]. In the limit that $m_H >> m_q$:

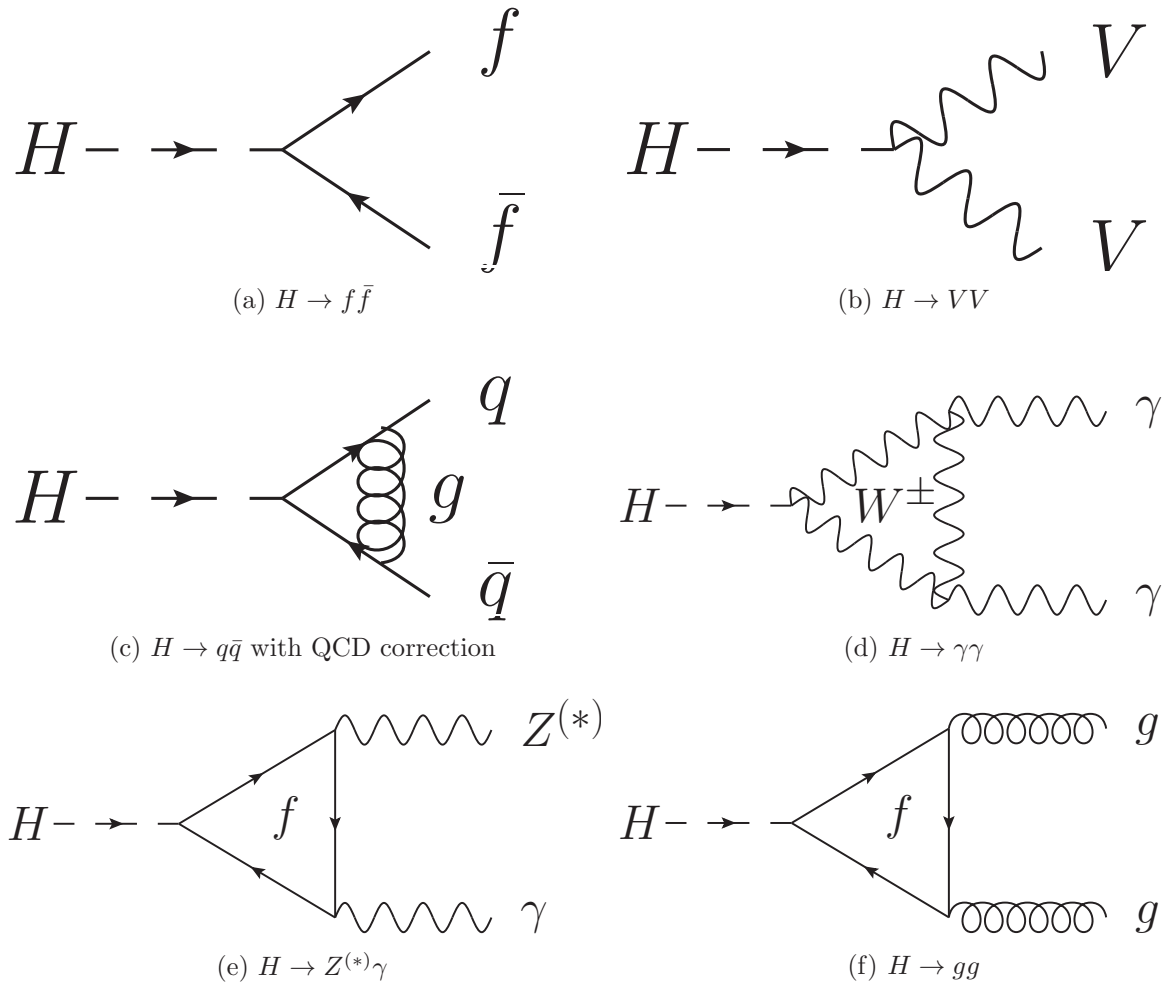
$$\Gamma_{\text{NLO}}(H \to q\bar{q}) = \frac{3}{4\pi\sqrt{2}} G_\mu m_H m_q^2 \left\{1 + \frac{4}{3}\frac{\alpha_s}{\pi}\left(\frac{9}{4} + \frac{3}{2}\log\frac{m_q^2}{m_H^2}\right)\right\} \tag{2.47}$$

The decay to on-shell electroweak gauge bosons can be written similarly [54, 67]:

$$\Gamma(H \to VV) = \frac{1}{16\pi\sqrt{2}} G_\mu m_H^3 \delta_V \sqrt{1-4x}(1 - 4x + 12x^2), x = \frac{M_V^2}{m_H^2} \tag{2.48}$$

Here $\delta_W = 2$ and $\delta_Z = 1$, and M_V is the mass of the boson. When the Higgs mass is small enough, the decay into vector bosons is still possible with one or both being virtual. For one virtual boson, the partial width is [53]:

$$\begin{aligned}
\Gamma(H \to VV^*) &= \tfrac{3}{16\pi} G_\mu^2 M_V^4 m_H \delta_V' R_T(x) \\
\delta_W' &= 1 \\
\delta_Z' &= \tfrac{7}{12} - 109\sin^2\theta_W + \tfrac{40}{9}\sin^4\theta_W \\
R_T(x) &= \tfrac{3(1-8x+20x^2)}{\sqrt{4x-1}}\arccos\left(\tfrac{3x-1}{2x^{3/2}}\right) - \tfrac{1-x}{2x}(2 - 13x + 47x^2) \\
&\quad -\tfrac{3}{2}(1 - 6x + 4x^2)\log x
\end{aligned} \tag{2.49}$$



(a) $H \to f\bar{f}$ (b) $H \to VV$

(c) $H \to q\bar{q}$ with QCD correction (d) $H \to \gamma\gamma$

(e) $H \to Z^{(*)}\gamma$ (f) $H \to gg$

Figure 2.3: Possible Higgs boson decays in the SM. (a) shows the leading-order decay into fermion pairs, (b) shows the two-body decay into vector bosons, (c) shows decay the into light quarks with QCD corrections, (d) shows the three-body decay into photon pairs through a $W^{\pm,*}$ loop, (e) shows the decay into a photon+$Z\gamma^*$ through a fermion loop, and (f) shows the decay into gluon pairs through a top-quark loop.

The decay into two off-shell bosons can be written in integral form [46]:

$$\begin{aligned}
\Gamma(H \to V^*V^*) &= \tfrac{1}{\pi^2} \int_0^{m_H^2} \frac{dq_1^2 M_V \Gamma_V}{(q_1^2 - M_V^2)^2 + M_V^2\Gamma_V^2} \int_0^{m_H^2} \frac{dq_2^2 M_V \Gamma_V}{(q_2^2 - M_V^2)^2 + M_V^2\Gamma_V^2} \Gamma_0(q_1^2, q_2^2) \\
\Gamma_0(q_1^2, q_2^2) &= \tfrac{1}{16\pi\sqrt{2}} G_\mu m_H^3 \delta_V \sqrt{\lambda(q_1^2, q_2^2; m_H^2)} \left\{ \lambda(q_1^2, q_1^2; m_H^2) + \tfrac{12 q_1^2 q_2^2}{m_H} \right\} \\
\lambda(x, y; z) &= (1 - x/z - y/z)^2 - 4xy/z^2
\end{aligned} \tag{2.50}$$

The loop-induced partial widths $H \to \gamma\gamma$ [41], $H \to Z\gamma$ [31, 26], and $H \to gg$ [45, 74] can also be written at leading order:

$$\Gamma(H \to \gamma\gamma) = \frac{G_\mu \alpha^2 m_H^3}{128\sqrt{2}\pi^3} \left| \sum_f N_C Q_f A_{1/2}^H(\tau_f) + A_1^H(\tau_W) \right|^2 \tag{2.51}$$

$A_{1/2}^H$ and A_1^H are form factors for spin-1/2 and spin-1 particles, respectively:

$$\begin{aligned}
A_{1/2}^H(\tau) &= 2[\tau + (\tau - 1) f(\tau)] \tau^{-2} \\
A_1^H(\tau) &= -[2\tau^2 + 3\tau + 3(2\tau - 1) f(\tau)] \tau^{-2}
\end{aligned} \tag{2.52}$$

The function $f(\tau)$ is defined as:

$$f(\tau) = \begin{cases} \arcsin^2(\sqrt{\tau}), & \tau \leq 1 \\ -\frac{1}{4} \left[\log\left(\frac{1 + \sqrt{1 - \tau^{-1}}}{1 - \sqrt{1 - \tau^{-1}}} \right) - i\pi \right]^2, & \tau > 1 \end{cases} \tag{2.53}$$

Finally, $\tau_i = \frac{m_H^2}{4M_i^2}$, with $i = f, W$. For $H \to Z\gamma$, the partial width is:

$$\Gamma(H \to Z\gamma) = \frac{G_\mu^2 M_W^2 \alpha m_H^3}{64\pi^4} \left(1 - \frac{M_Z^2}{m_H^2} \right)^3 \left| \sum_f N_C \frac{Q_f \hat{v}_f}{c_W} A_{1/2}^H(\tau_f, \lambda_f) + A_1^H(\tau_W, \lambda_W) \right|^2 \tag{2.54}$$

$\hat{v}_f = 2I_f^3 - 4Q_f s_W^2$ is the vector coupling of the Z to fermions, while $A_{1/2}^H$ and A_1^H are

again form factors as functions of $\tau_i = \frac{4M_i^2}{M_H^2}$ and $\lambda_i = \frac{4M_i^2}{M_Z^2}$:

$$\begin{aligned} A_{1/2}^H(\tau,\lambda) &= I_1(\tau,\lambda) - I_2(\tau,\lambda) \\ A_1^H(\tau,\lambda) &= c_W\left\{4(3-\tfrac{s_W^2}{c_W^2})I_2(\tau,\lambda) + [(1+\tfrac{2}{\tau})\tfrac{s_W^2}{c_W^2} - (5+\tfrac{2}{\tau})]I_1(\tau,\lambda)\right\} \end{aligned} \tag{2.55}$$

I_1 and I_2 are given by:

$$\begin{aligned} I_1(\tau,\lambda) &= \tfrac{\tau\lambda}{2(\tau-\lambda)} + \tfrac{\tau^2\lambda^2}{2(\tau-\lambda)^2}[f(\tau^{-1}) - f(\lambda^{-1})] + \tfrac{\tau^2\lambda}{(\tau-\lambda)^2}[g(\tau^{-1}) - g(\lambda^{-1})] \\ I_2(\tau,\lambda) &= -\tfrac{\tau\lambda}{2(\tau-\lambda)}[f(\tau^{-1}) - f(\lambda^{-1})] \end{aligned} \tag{2.56}$$

$f(\tau)$ is the same as Eq. 2.53, while $g(\tau)$ is given by:

$$g(\tau) = \begin{cases} \sqrt{\tau^{-1}-1}\arcsin\sqrt{\tau}, & \tau \geq 1 \\ -\frac{\sqrt{1-\tau^{-1}}}{2}\left[\log\left(\frac{1+\sqrt{1-\tau^{-1}}}{1-\sqrt{1-\tau^{-1}}}\right) - i\pi\right]^2, & \tau < 1 \end{cases} \tag{2.57}$$

Lastly, the partial width to gluons is:

$$\Gamma(H \to gg) = \frac{G_\mu \alpha_s^2 m_H^3}{36\sqrt{2}\pi^3}\left|\frac{3}{4}\sum_Q A_{1/2}^H(\tau_Q))\right|^2 \tag{2.58}$$

$A_{1/2}^H$ in this case is the same as Eq. 2.52 for $H \to \gamma\gamma$.

The branching ratio to any single mode is the ratio of the partial width to the total width, where the total width is the sum of all possible partial widths.

$$\mathcal{B}(H \to XX) = \frac{\Gamma(H \to XX)}{\sum_i \Gamma(H \to X_i X_i)} \tag{2.59}$$

The branching fraction for various modes as a function of the Higgs mass is shown in Figure 2.4. The values of the branching ratios for various decay modes assuming $m_H = 125$ are shown in Table 2.2, while the cross section for the major production processes is shown in Table 2.3 [55].

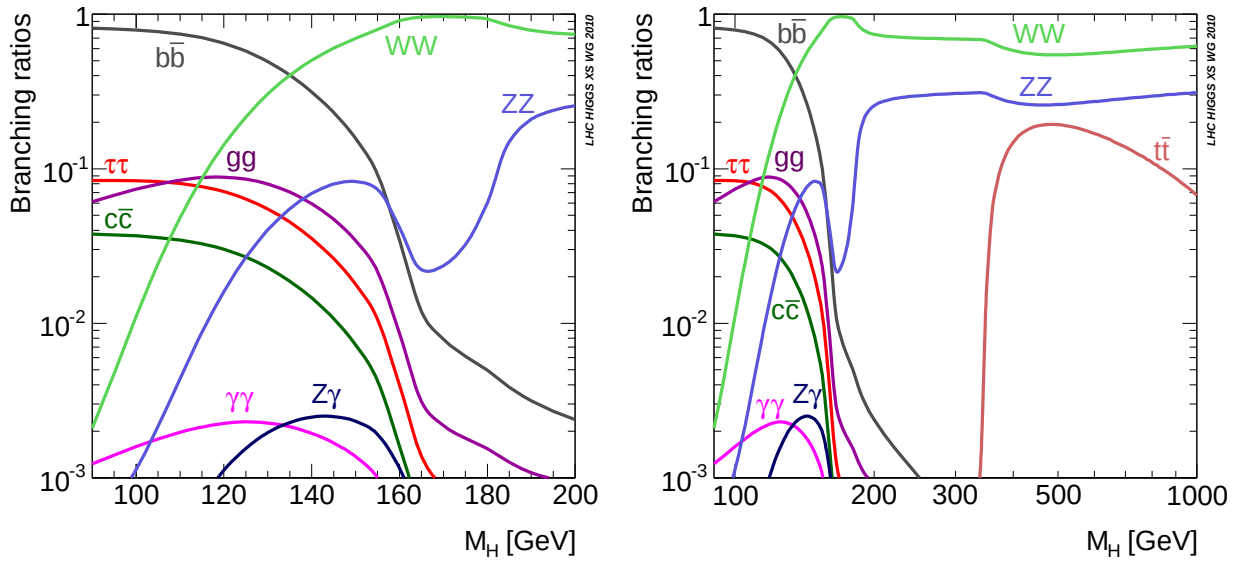


Figure 2.4: Branching ratio for various decay modes as a function of the Higgs mass m_H. The range of the left plot has an upper bound of 200 GeV to reveal the structure of the curves at low m_H.

Branching ratios ($m_H = 125$ GeV)						
$H \to WW^{(*)}$	$H \to \gamma\gamma$	$H \to ZZ^{(*)}$	$H \to b\bar{b}$	$H \to \tau^+\tau^-$	$H \to Z\gamma$	$H \to gg$
0.215	0.00228	0.0264	0.577	0.0632	0.00154	0.0857

Table 2.2: Branching ratios for various decay modes assuming $m_H = 125$ GeV

Production cross sections (pb, $m_H = 125$ GeV)					
	ggF	VBF	WH	ZH	$t\bar{t}H$
7 TeV	15.32	1.222	0.5729	0.3158	0.0863
8 TeV	19.52	1.578	0.6966	0.3943	0.1302
14 TeV	49.85	4.180	1.504	0.8830	0.6113

Table 2.3: Cross sections for the major production processes assuming $m_H = 125$ GeV

CHAPTER III

LHC & ATLAS

"Engineering is the art of organizing and directing men and controlling the forces and materials of nature for the benefit of the human race."

- Henry G. Stott

3.1 LHC

The LHC [43] is a proton-proton collider located in Geneva, Switzerland at the European Organization for Nuclear Research (CERN), which began operation in 2010. Data has been taken at the three center of mass energies $\sqrt{s}$ = 900 GeV, 7 TeV, and 8 TeV. The 26.7 km circumference collider will have a long shutdown from 2013 until 2014, at which time it will undergo upgrades aimed to increase the center of mass energy to 13 or 14 TeV. Beyond this, two additional long shutdowns in 2018 and 2022 are planned in order to increase the instantaneous luminosity to 2×10^{34} cm^{-2}s^{-1} and 5×10^{34} cm^{-2}s^{-1}, respectively.

There are four primary detectors along the ring, shown in Figure 3.1: ALICE, LHCb, ATLAS, and CMS. ALICE is designed to look at heavy ion collisions. LHCb is a dedicated b-physics detector that operates at low luminosity. ATLAS and CMS are general purpose detectors with design goals to find new physics in proton-proton interactions, but are also capable of reconstructing heavy ion collisions and b-quark decays. The work in this thesis

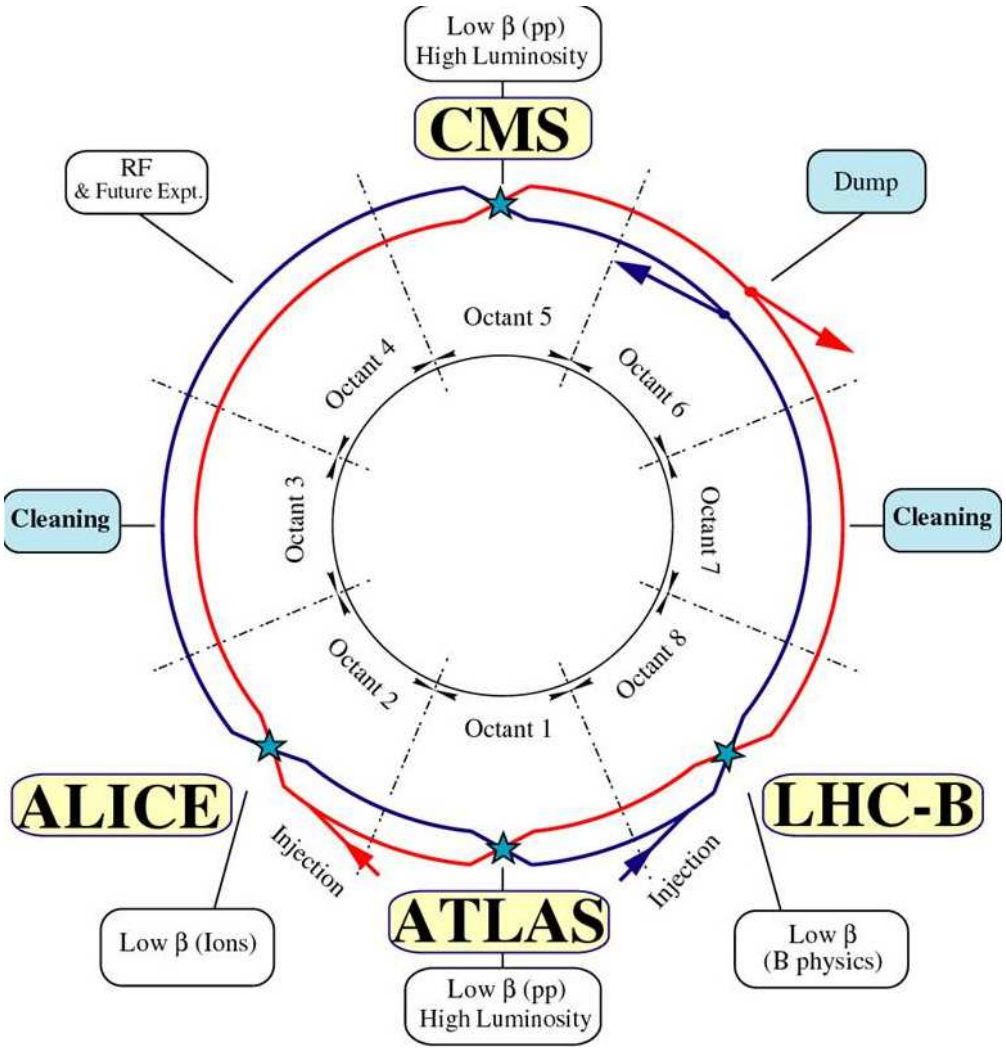


Figure 3.1: Layout of the LHC ring showing the positions of each experiment at the collision points as well as injection, beam dump, and beam cleaning regions.

was done with up to 4.8 fb^{-1} of 7 TeV and 21 fb^{-1} of 8 TeV data collected with the ATLAS detector.

The LHC uses protons derived from hydrogen atoms having been stripped of their valence electrons. The chain of accelerators used to inject the protons into the LHC is shown in Figure 3.2. Protons are first accelerated to 50 MeV in the Linac2 linear accelerator. From here, they are injected into the Proton Synchrotron Booster (PSB) and accelerated to 1.4 GeV. At this point the Proton Synchrotron (PS) is able to group protons into bunches of 1.6×10^{11} and accelerate them to 25 GeV. The final stages are the Super Proton Synchrotron

Parameter	Value	Definition
N_b	1.6×10^{11}	Protons per bunch
n_b	1368	Bunches per beam
f_{rev}	11.25 kHz	Revolution frequency [s^{-1}]
γ_r	4260	Relativistic γ factor
ϵ_n	2.5 μm	Transverse emmittance
β^*	0.6	β function at interaction point
θ_c	290 μrad	Crossing angle at interaction point
σ_z	9.4 cm	RMS bunch length
σ^*	19 μm	RMS transverse beam size

Table 3.1: Beam parameters for a typical 8 TeV run.

accelerating to 450 GeV and injection into the LHC. This will inject two beams, one travelling clockwise and one counter-clockwise. The LHC ramp to nominal energy of at least 3.5 TeV per beam takes around 20 minutes.

The schematics of the superconducting niobium-titanium dipole and quadrupole magnet assemblies used in the LHC ring are shown in Figure 3.3. The nominal magnetic field strength of 8.36 T generated from the 11,700 A electric current in each of the 1232 dipoles bends the path of the proton beams while they travel through the ring. This is made possible by cooling the niobiuim-titanium coils to 1.9 K with liquid helium. There are 392 quadrupole magnets that complement the dipoles by stabilizing and focusing the beams.

The instantaneous luminosity is given as a function of the beam parameters and is shown below. The definitions of the parameters and their values for a typical 8 TeV run are shown in Table 3.1.

$$\mathcal{L}_0 = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\epsilon_n \beta^*} \left\{ 1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right\}^{-1/2} \tag{3.1}$$

During collisions, the beam intensity will decay exponentially, such that the integrated luminosity of a run $\mathcal{L}_{\text{int}}$ is a function of the characteristic decay time τ_L and the length of the run T_{run}.

$$\mathcal{L}_{\text{int}} = \mathcal{L}_0 \tau_L \left\{ 1 - \exp\left(-\frac{T_{\text{run}}}{\tau_L} \right) \right\} \tag{3.2}$$

Figure 3.4 shows the total integrated luminosity by day for both 7 TeV and 8 TeV runs,

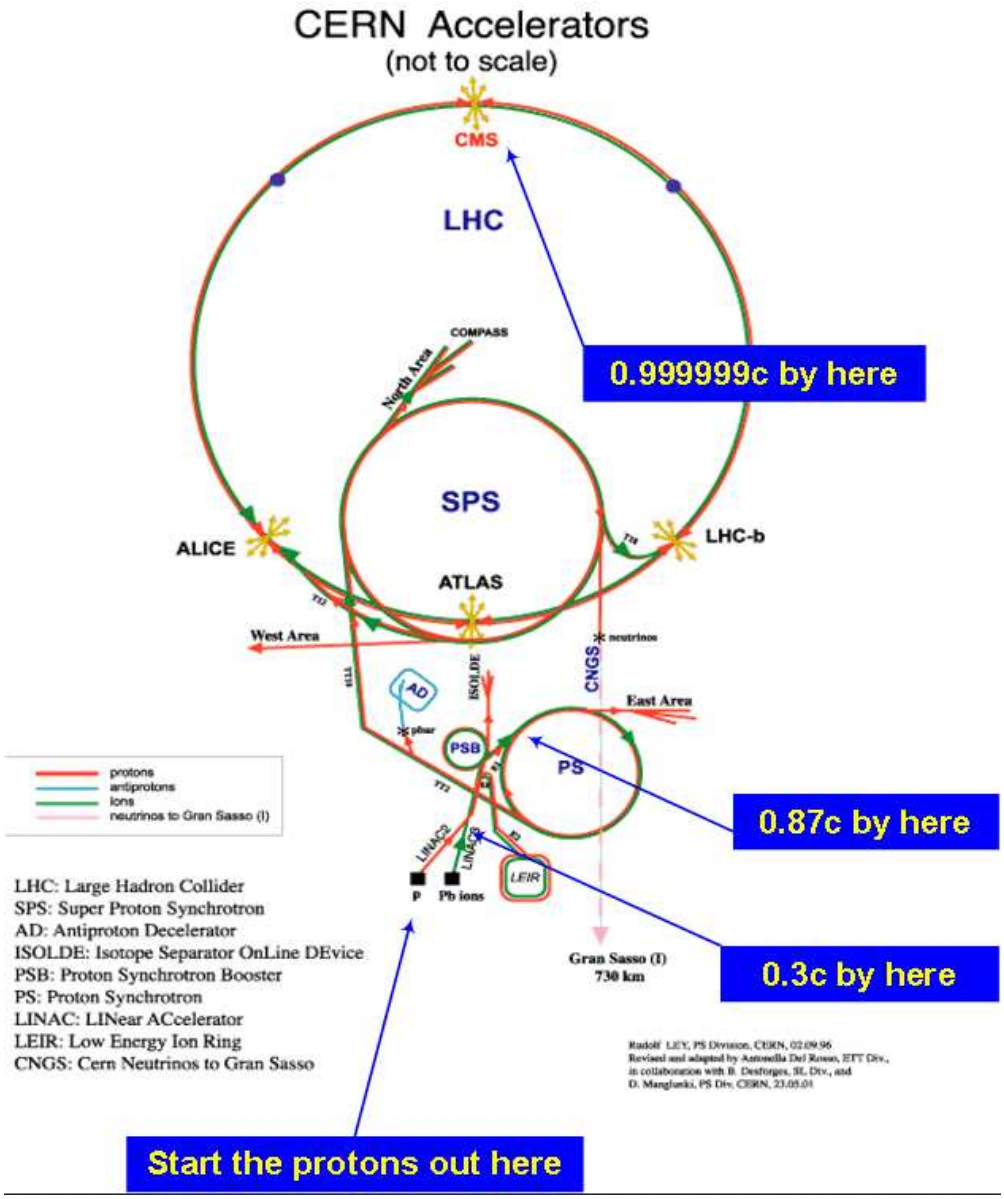


Figure 3.2: The chain of accelerators at CERN used to inject proton beams into the LHC.

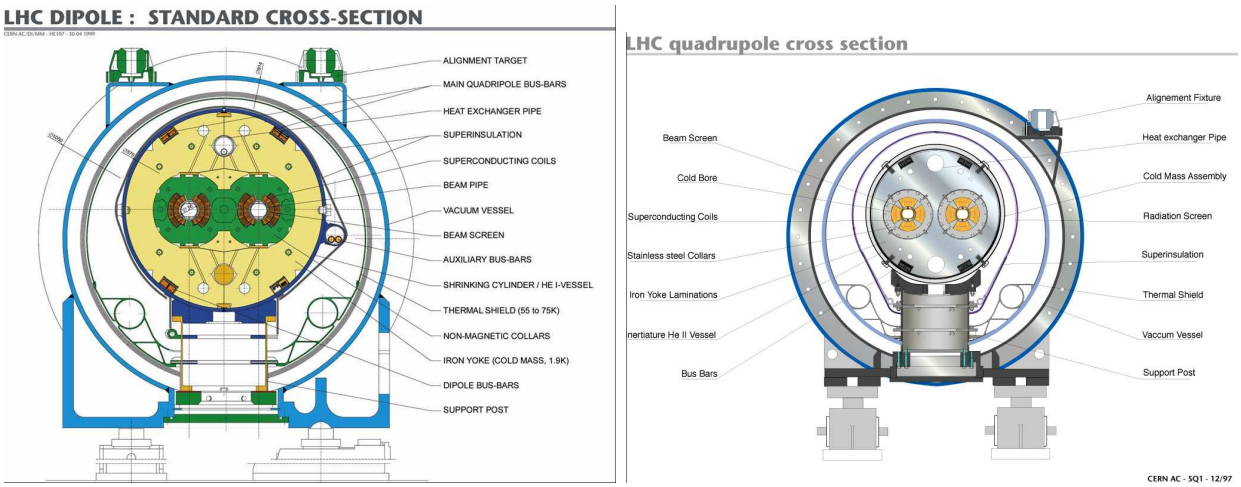


Figure 3.3: Cross sections of the superconducting dipole (left) and quadrupole (right) magnet assemblies used in the LHC ring.

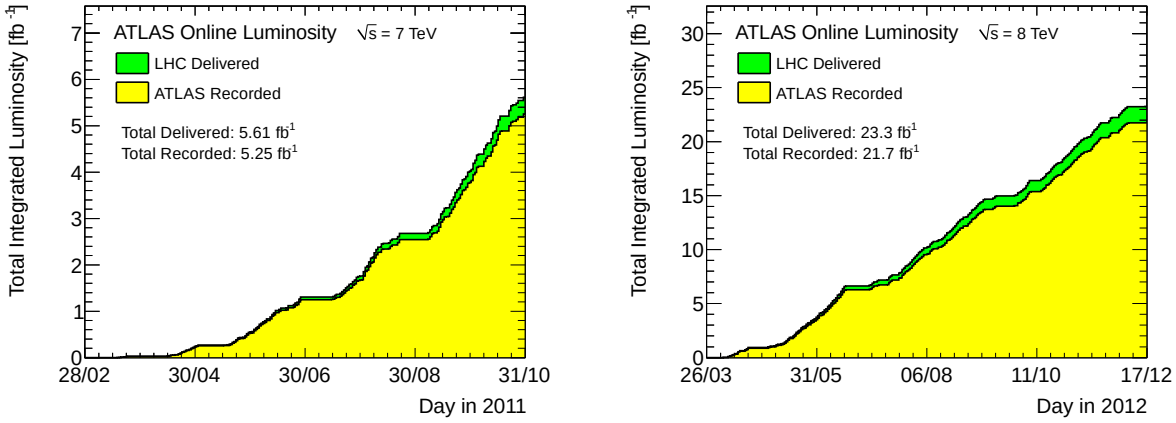


Figure 3.4: Integrated luminosity by day in the 7 (left) and 8 (right) TeV runs.

both the total delivered by the LHC and that which was recorded by ATLAS.

3.2 ATLAS Detector

The ATLAS (A Toroidal LHC Apparatus) detector [20] is one of two general purpose detectors on the LHC ring. It is a hermetic 4π coverage detector designed primarily to search for new physics in proton-proton interactions. The layout of the detector is shown in Figure 3.5. Starting from the beam pipe and working outward, ATLAS is comprised of an inner detector inside a superconducting solenoid magnet for vertex and track reconstruction,

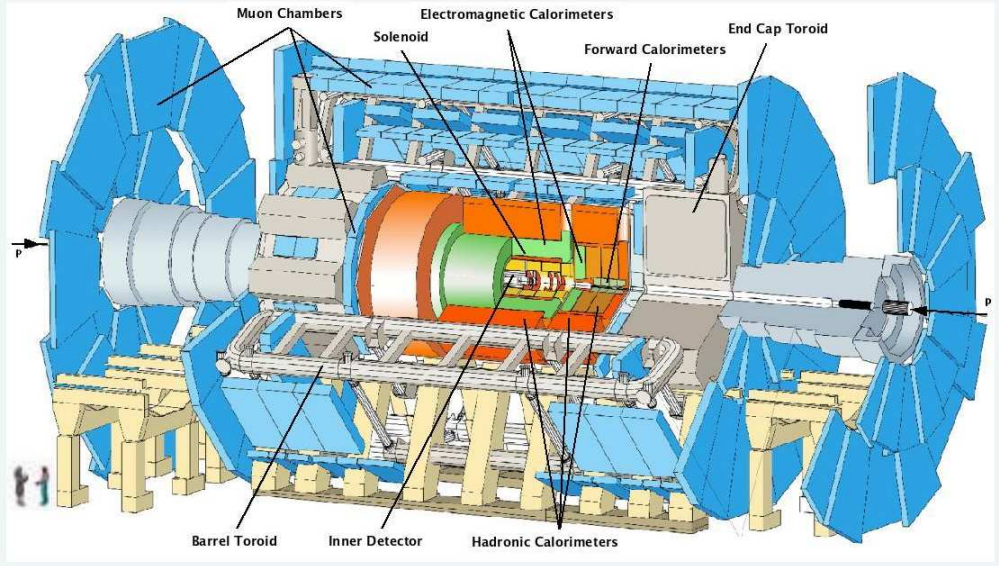


Figure 3.5: Layout of the ATLAS detector showing the major subdetectors.

a liquid argon (LAr) electromagnetic (EM) calorimeter for EM particle energy measurement, LAr and tile hadronic calorimeters for hadronic particle energy measurement, and a muon spectrometer for the reconstruction of muon tracks. The muon spectrometer is complemented by eight air-core magnet loops in the barrel and two toroidal magnets in the end-cap.

The coordinate system of the detector is denoted such that the origin is in the center of the detector at the nominal collision point. The transverse x-y plane lies perpendicular to the beam line, which is identical to the z-axis. The two halves of the detector split by the x-y plane are labeled A-side and C-side. **A**-side is a reference to the half being oriented towards the Geneva **A**eroport, while **C**-side references its orientation towards **C**harly's pub in Saint Genis-Pouilly, France. The positive y-axis points upwards, the positive x-axis points towards the center of the LHC ring, and the positive z-axis towards the A-side of the detector. The alternative r-ϕ-θ coordinate system is defined with $r = \sqrt{x^2 + y^2}$, $\cos\phi = \frac{x}{\sqrt{x^2+y^2}}$, and $\cos\theta = \frac{z}{\sqrt{x^2+z^2}}$. For relativistic particles, it is useful to map θ to the pseudo-rapidity $\eta = -\ln\tan\theta/2$, or alternatively to the rapidity $y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$ when the particle's mass is non-negligible compared to its momentum. $\Delta\eta$ and Δy are invariant under boosts in the z-axis. The quantity $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ is used to quantify the $\eta - \phi$ separation between

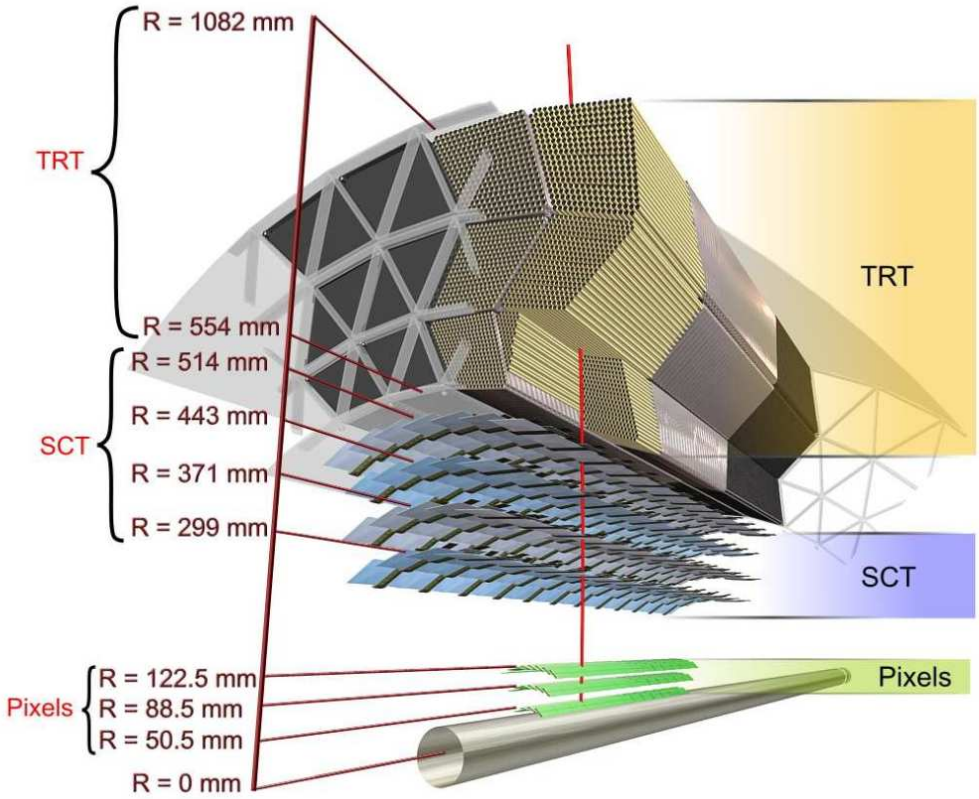


Figure 3.6: Layout of the inner detector in the barrel showing the Pixel, SCT, and TRT detectors and their geometrical configuration.

two particles.

3.2.1 Inner Detector

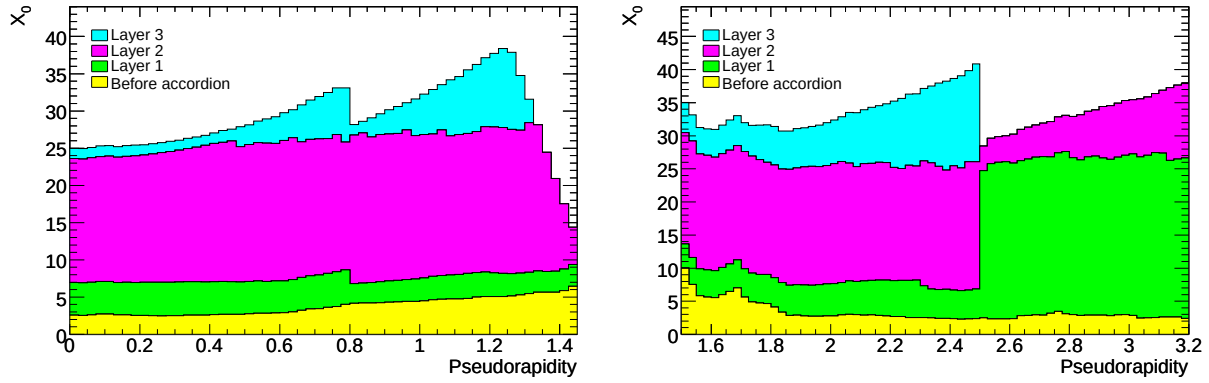
The inner detector provides tracking and vertex reconstruction for charged particles. It's layout is shown is Figure 3.6. It is composed of a high granularity Pixel detector, followed by a Semi-Conductor Tracker (SCT), and Transition Radiation Tracker (TRT). All three are divided into barrel and end-cap components, providing coverage within $|\eta| < 2.5$ in the silicon detectors and $|\eta| < 2$ in the TRT. The entire inner detector is surrounded by a nearly homogeneous 2T magnetic field produced by a solenoid magnet. The field points is oriented with the beam axis, such that positively charged particles passing perpendicular to the field are bent in the negative ϕ direction.

Being closest to the interaction point (IP), the Pixel detector has the highest granularity

and is designed to provide accurate primary and secondary vertex reconstruction. It has three concentric cylindrical layers in the barrel between $50.5 < R < 122.5$ mm and $0 < |z| < 400.5$ mm, and three disks in each end-cap between $88.8 < R < 149.6$ mm and $495 < |z| < 650$ mm. The pixel size in $R - \phi \times z$ is $50{\times}400$ μm^2 throughout, with a total of 80.4 million readout channels. This yields resolutions of 10 μm in R-ϕ and 115 μm in z for the barrel and 10 μm in R-ϕ and 115 μm in R for the end-cap.

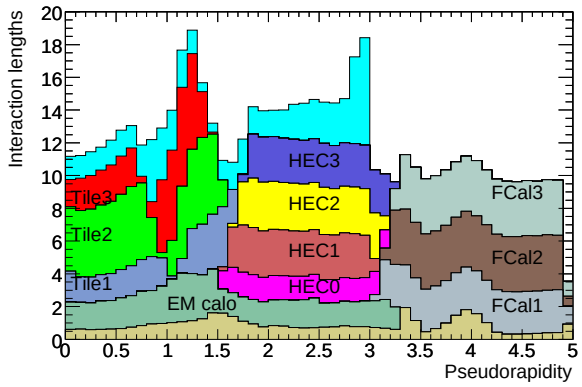
The SCT detector sits outside of the Pixel, with four cylindrical layers between $299 < R < 514$ mm and $0 < |z| < 749$ mm in the barrel, and nine disks in each end-cap between $275 < R < 560$ mm and $839 < |z| < 2735$ mm. The layers are composed of pairs of 6.4 cm daisy-chained silicon strips with a density of 80 μm per strip. The pairs are offset at an angle of 40 mrad, with one strip parallel to the beam-line. Each strip provides one dimensional resolution, such that the strips in stereo offer three dimensional space point measurements when considered with their position in R for the barrel or z for the end-cap. Having a larger lever arm than the Pixel, the SCT offers a higher momentum resolution.

The outer layer of the inner detector is the TRT, which is composed of straw tracker drift tubes filled with a 70%/27%/3% $Xe/CO_2/O_2$ gas mixture operating at two different thresholds to distinguish between track hits and transition radiation, and sits between 554-1082 mm from the IP. The 4 mm diameter tubes are aligned parallel to the beam line in the barrel and perpindicular in the end-cap. The tubes have a 31 μm tungsten wire as the anode strung through the center. They have a carbon fiber outer layer for support, the inside of which is 25 μm thick polyimide film with a 0.2 μm aluminum coating that acts as the cathode. Between the tubes are materials with varying indices of refraction to induce transition radiation from relativistic charged particles as they pass through. Because this radiation is proportional to the boost factor of the particle, it offers some discrimination between electrons and charged hadrons. The charged particle will also ionize the gas in the straw tube, and the drift time from the subsequent electron avalanche can be measured, effectively giving a measurement of the path that the particle traverses in the r-ϕ plane.



(a)

(b)



(c)

Figure 3.7: Breakdown of the radiation length by EM calorimeter layer (a, b) and interaction length by hadronic calormeter component (c) versus η.

3.2.2 Calorimeter

The ATLAS calorimeter is designed to measure the energy of all electromagnetic and hadronic particles produced in an event. It has a total radiation length (X_0) of > 22 (> 24) in the barrel (end-cap) for EM particles out to $|\eta| < 3.2$, and an interaction length (λ) of 9.7 at $\eta = 0$ for hadronic particles with a coverage out to $|\eta| < 4.9$. This is shown in more detail in Figure 3.7. All calorimeters used incorporate a sampling material to induce a particle shower and an absorbing material to measure the total energy of the shower. The technology used in each is optimized for the expected particle density with respect to θ and the intended interacting particle.

The EM targeted calorimeter is the inner most detector in both barrel and end-cap. The barrel extends to $|\eta| < 1.475$, and the end-cap between $1.375 < |\eta| < 3.2$. It is composed of alternating layers of lead absorbers, spacers, and copper electrodes, which use liquid argon as an ionizing medium. Liquid argon was chosen for radiation hardness, speed, and signal linearity. Non-magnetic stainless steel is inserted every few layers to supply rigidity to the modules. After sandwiching the layers together, they are bent into an accordion shape in the azimuthal and longitudinal directions for the barrel and end-cap, respectively, to ensure a gap-free environment for particles passing through. As particles pass through, the lead will induce a shower of secondary particles which ionize the liquid argon. A 2 kV electric field applied to the material by the electrodes allows this signal to be collected and read out so that one can measure the shower shape and total energy deposited in each cell. The number of radiation lengths of the calorimeter ranges between 22-30 X_0 for the electromagnetic barrel (EMB) and 24-38 X_0 for the electromagnetic end-cap (EMEC), depending on η. Because there is a cryostat with about 1.5 X_0 in front of the calorimeter, a presampler is installed to correct for energy loss in the region $|\eta| < 1.8$.

The technology used in the hadronic calorimeter is different for the barrel and end-cap. In the barrel is the tile calorimeter (TileCal), which uses plastic scintillators as the sampling material and steel as the absorber. Because the EMB sits in front of this and has two interaction lengths, most of the showering will already have taken place before the particles reach the tile. In any case, an additional eight interaction lengths of material are used, which provides sufficient containment for measurement of jets and missing energy. Fiber optic cables readout the scintillator signals to photomultipliers, from which the signal can be digitized.

The hadronic end-cap calorimeter (HEC) uses copper plates as the absorber, with sampling accomplished similar to the EM calorimeter. It is divided into two wheels, the front holding 25 copper plates, each 25mm thick, and the rear holding 17 plates at 50mm thickness. Including the EMEC, there is a total of 12 interaction lengths in the HEC.

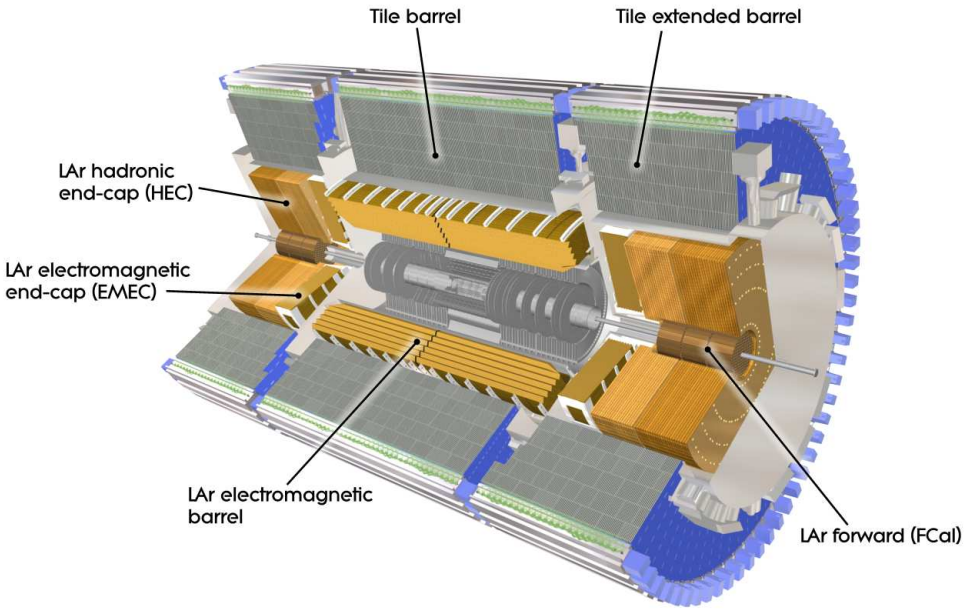


Figure 3.8: Layout of the calorimetry system

Finally, the forward calorimeter (FCAL) provides calorimetry in the region $3.2 < |\eta| < 4.9$. It is separated into one EM and two hadronic modules. Each cell is a copper cathode with a liquid argon medium. The EM module uses copper rods as the anode, while the hadronic modules use tungsten, providing between 9.5-15 interaction lengths.

3.2.3 Muon Spectrometer

The ATLAS Muon Spectrometer sits outside of the calorimeter, and was designed with the intent of measuring 1 TeV muons to 10% precision in $1/p_T$. Monitered Drift Tube (MDT) chambers are composed of 3 cm tubes filled with a 93% Ar/7% CO_2 mixture pressurized to four bars (three bars over atmosphere), with a 50 μm diameter tungsten wire anodes in the center. The anode is held at a potential of 3080 V with respect to the tube casing, which is at ground voltage. Muons passing through the tube ionize the gas, causing an electron avalanche that is collected by the anode. The collection time can be modeled as a function of the drift radius, giving the muon's closest approach to the wire, and thus one dimension of its position in the tube. The MDT chambers have three layers in the barrel

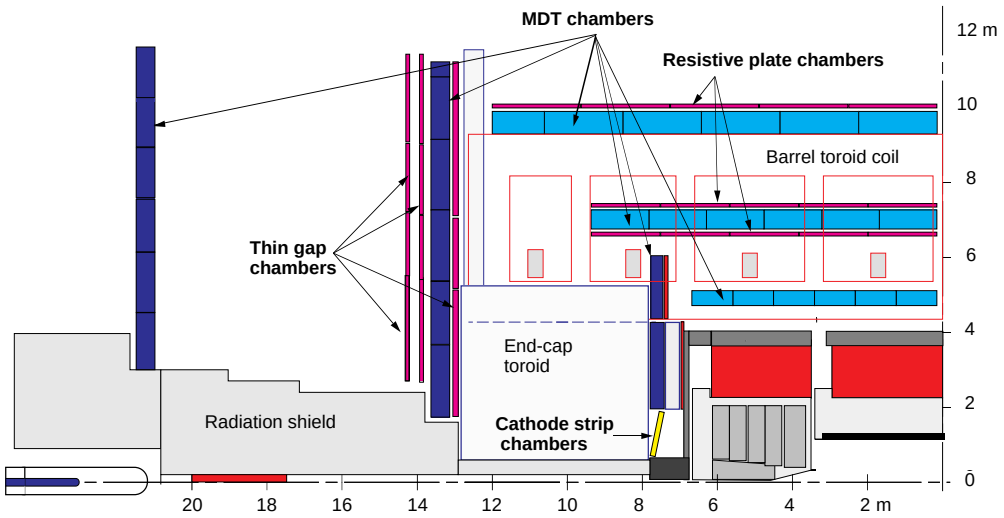


Figure 3.9: Cross section of the muon spectrometer in both barrel and end-cap.

between $R1 < R < R2$, and three wheels in each end-cap. This provides coverage out to $|\eta| < 2.7$, with the exception of the inner most end-cap wheel, which extends to 2.0 in $|\eta|$. In place of the MDTs here are high granularity Cathode Strip Chambers (CSC) in the region $2 < |\eta| < 2.7$. Each chamber consists of 192 cathode strips aligned perpendicular to anode wires, with 48 more strips parallel to them. Charged particles induce an avalanche of electrons in the gas that fills the chamber, which gives measurement of the hit position. In both barrel and end-cap, the MDT tubes are oriented perpindicular to both the beam pipe and the radial axis. Eight air core toroidal magnets in the barrel and two end-cap toroids produce a magnetic field in the ϕ direction, such that muons are bent in θ.

Resistive Plate Chamber (RPC) detectors are used in the barrel for the muon trigger. Resistive plates separated by 2 mm are held at a voltage of 9.8 kV. A 94.7/5/0.3% $C_2H_2F_4$/Iso-C_4H_{10}/SF_6 gas mixture between the plates ionizes as muons pass through it, giving the muon position with a $> 98.5\%$ reconstruction efficiency. These detectors sit above or below the MDT chambers as in Figure 3.9.

The triggering in the end-caps is done by Thin Gap Chambers (TGCs). These chambers also provide a measurement of the muon's azimuthal coordinate. 55/45% CO_2/n-pentane gas is held between two or three plate layers. Azimuthally segmented anode wires between

the plates are held at a voltage of 2.9 kV, which collects the ionization produced by muons passing through.

3.3 Forward Detectors

There are three special purpose forward detectors in ATLAS. LUCID (LUminosity measurement using Cerenkov Integrating Detector) is placed at ±17m from the IP at a radial distance of 10 cm ($|\eta| \approx 5.8$) from the beamline. It is composed of 20 aluminum gas vessels pointing to the IP, which contain C_4F_{10} pressurized to 1.2-1.4 bar. Photomultiplier tubes housed in the gas vessels measure the Cerenkov radiation of particles resulting from p-p inelastic scattering. ZDC (Zero-Degree Calorimeter) is ±140m from the IP and measures forward ($|\eta| > 8.3$) neutrons from heavy ion collisions, which is correlated with the centrality of the collisions. Finally, ALFA (Absolute Luminosity For ATLAS), sitting at ±240m inside Roman pots, is constructed from scintillating fiber trackers. A luminosity measurement is extracted via the optical theorem, which relates the elastic scattering amplitude in the forward direction with the total cross section.

CHAPTER IV

Event Reconstruction

"Measure what is measurable, and make measurable what is not so."

- Galileo Galilei

4.1 Trigger System

Given high instantaneous design luminosity at the LHC, $L \sim 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$, corresponding to 25ns per bunch crossing, the full amount of raw data produced would be impossible to record and analyze given today's technology and resources; even if possible to record, the vast majority of collisions are uninteresting. The ATLAS trigger system is designed to selectively reduce the information throughput by rapidly filtering out uninteresting collisions at high efficiency. This is achieved through a three part sequential process consisting of a Level 1 (L1) hardware trigger and a software based High Level Trigger (HLT), which is itself divided into a Level 2 (L2) and Event Filter (EF). The full chain reduces the 40 MHz nominal collision rate down to ∼400 Hz with a bandwidth of 500 MB/s.

4.1.1 L1 Trigger

The L1 hardware trigger is designed to pass events at around 75 kHz using information from the calorimeter and muon spectrometer. Raw data is cached in Read Out Buffer (ROB)

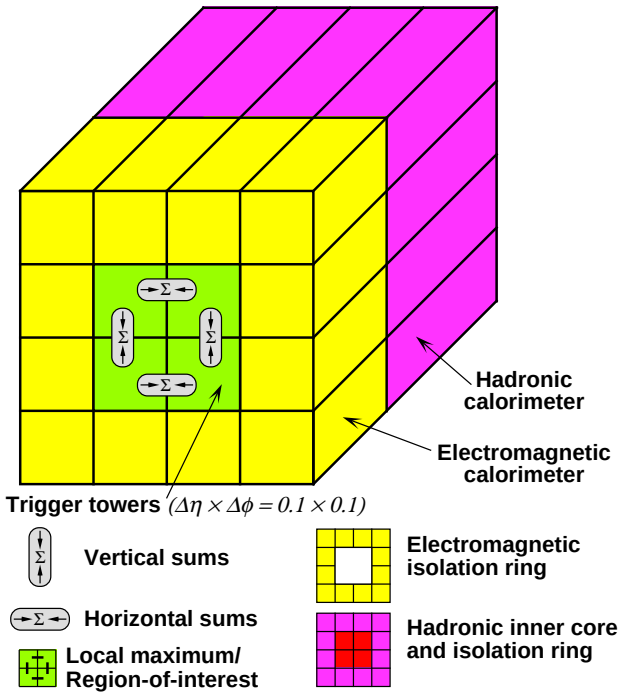


Figure 4.1: Representation of the trigger towers used in the e/γ trigger algorithm.

queues in the detector electronics. One or more ROBs are grouped into a Read Out System (ROS), from which the trigger logic can pull information. Several Regions of Interest (RoI) are defined for each event within the L1 algorithms for use in the HLT.

The L1 calorimeter trigger coarsely sums transverse energy in towers of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the central region and in larger and more irregular sizes in the $2.5 < |\eta| < 4.9$ region. The energy is summed separately for the EM and hadronic calorimeters. Two processor systems, the cluster processor and the jet and energy-sum processor, run in parallel to make a trigger decision. The cluster processor is used for e/γ/τ based triggers. It identifies clusters of 2×2 EM towers as RoIs, in which any of the 1×2 or 2×1 tower clusters must pass a programmable threshold, as shown in Figure 4.1. The 12 tower ring surrounding each RoI is also used for isolation thresholds. The jet and energy-sum processor works similarly, but identifies larger 4×4, 6×6, and 8×8 tower RoIs centered around local maximums of 2×2. This processor sums transverse energy from both EM and hadronic towers in order to form the trigger decision.

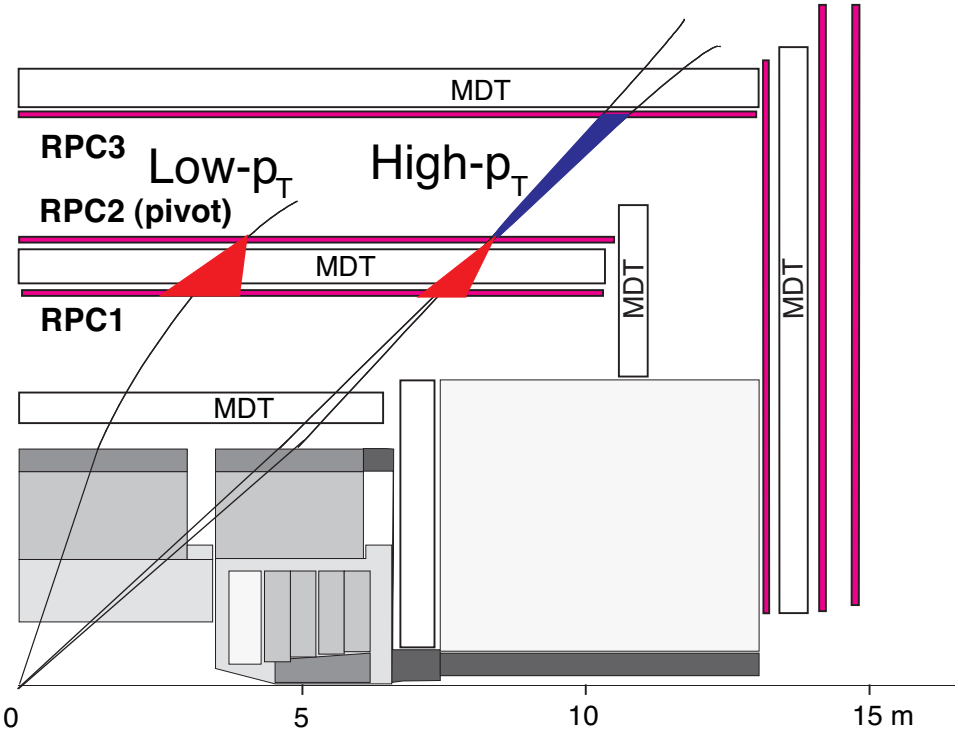


Figure 4.2: Graphical view of the muon trigger algorithm, showing how the muon p_T can be estimated based on hits in the chambers.

The L1 muon trigger uses information from the RPC in the barrel and TGC in the end-caps, which are composed of planes of between two and four chambers. Hits in these detectors are used to give a rough reconstruction of the p_T, η, and ϕ of the muon, as shown in Figure 4.2.

4.1.2 High Level Trigger

The HLT uses full inner detector, calorimeter, and muon spectrometer information within the RoIs defined at L1, with few exceptions. At L2, the inner detector uses two inside-out algorithms starting from the Pixel and SCT detectors to reconstruct tracks. The tracks are then processed by a Kahlman filter, after which TRT information is used to improve the p_T resolution and electron identification performance. The inner detector EF reconstruction shares software with the offline reconstruction, which is extended to support RoIs. The algorithm is similar to L2: track seeds from the Pixel and SCT are preselected with momentum

and impact parameter requirements, to which a combinatorial Kahlman filter is applied. Low quality tracks are rejected, and the remaining ones use additional TRT information to perform a final fit, from which the track parameters are extracted. The final trigger decision at the EF level is able to use the full information from the event to form a decision and is not restricted to the RoIs as in L2.

4.2 Electrons

Electrons are reconstructed based on energy deposited in the EM calorimeter as well as inner detector tracks. Information from the hadronic calorimeter is also used to discriminate from hadronic particles faking electrons. Two primary algorithms are used. The standard algorithm performs an outside-in reconstruction, first reconstructing EM clusters then matching to inner detector tracks. The second non-standard algorithm is inside-out, first reconstructing tracks and second projecting to the EM calorimeter to search for a matching cluster. The non-standard algorithm is not covered here, though it should be mentioned that it is useful for low energy electron reconstruction and for finding electrons within jets.

The EM seeded algorithm uses clusters with $E_T > 3$ GeV. Inner detector tracks are extrapolated to the EM calorimeter. Tracks are required to match the cluster position within a window of $\Delta\eta \times \Delta\phi = 0.05 \times 0.1$. The ratio of energy deposited in the EM calorimeter to track momentum (E/p) is required to be less than 10. At this point various quality selections can be applied to clusters and tracks. A simple cut based method classifies electrons into three classes of quality, each with various degrees of true electron efficiency and fake rejection. The exact cut based selections are optimized in seven η bins and six p_T bins.

- Loose electrons are defined as electron candidates passing requirements on hadronic leakage as well as lateral shower shape and cluster width variables based on the middle layer of the EM calorimeter. These offer the best efficiency, but the lowest background

rejection.

- In addition to the selections for loose electrons, medium electrons are required to pass quality cuts on strips in the first layer of the EM calorimeter as well as inner detector tracks. In particular, cuts are applied to $E_{max2} - E_{min}$, the difference between the second most energetic and least energetic cell in the first EM layer; to $R_{max2} = E_{max2}/(1 + 9 \times 10^{-3} \frac{E_T}{\text{GeV}})$; to w_{stot}, the shower width over strips covering 2.5 cells of the second layer; to w_{s3}, the shower width over three strips centered around the most energetic cell; to F_{side}, the fraction of energy deposited around the shower core; to the number of hits in the pixel detector; to the sum of hits in the silicon detectors (pixel + SCT); and to d_0, the transverse impact parameter of the track. These selections increase background rejection by a factor of 3 or 4 and reduce reconstruction efficiency by $\sim 10\%$.

- In addition to selections on medium electrons, tight electrons must satisfy cuts on the number of vertex layer hits; on the number of TRT hits; on the ratio of high threshold to total TRT hits; on the difference in η and ϕ between the cluster and track; and on E/p. At this point two different selections can be chosen depending on the expected electron topology: A tight calorimeter isolation requirement based on cells within $\Delta R < 0.2$ of the EM cluster, or tighter TRT selections based on already selected TRT variables.

4.2.1 Gaussian Sum Filter

The energy loss an electron incurs while traversing through matter has a significant impact on both the efficiency of reconstruction and the quality of the reconstructed track parameters. The material budget of ATLAS as shown in Figure 4.3 highlights the need to mitigate this effect. An analytical parametrization developed by Beth and Heitler [50] describes the probability for an electron to retain a fraction $z = \frac{E_f}{E_i}$ of it's initial energy after

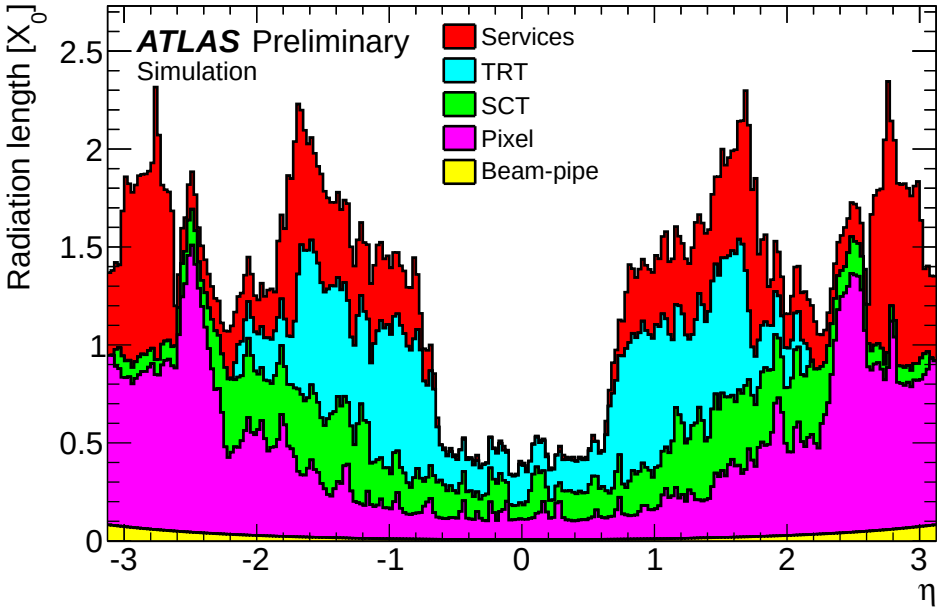


Figure 4.3: Composition of material in the ATLAS inner detector in number of radiation lengths versus η.

traversing t radiation lengths though a material:

$$f(z) = \frac{1}{\Gamma(\frac{t}{\ln 2})}[-\ln z]^{\frac{t}{\ln 2}-1} \tag{4.1}$$

Secondary processes such as the Landau-Pomeranchuk-Migdal or Ter-Mikaelian effects become important at higher energies, such that the validity of Eq. 4.1 breaks down and GEANT4 simulations must be used to correctly describe the energy loss. A Gaussian Sum Filter (GSF) technique [4] can be employed to refit the track to better account for electron bremsstrahlung radiation. Assuming the track's trajectory can be approximated by a sum of weighted Gaussian functions, a Kahlman filter can be run in parallel on each, such that each component represents a different contribution to the full Beth-Heitler spectrum. The standard electron reconstruction is then performed with the refitted track. Figure 4.4 show the impact on the momentum and impact parameter resolution, respectively.

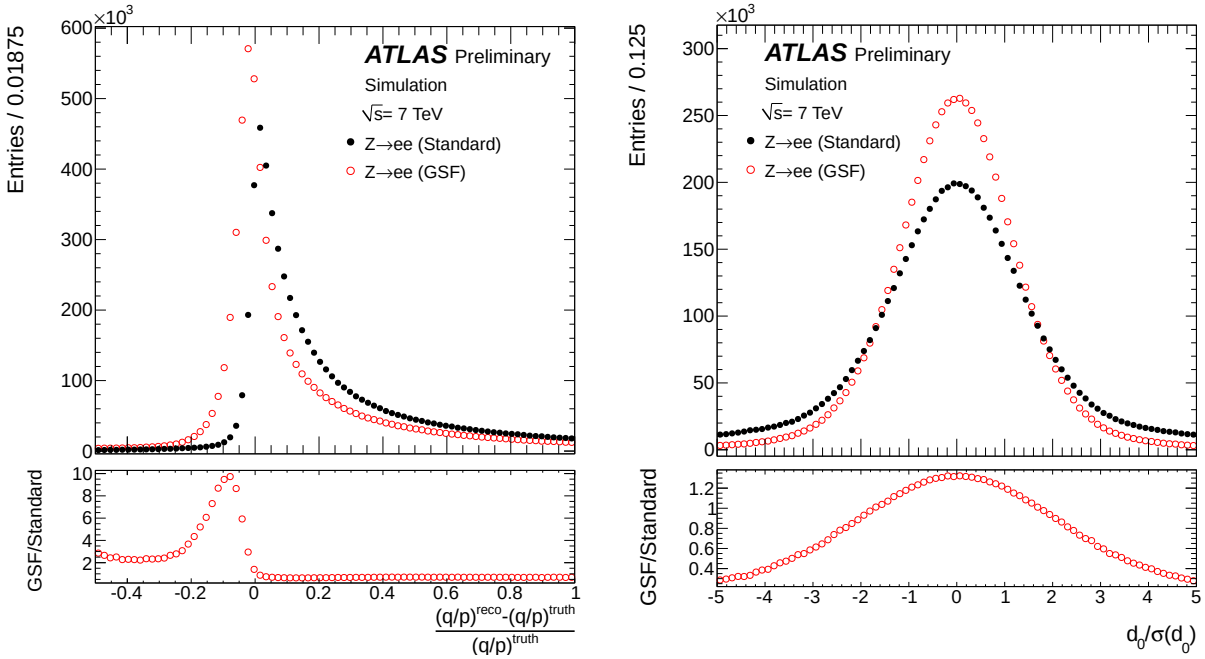


Figure 4.4: Impact of the GSF algorithm on the electron track resolution and impact parameter.

4.2.2 Electron Performance

Figure 4.5 shows the electron reconstruction efficiency versus p_T and η for both 7 TeV and 8 TeV data. Figure 4.6 shows the identification efficiency versus the number of primary vertices for the three cutbased quality criteria for 8 TeV data. This was measured using the tag-and-probe method on $Z \to e^+e^-$ events in data [21].

4.3 Muons

Muons are reconstructed using information in the muon spectrometer, inner detector, and calorimeter. Stand Alone (SA) muon trajectories use hits in the three Muon Spectrometer (MS) stations to seed the reconstruction algorithm. The trajectory is extrapolated back to the beam line to obtain the final angular and impact parameters. The muon's energy is further corrected based on parametrized calorimeter energy loss. SA tracks can be combined

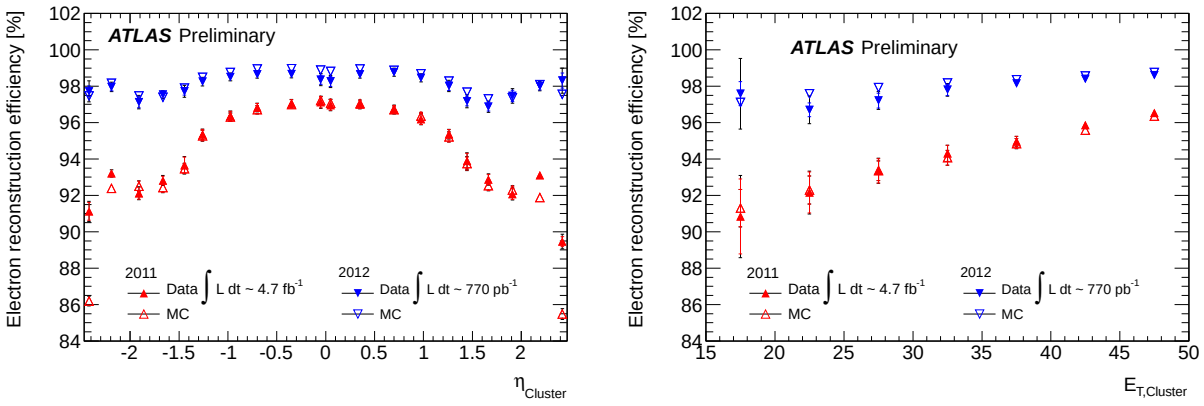


Figure 4.5: Electron reconstruction efficiency as a function of η and cluster E_T in data and MC for both 7 and 8 TeV data.

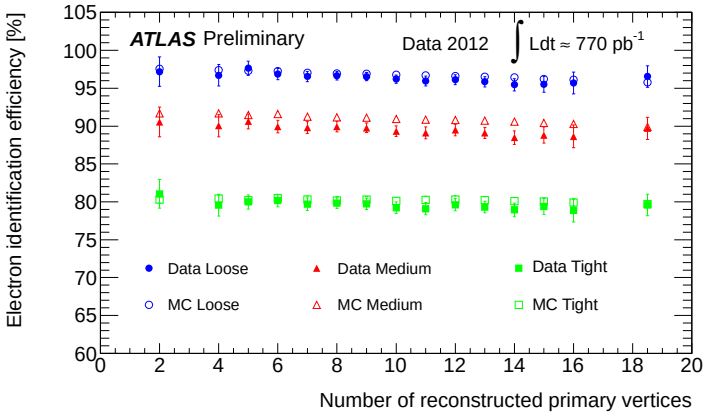


Figure 4.6: Electron identification efficiency as a function of the number of primary vertices for the three loose, medium, and tight selection criteria.

with reconstructed ID tracks and refit to give a final combined muon. The STACO (STAtistical COmbination) algorithm uses a χ^2_{match} as a figure of merit to determine the best pairings between SA and ID tracks, where $\chi^2_{\text{match}} = (\mathbf{T_{MS}} - \mathbf{T_{ID}})^{\mathbf{T}}(\mathbf{C_{MS}} - \mathbf{C_{ID}})^{-\mathbf{1}}(\mathbf{T_{MS}} - \mathbf{T_{ID}})$. $\mathbf{T}$ is the vector of five track parameters and $\mathbf{C}$ is its covariance matrix. The final refit track vector is then computed as $\mathbf{T} = (\mathbf{C_{ID}^{-1}} + \mathbf{C_{MS}^{-1}})^{-\mathbf{1}}(\mathbf{C_{ID}^{-1}T_{ID}} + \mathbf{C_{MS}^{-1}T_{MS}})$.

Two additional reconstruction algorithms are available to further increase the muon efficiency. One algorithm identifies segmented tagged muons by extrapolating ID tracks to the MS and identifying straight track segments (SA tracks missing hits in at least one MS station) consistent with the extrapolated hypothesis. A second algorithm extrapolates ID tracks to the calorimeter and searches for energy deposits consistent with a minimum ionizing particle. These algorithms are able to increase muon efficiency, but can potentially increase background rates due to the less stringent requirements on the MS track.

4.3.1 Muon Performance

Figure 4.7 shows the muon reconstruction efficiency versus η for combined muons based on tag and probe studies on early 7 TeV data. As can be seen in the figure, the efficiency loss in the $|\eta| \approx 0$ and $|\eta| \approx 1.4$ regions is high, which is the motivation for the additional reconstruction algorithms outlined at the end of the previous section.

4.4 Jets and Missing Transverse Energy

There are multiple jet reconstruction algorithms used in ATLAS in order to accommodate the needs of different physics analyses. Certain properties can be attributed to the algorithms that describe their stability from a theoretical standpoint. The property that the reconstructed jet multiplicity of an event is insensitive to additional soft particles not coming from the fragmentation of a hard scattered parton is called infrared safety. Collinear safety describes an algorithm that is insensitive to how the transverse momentum is distributed among collinear decay products; that is, jet reconstruction should not depend on the decay

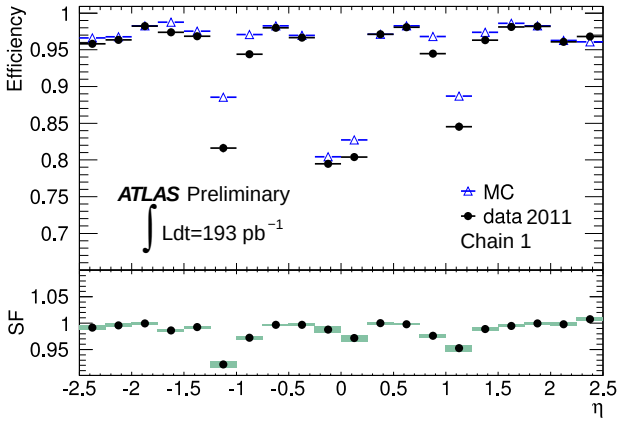


Figure 4.7: Combined muon reconstruction efficiency with respect to the inner tracking efficiency as a function of the pseudorapidity of the muon for muons with $p_T > 20$ GeV. The panel at the bottom shows the ratio between the measured and predicted efficiencies.

of a particle into two additional collinear particles. The jet algorithm must also be computationally practical, such that it scales well with event particle multiplicity. Before describing the algorithms, the procedure for reconstructing individual clusters within the calorimeter will first be mentioned.

The ATLAS calorimeters are comprised of about 200,000 cells in which energy can be deposited. Individual cell energies in an event are lumped together to yield reconstructed particle level objects. There are two methods for doing so. The first, topological cell clusters, is seeded by cells with a signal significance $\Gamma = E_{\text{cell}}/\sigma_{\text{noise,cell}}$ above some threshold S. Adjacent cells in all three dimensions are added indiscriminately of their Γ. Cells neighboring these are then added if their signal significance is larger than a secondary threshold N. A ring of guard cells with significance $\Gamma > P$ is finally added. Typical values of S, N, and P are 4, 2, and 0, respectively. After the cell has been reconstructed, local maxima of energy within the cell is searched for. If multiple maxima are present, the cluster is split using a splitting algorithm.

An alternative to topological cell based clusters are signal tower based reconstruction. Calorimeter tower signals are reconstructed using 2-d towers with a grid size of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. Energy in cells in the radial direction are summed indiscriminately into towers. Similar to topological clusters, signal towers are grouped into clusters based on a 4-2-0

approach. Starting from a seed tower with large signal significance, neighboring towers with a secondary significance threshold are added, followed by neighboring neighbors with a tertiary threshold. Being a two dimensional cluster, signal tower clusters are simpler but have no defined angular direction without additional vertex information.

4.4.1 Jets

Many localized streams of particles result from the fragmentation of hadronized partons produced within an event. The tracks and calorimeter deposits left by these particles can be reconstructed as jets. There are two common jet reconstruction algorithms in ATLAS. Fixed cone algorithms directly sum the four vectors of particles within a cone of some ΔR, typically 0.4 or 0.6, into a jet object as follows. Reconstructed calorimeter clusters are sorted by p_T. The largest of these, if above the seed threshold, is taken as the jet center. All objects within a cone of the specified ΔR are added to the jet. The jet center is recomputed using all objects inside the cone, and the summation is repeated with this new jet center. This is repeated until the jet center does not change. The final set of particles within the cone is removed from the initial sorted list, and the procedure is repeated for the remaining jets until no more seeds above threshold are available. This algorithm is not infrared safe, as soft radiation between two jets can potentially cause them to merge and change the jet multiplicity.

An alternative to the fixed cone algorithm is a k_T algorithm. k_T algorithms iteratively sum individual particles into a jet object based on a distance metric between each particle until some condition is met. For each pair i, j of calorimeter objects, the metric $d_{ij} = \min(p_{T,i}^2, p_{T,j}^2)\frac{\Delta R_{ij}}{R}$ is computed, as well as $d_i = p_{T,i}^2$ for each single object. If the minimum of all d_{ij} and d_i is from a pair, the i, j pair is merged into a single object. If the minimum is a single object, that object is termed a jet and removed from the list of objects. This process is repeated until no objects are left. The parameter R is typically chosen to be 0.4 or 0.6, depending on the needs of the physics analysis.

Similar to the k_T algorithm is an anti-k_T algorithm, in which the metrics are defined to be $d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2})\frac{\Delta R_{ij}}{R}$ and $d_i = p_{T,i}^{-2}$. For the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis, for example, the anti-k_T algorithm with $R = 0.4$ is used. The k_T and anti-k_T algorithms are both infrared and collinear safe.

4.4.2 Missing Energy

Following the conservation of momentum, the magnitude of the vector sum of the transverse momentum in an event must be exactly zero. In the case of massless particles, this is equivalent to the transverse energy, though the latter is more commonly used. The presence of non-interacting particles such as neutrinos or, potentially, dark matter in an event can therefore be indirectly detected by the presence of a non-zero transverse energy vector. The magnitude of the negative of this vector is called the missing energy of an event, or $\not{E}_T \equiv |\sum_i \vec{p}_{T,i}|$. There are two primary methods for reconstructing $\not{E}_T$ within ATLAS.

4.4.2.1 Cell Based $\not{E}_T$

Cell-based $\not{E}_T$ reconstruction uses calorimeter cell energies with noise suppression requirements. Cell energies are calibrated with global calibration weights. As muons escape the calorimeter with minimal energy loss, corrections from muon energy are required. A non-negligible component of energy is lost in the cryostat as it is about half an interaction length deep and must be taken into account. The x, y missing energy components can be written as the sum of these three terms.

$$\not{E}_{x,y}^{\text{Final}} = \not{E}_{x,y}^{\text{Calo}} + \not{E}_{x,y}^{\text{Muon}} + \not{E}_{x,y}^{\text{Cryo}} \tag{4.2}$$

The $\not{E}_{x,y}^{\text{Calo}}$ term is the negative sum of transverse energies in reconstructed topological cluster cells. These terms require an additional calibration based on the object associated with the cluster, with priority given in the order of electrons, photons, muons, hadronic taus,

b-jets, and light jets. This results in a more precise $\not\!\!E_T$ given the more relevant object level calibrations. Cells not associated with a reconstructed high-p_T object default to the global calibration scheme.

The $\not\!\!E_{x,y}^{\text{Muon}}$ term is derived from the p_T of stand alone muon tracks reconstructed in the muon spectrometer: $\not\!\!E_{x,y}^{\text{Muon}} = -\sum_{\text{RecMuons}} E_{x,y}$. Muons entering the calculation are required to have an inner detector track match if they're within the ID coverage of $|\eta| < 2.5$ in order to reduce the contribution from fake muons, as these can induce substantial missing energy up to the p_T of the fake muon.

The $\not\!\!E_{x,y}^{\text{Cryo}}$ is determined indirectly through the correlation of energy between the last layer of the LAr calorimeter (E_{EM3}) and the first layer of the hadronic calorimeter (E_{HAD}): $\not\!\!E_{x,y}^{\text{Cryo}} = -\sum_{\text{recJets}} w^{\text{Cryo}} \sqrt{E_{EM3} \times E_{HAD}}$, where w^{Cryo} is a calibration weight. A similar procedure is used for the end-cap cryostats.

4.4.2.2 Object Based $\not\!\!E_T$

An object based missing energy calculation can be used to reduce the impact of low-p_T objects, underlying event, and pileup. The $\not\!\!E_T$ can be considered a sum of high-p_T and low-p_T terms. The high-p_T term derives from the energy of fully reconstructed electrons, photons, muons, and taus, each at their respective calibrations. Hadronic jets above some threshold not associated with any of these objects are then added to this. Cells that are not associated with any high-p_T reconstructed objects are finally added to obtain the low-p_T term.

4.4.2.3 Missing Energy Performance

Figure 4.8 shows the missing energy resolution versus the number of primary vertices for $Z \to \mu\mu$ events in 8 TeV data.

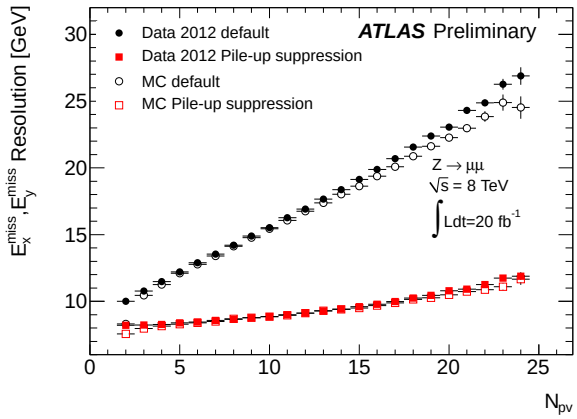


Figure 4.8: Resolution of x and y missing ET components as a function of the number of primary vertices for data and MC in $Z \to \mu\mu$ candidates.

4.4.3 Flavor Tagging

The identification of jets originating from b-quarks is important for the efficient selection of heavy flavor processes such as $t\bar{t}$, as well as for the rejection of such processes as backgrounds to other many analyses. Due to the relatively long lifetime of B-hadrons, jets originating from the fragmentation of b-quarks have a characteristic displaced secondary vertex as well as tertiary vertices from the subsequent c-parton fragmentation. As a result, the tracks within these jets will tend to have impact parameters inconsistent with the event's primary vertex (PV), where the impact parameter is defined to be the point of closest approach of the reconstructed track to the PV. The efficient reconstruction and precision identification of tracks and vertices is therefore essential for tagging b-jets. Several algorithms are used within ATLAS to perform this task [2].

Tracks used as inputs to the b-tagging algorithms are required to pass several quality selections in order to reject tracks originating from other long lived particles such as K_S and Λ. Tracks are required to have $p_T > 1$ GeV. At least two hits in the pixel are required, with a minimum of one being in the innermost layer. A cut of $|d_0| < 1$ mm is made on the transverse impact parameter, and $|z_0| \sin\theta < 1.5$ mm on the longitudinal impact parameter. SV1, one of the secondary vertex based algorithms, uses a looser set of selections: $p_T > 400$

MeV, $|d_0| < 3.5$ mm, and only one hit in the pixel detector. Further, no more than one track hit can be shared among any two tracks.

There are two algorithms based solely on the impact parameters of tracks. The JetProb algorithm performs a simple quadrature sum of the impact parameter significance of all candidate tracks in the jet. The more advanced IP3D algorithm uses a likelihood ratio (LR) technique using 2D smoothed histograms of d_0/σ_{d_0} and z_0/σ_{z_0} taken from Monte Carlo (MC) simulation for b- and light-jets.

Two tagging algorithms based on secondary vertices in the event are SV0 and SV1. Secondary vertices are identified starting with the set of all two-track pairs associated to the jet and far from the PV. Track pair vertices consistent with the PV or material interactions are removed. The remaining pairs are formed into an inclusive vertex. The track with the poorest χ^2 is iteratively removed until the χ^2 of the inclusive vertex is above some threshold. SV0 uses the signed 3D decay length significance $L_{3D}/\sigma_{L_{3D}}$ of the vertex as a discriminator. SV1 is LR based, using 1D histograms from $L_{3D}/\sigma_{L_{3D}}$, the number of two track vertices, and the ΔR between the jet axis and the line between the primary and secondary vertices. 2D histograms are also used based on the invariant mass of all tracks associated with the vertex and the ratio of the sum of the energies of the tracks in the vertex to the sum of the energies of all tracks in the jet.

JetFitter is an algorithm based on the decay chain inside the jet. This applies a Kahlman filter to find a common line of flight between the b- and c-quark decay vertices and the PV. A LR approach is then taken using variables similar to the SV1 algorithm, using the additional information from the b- and c-quark decay vertices.

Combinations of the above variables can be made to improve performance. In particular, the IP3D and SV1 discriminating variables can be directly added since they are both LR based. An IP3D and JetFitter combination can also be formed using an artificial neural network.

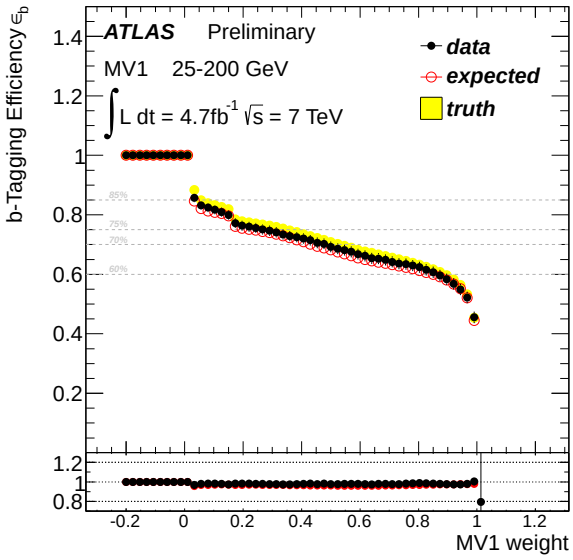


Figure 4.9: The efficiency of the MV1 b-tagging algorithm. The expected distribution is obtained by performing the kinematic fit method on simulated events while the truth distribution corresponds to that for simulated b-jets.

4.4.3.1 Flavor Tagging Performance

Figure 4.9 shows the efficiency of the MV1 tagger. MV1 is the output of a neural network that uses JetFitter+IP3D, IP3D, and SV1 as inputs.

CHAPTER V

Statistics

"It is somewhat like having four girlfriends at once ... There are certain advantages, because some of them will do things that others won't, but it is really really important to remember which one you're dealing with at any one time."

- Roger Barlow, on the definitions of probability

This chapter seeks to give a comprehensive overview of the frequentist statistics methods used at the LHC, and of particular modeling strategies incorporated within ATLAS.

5.1 Formalism

5.1.1 Likelihood Function

At the heart of any statistical analysis is the likelihood function. The likelihood function should condense all details of an analysis into a single equation, from which the minimal set of information required to describe that analysis can be sufficiently reconstructed. A simple example is the single bin number counting experiment with expected signal S, background B, and observed events N. A single parameter of interest μ normalizes S such that $\mu = 0$ corresponds to the null hypothesis and $\mu = 1$ corresponds to the nominal signal hypothesis. S, B, and N are simple constants, so the likelihood is a function of only μ. In this case the

likelihood is just the Poisson probability of the observed and expected events:

$$\mathcal{L}(\mu) = P(N|\mu S + B) \tag{5.1}$$

An extension of this is to normalize B from a Control Region (CR) of pure background with observed events N_{CR} and expected background B_{CR} using a normalization factor θ:

$$\mathcal{L}(\mu, \theta) = P(N|\mu S + \theta B)P(N_{CR}|\theta B_{CR}) \tag{5.2}$$

This is a minimal form of a likelihood in which all moving parts are present: The first Poisson is the part of the likelihood is the signal-rich component, in which all of the interesting information about μ is contained; μ is a parameter of interest (POI), θ is a nuisance parameter (NP), and N is the observed data. The second Poisson is the auxiliary constraint that constrains θ, and N_{CR} is an auxiliary measurement, commonly referred to as a global observable and alternatively denoted as $\tilde{\theta}$.

If the true values of the parameters of the likelihood (POI and NP) are unknown, common practice is to estimate them by finding the value that maximizes the likelihood. Explicity, the Maximum Likelihood Estimator (MLE) of θ is denoted as $\hat{\theta}_\mu$ and is a function of μ. Likewise, the true value of μ can also be estimated with $\hat{\mu}$. The profiled likelihood $\mathcal{L}(\mu, \hat{\theta}_\mu)$ will be of significant importance when constructing test statistics. For large models, numerical techniques must be used to find $\hat{\theta}$ and $\hat{\mu}$. In the minimal likelihood, the unconditional $\hat{\mu}$ and $\hat{\theta}$ can be solved analytically:

$$\begin{aligned} \hat{\theta} &= \frac{N_{CR}}{B_{CR}} \\ \hat{\mu} &= (N - N_{CR}\frac{B}{B_{CR}})/S \end{aligned} \tag{5.3}$$

The conditional solution for $\hat{\theta}_\mu$, where μ is held at some constant value, can also be found,

leading to a quadratic equation:

$$\hat{\theta}_\mu = \frac{B(N_{CR}+N)-\mu S(B+B_{CR})}{2B(B+B_{CR})} \times \left\{ 1 + \sqrt{1 - \frac{4\mu S N_{CR} B(B+B_{CR})}{(B(N_{CR}+N)-\mu S(B+B_{CR}))^2}} \right\} \quad (5.4)$$

This simple model will be used frequently throughout this chapter. A more general form of the likelihood can be written:

$$\mathcal{L}(\mu,\theta) = \mathcal{L}_{data}(N|\vec{\mu},\theta) \times \mathcal{A}(\tilde{\theta}|\theta) \quad (5.5)$$

$\mathcal{L}_{data}(N|\mu,\theta)$ represents the portion of the likelihood particular to your local data $N = \{\mathbf{x}_1,...,\mathbf{x}_q\}$ with corresponding weights $w = \{w_1,...,w_q\}$, which, in general, can represent a set of q data points on a multi-dimensional observable $\mathbf{x}$ each with some weight w_i. This is elaborated upon in Section 5.1.1.1. The n parameters of interest $\mu = \{\mu_1,...,\mu_n\}$ are the parameters one wishes to make a statement about. The m nuisance parameters $\theta = \{\theta_1,...,\theta_m\}$ are free variables that represent other unknown quantities. $\mathcal{A}(\tilde{\theta}|\theta) = \prod_i^p \mathcal{A}_i(\tilde{\theta}_i|\theta_i)$ represents an independent auxiliary likelihood that condenses measurements about the p global observables $\tilde{\theta} = \{\tilde{\theta}_1,...,\tilde{\theta}_p\}$ into smaller forms. As long as $\mathcal{A}(\tilde{\theta}|\theta)$ is statistically independent from N and is unaffected by μ this is a valid approximation. An individual $\mathcal{A}_i(\tilde{\theta}_i|\theta_i)$ could represent, for example, the measurement of the jet energy scale (JES) $\tilde{\theta}_{JES}$.

5.1.1.1 Extended Likelihoods

The likelihood formalism can be extended to parametric models where the expected number of events is a function of not only the parameters $\alpha = \{\mu,\theta\}$ but some observables $\mathbf{x}$. $\mathbf{x}$ might be, for example, some invariant mass, angular quantity, time of flight, etc... . The differential distribution of expected events can be defined as $f(\mathbf{x},\alpha)$ such that the total expected is $E(\alpha) = \int f(\mathbf{x},\alpha)d\mathbf{x}$. For each observed data point $\mathbf{x}_i$ with weight w_i, the differential probability of observing such event is $(f(\mathbf{x}_i,\alpha)/E(\alpha))^{w_i}$. The likelihood is then the product over the differential probability for the q observed events, along with the Poisson

probability for the total number of events $N = \sum_i^q w_i$ and any auxiliary constraints $\mathcal{A}(\tilde{\theta}|\theta)$ on α:

$$\mathcal{L}(\alpha) = P(N|E(\alpha)) \times \mathcal{A}(\tilde{\theta}|\theta) \times \prod_i^q \left(\frac{f(\mathbf{x}_i, \alpha)}{E(\alpha)} \right)^{w_i} \tag{5.6}$$

A common technique is to use binned histograms templates for expected signal and background rates for some observable distribution. This can be conceptualized as an extended likelihood with N_{bins} discontinous PDFs, each containing N_i observed events:

$$\mathcal{L}_{\text{binned}}(\alpha) = P(N|E(\alpha)) \times \mathcal{A}(\tilde{\alpha}|\alpha) \times \prod_i^{N_{bins}} \left(\frac{f(\mathbf{x}_i, \alpha)}{E(\alpha)} \right)^{N_i} \tag{5.7}$$

This is mathematically equivalent to a simple product over Poisson probabilities up to a multiplicitive constant:

$$\begin{aligned} P(N|E) \times \prod_i^{N_{bins}} \left(\frac{f(\mathbf{x}_i)}{E} \right)^{N_i} &= \left\{ \tfrac{E^N e^{-E}}{N!} \right\} \times \left\{ \prod_i^{N_{bins}} \left(\frac{E_i}{E} \right)^{N_i} \right\} \\ &= \left\{ \tfrac{E^N}{N!} \prod_i^{N_{bins}} e^{-E_i} \right\} \times \left\{ \tfrac{1}{E^N} \prod_i^{N_{bins}} E_i^{N_i} \right\} \\ &= \tfrac{1}{N!} \times \prod_i^{N_{bins}} E_i^{N_i} e^{-E_i} \\ &= \tfrac{1}{N!} \times \prod_i^{N_{bins}} N_i! \times \prod_i^{N_{bins}} \frac{E_i^{N_i} e^{-E_i}}{N_i!} \\ &= const \times \prod_i^{N_{bins}} P(N_i|E_i) \end{aligned} \tag{5.8}$$

5.2 Statistical Methods

5.2.1 Test Statistics

As the value of a likelihood function is arbitrary up to a multiplicitive constant, it holds little meaning on its own (not to be confused with the meaning held in the interpretation of its functional form). It is typical to bring meaning to its value by comparing it at two different points in parameter space. In particular, a test statistic $Q = -2 \ln \frac{\mathcal{L}(N|\alpha_1)}{\mathcal{L}(N|\alpha_2)}$ holds

meaning as a metric to distinguish between two different hypotheses α_1 and α_2 given the same set of data N. The hypothesis in the numerator (α_1) is usually refered to as the null, and the denominator (α_2) the alternate. Although any function of the data can technically be used as a test statistic, the use of the likelihood ratio is motivated by the Neyman-Pearson lemma [60], which states that in the absence of unknown parameters, the likelihood ratio has the highest statistical power. Power in this case is defined as the probability to reject the null hypothesis when the null hypothesis is false.

When considered in relation to its corresponding conditional ensemble of all possible values, the value of the test statistic can be mapped to a probability, commonly refered to as a p-value, that quantifies the consistency of the two hypotheses with the data. The definition of consistency depends on the conditional parameters under which the ensemble is generated and the integrated range of the sampling distribution (see section 5.2.3).

5.2.2 Test Statistics used at the LHC

Three likelihood ratios are commonly used at the LHC. The most common is the Profile Likelihood Ratio (PLR) $\lambda(\mu)$, and the alternate PLR $\tilde{\lambda}(\mu)$:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \tag{5.9}$$

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{\mathcal{L}(\mu,\hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu},\hat{\theta})}, & \hat{\mu} > 0 \\ \frac{\mathcal{L}(\mu,\hat{\theta}_\mu)}{\mathcal{L}(0,\hat{\theta}(0))}, & \hat{\mu} \leq 0 \end{cases} \tag{5.10}$$

The third is the Ratio of Profiled Likelihoods (RPL) λ_{RPL}, which can in fact be written as a function of the PLR:

$$\begin{aligned} \lambda_{\mathrm{RPL}} &= \frac{\mathcal{L}(1,\hat{\theta}(1))}{\mathcal{L}(0,\hat{\theta}(0))} \\ &= \frac{\mathcal{L}(1,\hat{\theta}(1))}{\mathcal{L}(\hat{\mu},\hat{\theta})} \frac{\mathcal{L}(\hat{\mu},\hat{\theta})}{\mathcal{L}(0,\hat{\theta}(0))} \\ &= \frac{\lambda(1)}{\lambda(0)} \end{aligned} \tag{5.11}$$

The standard test statistic used at the LHC is the Profile Likelihood Test Statistic

(PLTS), which is constructed from the PLR. The PLTS has several variations depending on the hypothesis being tested and the allowed range of the parameter of interest.

The test statistic t_μ uses the PLR $\lambda(\mu)$ and has no restrictions on μ. This test statistic is generally used to quantify compatibility of the data with the μ hypothesis:

$$t_\mu = -2 \ln \lambda(\mu) \tag{5.12}$$

The alternate test statistic $\tilde{t}_\mu$ is similar to t_μ, but uses the the alternate PLR $\tilde{\lambda}(\mu)$. This is used when one is interested in restricting the range of $\hat{\mu}$ to be physical:

$$\tilde{t}_\mu = -2 \ln \tilde{\lambda}(\mu) \tag{5.13}$$

The discovery test statistic q_0 is used to quantify the rejection of the null hypothesis in the case of an excess in the data:

$$q_0 = \begin{cases} -2 \ln \lambda(0), & \hat{\mu} > 0 \\ 0, & \hat{\mu} \leq 0 \end{cases} \tag{5.14}$$

When one is interested in setting an upper limit on μ, the test statistic q_μ can be used. This test statistic does not penalize cases where $\hat{\mu}$ is larger than the tested μ. That is, an excess of signal-like events should not be considered evidence against the signal:

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu), & \hat{\mu} < \mu \\ 0, & \hat{\mu} \geq \mu \end{cases} \tag{5.15}$$

Again, $\tilde{q}_\mu$ can be written for cases in which one wishes to restrict $\hat{\mu}$ to be physical:

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu), & \hat{\mu} < \mu \\ 0, & \hat{\mu} \geq \mu \end{cases} \tag{5.16}$$

λ_{RPL} can be used to form the final test statistic considered here, which quantifies compatibility between two fixed hypotheses:

$$\begin{aligned} q_{RPL} &= -2\ln\lambda_{RPL} \\ &= -2\ln\lambda(1) + 2\ln\lambda(0) \end{aligned} \tag{5.17}$$

5.2.2.1 Uncapped Test Statistics

One downside of the q_0, q_μ, and $\tilde{q}_\mu$ test statistics just shown, is that information is lost when the conditions that lead to $q = 0$ are met. This problem can be alleviated by 'uncapping' the test statistic, allowing it to take on negative values when these conditions are met:

$$q_0 \to r_0 = \begin{cases} -2\ln\lambda(0), & \hat{\mu} > 0 \\ +2\ln\lambda(0), & \hat{\mu} \leq 0 \end{cases} \tag{5.18}$$

$$q_\mu \to r_\mu = \begin{cases} -2\ln\lambda(\mu), & \hat{\mu} < \mu \\ +2\ln\lambda(\mu), & \hat{\mu} \geq \mu \end{cases} \tag{5.19}$$

$$\tilde{q}_\mu \to \tilde{r}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu), & \hat{\mu} < \mu \\ +2\ln\tilde{\lambda}(\mu), & \hat{\mu} \geq \mu \end{cases} \tag{5.20}$$

Note that both $\lambda(\mu)$ and $\tilde{\lambda}(\mu)$ are only able to take on values between 0 and 1, such that $-2\ln\lambda(\mu)$ and $-2\ln\tilde{\lambda}(\mu)$ are positive definite and $+2\ln\lambda(\mu)$ and $+2\ln\tilde{\lambda}(\mu)$ are negative definite.

As will be seen in the following section, this is equivalent to populating the negative side of the ensemble of q, while before this information was lost in the delta function at $q = 0$.

5.2.3 Procedure for Computing P-Values

This section outlines the frequentist procedures used at the LHC for generating pseudo-experiments with the purpose of computing p-values and calibrating confidence intervals.

Starting with a test statistic Q, we are interested in obtaining the distribution of Q, namely $f(Q|\alpha')$ (read: probability distribution of Q assuming the α' hypothesis is true) under repeated experiments. It is assumed that Q is constructed from a likelihood $\mathcal{L}(N|\mu,\theta)$ with n parameters of interest $\mu = \{\mu_1, ..., \mu_n\}$, nuisance parameters θ, and auxiliary measurements $\tilde{\theta}$ with associated auxiliary constraints $\mathcal{A}(\tilde{\theta}|\theta)$. $\alpha' = \{\mu', \theta'\}$ is the set of hypothesized parameters used to generate the ensemble. N represents the dataset on which $\mathcal{L}$ is computed. For each pseudo-experiment, pseudo-data N_{pseudo} is generated by sampling from its associated PDF within $\mathcal{L}$. Similarly, pseudo-auxiliary-measurements $\tilde{\theta}_{\text{pseudo}}$ are generated by sampling from its associated auxiliary constraint $\mathcal{A}$.

In this way, the meaning of the auxiliary measurements and constraints is made clear. All information of some more detailed independent likelihood $\mathcal{L}_{\mathcal{A}} \sim \mathcal{A}$ used to extract the nominal value of a parameter, which is now the auxiliary measurement used in $\mathcal{L}$, is condensed into the auxiliary constraint $\mathcal{A}$. Randomizing the auxiliary measurements according to its auxiliary constraint should be equivalent to including the full likelihood $\mathcal{L}_{\mathcal{A}}(N_{\mathcal{A}}|\theta)$ in place of $\mathcal{A}$, generating $N_{\mathcal{A}}$ through this likelihood, and using this in subsequent estimates of the corresponding nuisance parameter θ.

A subtle point is that a value of the nuisance parameters must be chosen around which to randomize the pseudo-data and the auxiliary measurements. For parameters that have no a auxiliary constraint, such as the normalization of a background from a control region, an obvious choice is to profile the observed data to extract an estimate of the parameter. The general recommendation for LHC physics is to profile all nuisance parameters on the observed data while fixing the parameters of interest to their hypothesized value μ'. In this way one can write the distribution of Q as $f(Q|\mu', \hat{\theta}(\mu'))$.

The probability for obtaining a value of Q greater than the observed value Q_{obs}, the p-value, is obtained from integrating $f(Q|\alpha')$ upwards: $p = \int_{Q_{obs}}^{\infty} f(Q|\alpha')dQ$. In some cases, the p-value is defined to be the integral of the complementary side: $p' = 1 - p = \int_{-\infty}^{Q_{obs}} f(Q|\alpha')dQ$. The p-value, p, is a frequentist statement about the conditional probability of Q under

repeated measurements assuming the α' hypothesis is true.

For quantifying an excess, the test statistic r_0 can be used to obtain the p-value p_0. p_0 is the probability of obtaining a value of r_0 more discrepant than the observed $r_{0,obs}$ under the assumption that the null hypothesis $\mu' = 0$ is true. The median expected p-value $p_{0,exp}$ can be obtained from the same distribution but integrating from the median expected value of r_0, namely $r_{0,exp}$. $r_{0,exp}$ is obtained from the median of the distribution of r_0 under the nominal alt hypothesis $\mu' = 1$. This is shown more clearly in Figure 5.1:

$$\begin{aligned} p_0 &= \int_{r_{0,obs}}^{\infty} f(r_0|0,\hat{\theta}(0))dr_0 \\ p_{0,exp} &= \int_{r_{0,exp}}^{\infty} f(r_0|0,\hat{\theta}(0))dr_0 \\ r_{0,exp} &= \text{med}\{\text{r}_0|\mu' = 1,\hat{\theta}(1)\} \end{aligned} \tag{5.21}$$

For upper limits, the test statistics r_μ and $\tilde{r}_\mu$ can be used. The procedure for each is the same, but $\tilde{r}_\mu$ is used more frequently, so this will be used in all notations. For an observed $\tilde{r}_{\mu,obs}$, the p-value p_μ is the probability of rejecting the μ hypothesis assuming μ is true if $\tilde{r}_\mu > \tilde{r}_{\mu,obs}$ is the criteria for rejection. The expected under the null is defined similar to the expected p_0:

$$\begin{aligned} p_\mu &= \int_{\tilde{r}_{\mu,obs}}^{\infty} f(\tilde{r}_\mu|\mu,\hat{\theta}_\mu)d\tilde{r}_\mu \\ p_{\mu,exp} &= \int_{\tilde{r}_{\mu,exp}}^{\infty} f(\tilde{r}_\mu|\mu,\hat{\theta}_\mu)d\tilde{r}_\mu \\ \tilde{r}_{\mu,exp} &= \text{med}\{\tilde{\text{r}}_\mu|\mu' = 0,\hat{\theta}(0)\} \end{aligned} \tag{5.22}$$

For the p-value p_b, $1 - p_b$ represents the power of the test, with larger values of $1 - p_b$ indicating a larger power. This is defined from the same distribution that $\tilde{r}_{\mu,exp}$ is obtained:

$$p_b = \int_{-\infty}^{\tilde{r}_{\mu,obs}} f(\tilde{r}_\mu|0,\hat{\theta}(0))d\tilde{r}_\mu \tag{5.23}$$

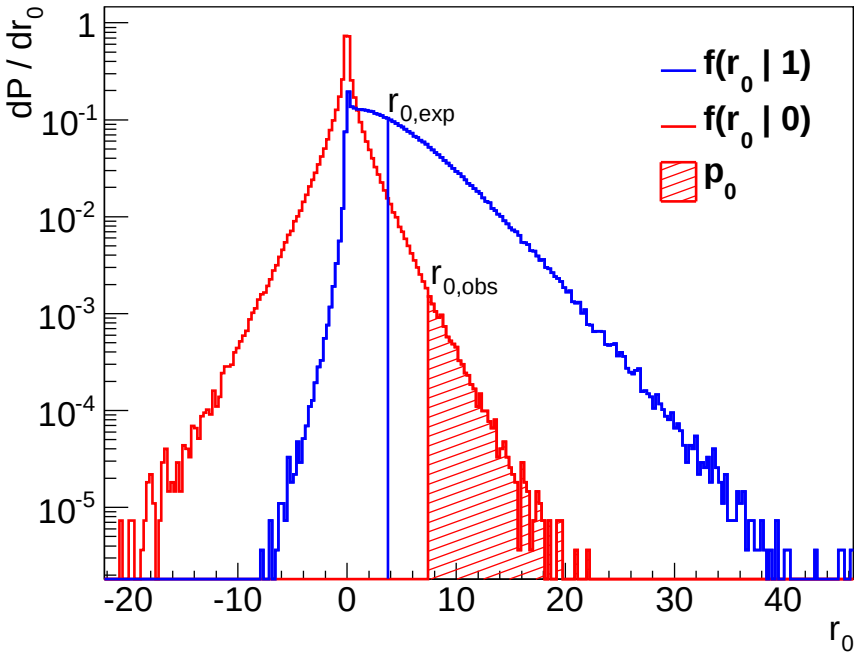


Figure 5.1: Distribution of the test statistic r_0 for both null ($\mu = 0$) and alternate ($\mu = 1$) hypotheses. The observed p-value p_0 is shown in the shaded area. The observed and expected value of the test statistic are also shown.

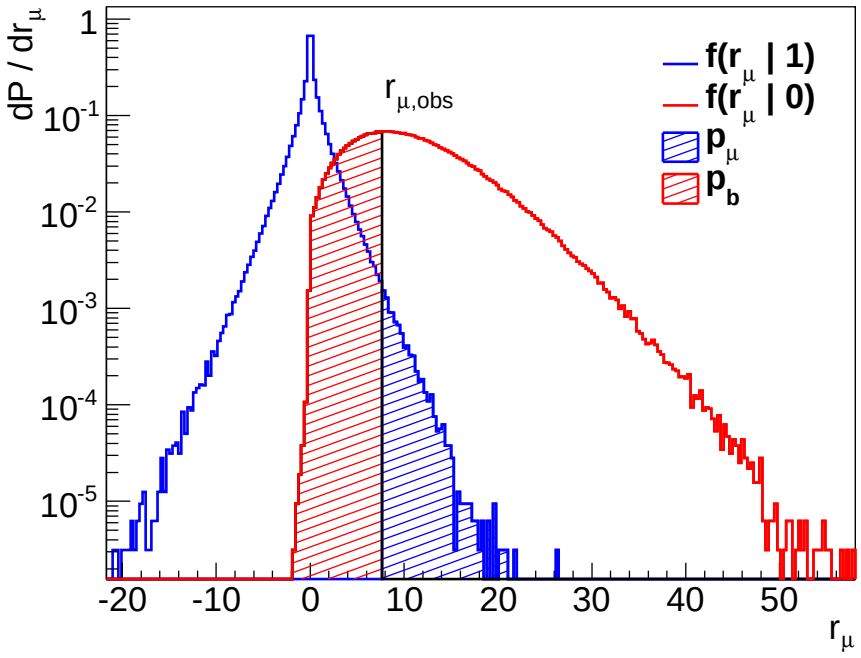


Figure 5.2: Distributions of the test statistic r_μ for both $\mu = 0$ and $\mu = 1$ hypotheses, along with the p-values p_b and p_μ.

The quantity CLs $= \frac{p_\mu}{1-p_b}$ is frequently used for upper limits in place of p_μ. The motivation is clear from the formula. In the case of strong downward fluctuations of the data, the p-value p_μ will tend to reject μ even if the test has no power to. By normalizing to the power, this feature is alleviated, but by no means removed. In the case of testing $\mu = 0$, CLs$\rightarrow$1, so that CLs can never reject the null, whereas p_μ would by definition reject the null with a probability α when $\alpha < p_\mu$ is the condition for rejection. The expected value of CLs under the null is exactly twice $p_{\mu,exp}$, given that the expected value of p_b is 0.5. For increasingly strong upward fluctuations of the data, $p_b \rightarrow 0$, such that CLs$\rightarrow p_\mu$.

Figure 5.2 shows graphical representations of the p-values p_μ and p_b, and how they are obtained from the sampling distribution of the test statistic r_μ.

5.3 Systematics

Uncertainties are included in the likelihood in two parts. The first is the auxiliary constraint on the nuisance parameter that represents the uncertainty. The second is the parametrization of how the terms in the data part of the likelihood respond to changes in the nuisance parameter. The expected rates are typically represented as the nominal rate multiplied by a response function:

$$E(\theta) = E_0 \nu(\theta) \tag{5.24}$$

Here E_0 is the nominal expected rate, and $\nu(\theta)$ is the response function due to the nuisance parameter θ. The exact form of the auxiliary constraint and response function depend on the source of the uncertainty. The typical cases used in Higgs analyses are outlined here.

5.3.1 Statistical Uncertainties

Statistical uncertainties follow a Poisson distribution and can therefore use a Poisson as an auxiliary constraint. These uncertainties could be due to MC statistical uncertainties, for example.

$$\begin{aligned} \mathcal{A}(\tilde{\theta}|\theta) &= P(\tilde{\theta}|\theta M) \\ \nu(\theta) &= \theta \end{aligned} \tag{5.25}$$

M is a constant, and is the nominal value of $\tilde{\theta}$. In the case of MC statistical uncertainties, M would be the effective number of MC events in the selected sample.

5.3.1.1 Barlow Beeston

In a binned likelihood, each bin can have an associated nuisance parameter θ representing the MC statistical uncertainty for a subset of the expectation in that bin. Because the likelihood can be factorized into a θ dependent and θ independent component, the MLE of

θ can be found analytically:

$$\mathcal{L} = \mathcal{L}_{other} \times P(N|\theta E + E_0)P(\tilde{\theta}|\theta M) \tag{5.26}$$

The likelihood $\mathcal{L}_{other}$ is independent of the particular θ we're interested in. The first Poisson represents the bin in the histogrammed distribution, while the second is the auxiliary constraint on the nuisance parameter θ. M is the effective number of MC events for the expectation E. E_0 some extra expectation in this bin that is not associated with M. Maximizing the likelihood is equivalent to minimizing the Negative Log Likelihood (NLL). Differentiating the NLL with respect to θ and evaluating at $\hat{\theta}$ gives a quadratic equation:

$$\begin{aligned} -\tfrac{d\ln L}{d\theta}|_{\hat{\theta}} = -\tfrac{EN}{E\hat{\theta}+E_0} + E - \tfrac{\tilde{\theta}}{\hat{\theta}} + M &= 0 \\ E(E+M)\hat{\theta}^2 + (E_0(E+M) - E\tilde{\theta})\hat{\theta} - EN - E_0\tilde{\theta} &= 0 \end{aligned} \tag{5.27}$$

$\hat{\theta}$ can then be solved for using the usual quadratic formula.

5.3.2 Normalization Systematics

Normalization systematics are systematics which strictly affect the normalization of a signal or background in a distribution. Sources of systematics can also have both normalization and shape components with each treated in a factorized way. The normalization components are treated with a unit Gaussian distributed nuisance parameter and an exponential response function:

$$\begin{aligned} \mathcal{A}(\tilde{\theta}|\theta) &= G(\tilde{\theta}|\theta, 1) \\ \nu(\theta) &= \kappa^{\theta} \end{aligned} \tag{5.28}$$

In this way, the sampling distribution of some expected number of events $E(\theta) = E_0\kappa^{\theta}$, namely $\frac{dP}{dE}$, is a log-normal centered at E_0, as shown in Figure 5.3. This follows from a

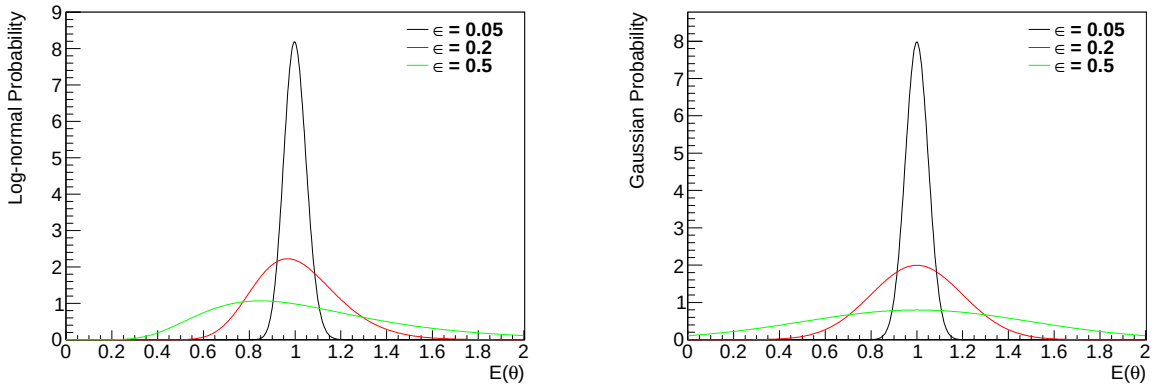


Figure 5.3: Prior probability versus the expected number of events $E(\theta)$ for both log-normal ($E(\theta) = E_0(1+\epsilon)^\theta$) and Gaussian ($E(\theta) = E_0(1+\epsilon\theta)$) treatments for various values of ϵ.

change of variables of the normally distributed θ:

$$\begin{aligned} \frac{dP}{dE} &= \frac{dP}{d\theta}\frac{d\theta}{dE} \\ &= \frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}\theta^2)\frac{1}{E\ln\kappa} \\ &= \frac{1}{\sqrt{2\pi}E\ln\kappa}\exp\left\{-\tfrac{1}{2}\left(\frac{\ln E - \ln E_0}{\ln\kappa}\right)^2\right\} \end{aligned} \tag{5.29}$$

κ is typically determined asymmetrically by measuring $E(1)$ and $E(-1)$. In the case that $E(1) = \frac{1}{E(-1)}$, the solution for κ is the same for the two boundary conditions. $E(1) \neq \frac{1}{E(-1)}$ represents an asymmetric uncertainty and can be dealt with in a bifurcated way. The asymmetric values of κ are $\kappa_+ = \frac{E(1)}{E_0}$ and $\kappa_- = \frac{E_0}{E(-1)}$, such that $E(\theta)$ is conditional up to the sign of θ:

$$E(\theta) = \begin{cases} E_0\kappa_+^\theta & \theta \geq 0 \\ E_0\kappa_-^\theta & \theta < 0 \end{cases} \tag{5.30}$$

This treatment is factorized for each systematic, such that the total expectation due to N normalization systematics is written as $E(\vec{\theta}) = E_0\prod_{i=1}^{N}\nu_i(\theta_i)$.

An alternative is to use expectations distributed as truncated Gaussians:

$$E(\theta) = \begin{cases} E_0(1+\delta_+\theta) & \theta \geq 0 \\ E_0(1+\delta_-\theta) & \theta < 0 \end{cases} \tag{5.31}$$

δ_+ and δ_- can be similarly determined by computing $E(1)$ and $E(-1)$. Log-normally distributed expectations have an advantage over truncated Gaussians in that they're positive definite over the full range of θ, while truncated Gaussians lead to negative expected events when $\theta < -\frac{1}{\delta}$, hence $E(\theta)$ must be truncated at zero in cases that θ satisfies this condition. Log-normals are the default sytematics in Higgs analyses due to this property.

5.3.3 Shape Systematics

Shape systematics must be treated differently for binned and unbinned distributions. Unbinned distributions need consideration on a case-by-case basis and are not covered in this section. For binned distributions, shape systematics are treated with a unit Gaussian distributed nuisance parameter and a piecewise linear response function. The binned distribution for a discrete observable x is computed for the nominal expectation as $E_0(x)$, and in generally separately for the two boundaries $\theta_j = \pm 1$, yeilding $E_j^{\pm}(x)$. For N shape systematics, the expected events in the bin corresponding to the observable value x_i can be written like so:

$$\begin{aligned} E(x_i, \vec{\theta}) &= \max(0, E_0(x_i) + \sum_{j=1}^{N} \delta E_j(x_i, \theta_j)), \\ \delta E_j(x_i, \theta_j) &= \begin{cases} (E_j^+(x_i) - E_0(x_i))\theta_j, & \theta_j \geq 0 \\ (E_j^-(x_i) - E_0(x_i))\theta_j, & \theta_j < 0 \end{cases} \end{aligned} \tag{5.32}$$

For each normally distributed nuisance parameter θ_j, the distribution $\frac{dP}{dE(x_i,\theta_j)}$ is a truncated Gaussian as shown in Figure 5.3, where the truncation is in place to keep the expected events positive definite. The motivation for piecewise linear (truncated Gaussian) rather than piecewise exponential (log-normal) interpolation is mostly historical. Earlier versions of RooFit could not elegantly handle the integration required to normalize the piecewise exponential interpolation scheme, and so the piecewise linear scheme was adopted. The two methods are equivalent to first order, so for small shape systematics the difference should also be small.

5.3.4 Nuisance Parameter Interpolation

If the response function $\nu(\theta)$ for a normally distributed nuisance parameter θ is asymmetric around $\theta = 0$, a kink in the likelihood and a discontinuity in its first derivative appear at $\theta = 0$. This can lead to numerical instabilities in the maximum likelihood fit and a delta function in the sampling distribution of $\hat{\theta}$ at $\hat{\theta} = 0$. A solution to this is to use a smoothing polynomial $\nu_0(\theta)$ for the interval $-\theta_b < \theta < \theta_b$ such that $\nu(\theta)$ and at its l^{th} derivative $\nu^{(l)}(\theta)$ are continuous at the boundaries $\theta = \pm\theta_b$ and at $\theta = 0$. Given that computing the boundaries $\nu(\pm 1)$ is common practice, θ_b is usually taken to be 1. The response function $\nu(\theta)$ is then divided into three regions.

$$\nu(\theta) = \begin{cases} \nu_+(\theta) & \theta \geq \theta_b \\ \nu_0(\theta) & -\theta_b < \theta < \theta_b \\ \nu_-(\theta) & \theta \leq -\theta_b \end{cases} \tag{5.33}$$

Continuity up to the $(n-1)^{th}$ derivative requires a polynomial of degree $2n$ (n is taken to be 1 greater than the derivative to simplify the algebra). This requirement stems from imposing the boundary conditions $\nu_\pm^{(\ell)}(\pm\theta_b) = \nu_0^{(\ell)}(\pm\theta_b)$ for $\ell = 0, .., n-1$. The resulting polynomial and its ℓ^{th} derivative can be written in the form of a series:

$$\begin{aligned} \nu_0(\theta) &= 1 + \sum_{p=1}^{2n} c_a \theta^a \\ \nu_0^{(\ell)}(\theta) &= \sum_{p=\ell}^{2n} c_p p(p-1)(...)(p-\ell+1)(\theta)^{p-\ell} \end{aligned} \tag{5.34}$$

The continuity requirements $\nu_\pm^{(\ell)}(\pm\theta_b) = \nu_0^{(\ell)}(\pm\theta_b)$ can be represented by a $(2n \times 2n)$

matrix equation:

$$\begin{pmatrix} \theta & \theta_b^2 & ... & \theta_b^{2n} \\ 1 & 2\theta_b & ... & 2n\theta_b^{2n-1} \\ & \vdots & & \\ ... & 0 & n! & ... & \frac{(2n+1)!}{(n+1)!}\theta_b^{n+1} \\ -\theta & (-\theta_b)^2 & ... & (-\theta_b)^{2n} \\ 1 & -2\theta_b & ... & 2n(-\theta_b)^{2n-1} \\ & \vdots & & \\ ... & 0 & n! & ... & \frac{(2n+1)!}{(n+1)!}(-\theta_b)^{n+1} \end{pmatrix} \times \begin{pmatrix} c_1 \\ \vdots \\ c_{2n} \end{pmatrix} = \begin{pmatrix} \nu_+(\theta_b) - 1 \\ \nu_+^{(1)}(\theta_b) \\ \vdots \\ \nu_+^{(n)}(\theta_b) \\ \nu_-(-\theta_b) - 1 \\ \nu_-^{(1)}(-\theta_b) \\ \vdots \\ \nu_-^{(n)}(-\theta_b) \end{pmatrix} \tag{5.35}$$

The $(\ell, p)^{th}$ element of the $(2n \times 2n)$ matrix can be condensed into a more convenient form:

$$A_p^\ell = \begin{cases} \frac{(p+1)!}{(p-\ell+1)!}(\theta_b)^{p-\ell+1} & p \geq \ell - 1, \quad \ell \leq n \\ 0 & p < \ell - 1, \quad \ell \leq n \\ \frac{(p+1)!}{(p-\ell+1)!}(\theta_b)^{p-\ell+1}(-1)^{p+1} & p + n \geq \ell - 1, \quad \ell > n \\ 0 & p + n < \ell - 1, \quad \ell > n \end{cases} \tag{5.36}$$

Inverting the matrix $\mathbf{A}$ and multiplying both sides of the matrix equation yields an equation for the coefficients c_p:

$$c_p = \sum_{\ell=0}^{n} (\mathbf{A}^{-1})_p^\ell \nu_+^{(\ell)} + (\mathbf{A}^{-1})_p^{\ell+n+1} \nu_-^{(\ell)} \tag{5.37}$$

The minimum polynomial order required to avoid large numerical instabilities is four, corresponding to continuity up to the first derivative. However, the more robust MINUIT algorithms require computing the second derivative also, so a sixth order polynomial has become the standard within Higgs analyses. The two most common forms of $\nu_\pm(\theta)$ are exponential, corresponding to the normalization systematics outlined in section 5.3.2, and linear, corresponding to Section 5.3.3. The coefficients of the sixth order solution for the

Coefficient	Linear	Exponential
c_1	$\frac{1}{2}(\epsilon^+ + \epsilon^-)$	$\frac{1}{8\theta_b}\left(15A_0 - 7\theta_b S_1 + \theta_b^2 A_2\right)$
c_2	$\frac{15}{16\theta_b}(\epsilon^+ - \epsilon^-)$	$\frac{1}{8\theta_b^2}\left(-24 + 24S_0 - 9\theta_b A_1 + \theta_b^2 S_2\right)$
c_3	0	$\frac{1}{4\theta_b^3}\left(-5A_0 + 5\theta_b S_1 - \theta_b^2 A_2\right)$
c_4	$-\frac{10}{16\theta_b^3}(\epsilon^+ - \epsilon^-)$	$\frac{1}{4\theta_b^4}\left(12 - 12S_0 + 7\theta_b A_1 - \theta_b^2 S_2\right)$
c_5	0	$\frac{1}{8\theta_b^5}\left(3A_0 - 3\theta_b S_1 + \theta_b^2 A_2\right)$
c_6	$-\frac{3}{16\theta_b^5}(\epsilon^+ - \epsilon^-)$	$\frac{1}{8\theta_b^6}\left(-8 + 8S_0 - 5\theta_b A_1 + \theta_b^2 S_2\right)$

Table 5.1: Coefficients of the solution for the sixth order polynominal interpolation for both linear and exponential extrapolation schemes.

two extrapolation cases are shown in Table 5.1. In the exponential case, symmetric and asymmetric terms are first defined to simplify the algebra:

$$\begin{aligned} S_0 &= \tfrac{1}{2}\left(\kappa^+ + \kappa^-\right) \\ A_0 &= \tfrac{1}{2}\left(\kappa^+ - \kappa^-\right) \\ S_1 &= \tfrac{1}{2}\left(\kappa^+ \log\kappa^+ - \kappa^- \log\kappa^-\right) \\ A_1 &= \tfrac{1}{2}\left(\kappa^+ \log\kappa^+ + \kappa^- \log\kappa^-\right) \\ S_2 &= \tfrac{1}{2}\left\{\kappa^+ \left(\log\kappa^+\right)^2 + \kappa^- \left(\log\kappa^-\right)^2\right\} \\ A_2 &= \tfrac{1}{2}\left\{\kappa^+ \left(\log\kappa^+\right)^2 - \kappa^- \left(\log\kappa^-\right)^2\right\} \end{aligned} \tag{5.38}$$

The interplolation for normalization systematics is tested for the four models shown in Table 5.2. These have varying degrees of asymmetrical systematics. Model 4 is symmetric and acts as a control model for the interpolation, since the response function does not necessarily reduce to the standard exponential interpolation in the case that the asymmetry is zero. The sampling distributions of the nuisance parameter for the four models are shown in Figure 5.4. It can be seen that even in cases of very small asymmetries, such as Model 3, the kink in the likelihood can still cause a substantial delta function in the sampling distribution of the nuisance parameter. The control model shows no strange features in either interpolation schemes, as desired.

One shortcoming of this interpolation method is that it is not positive definite for all

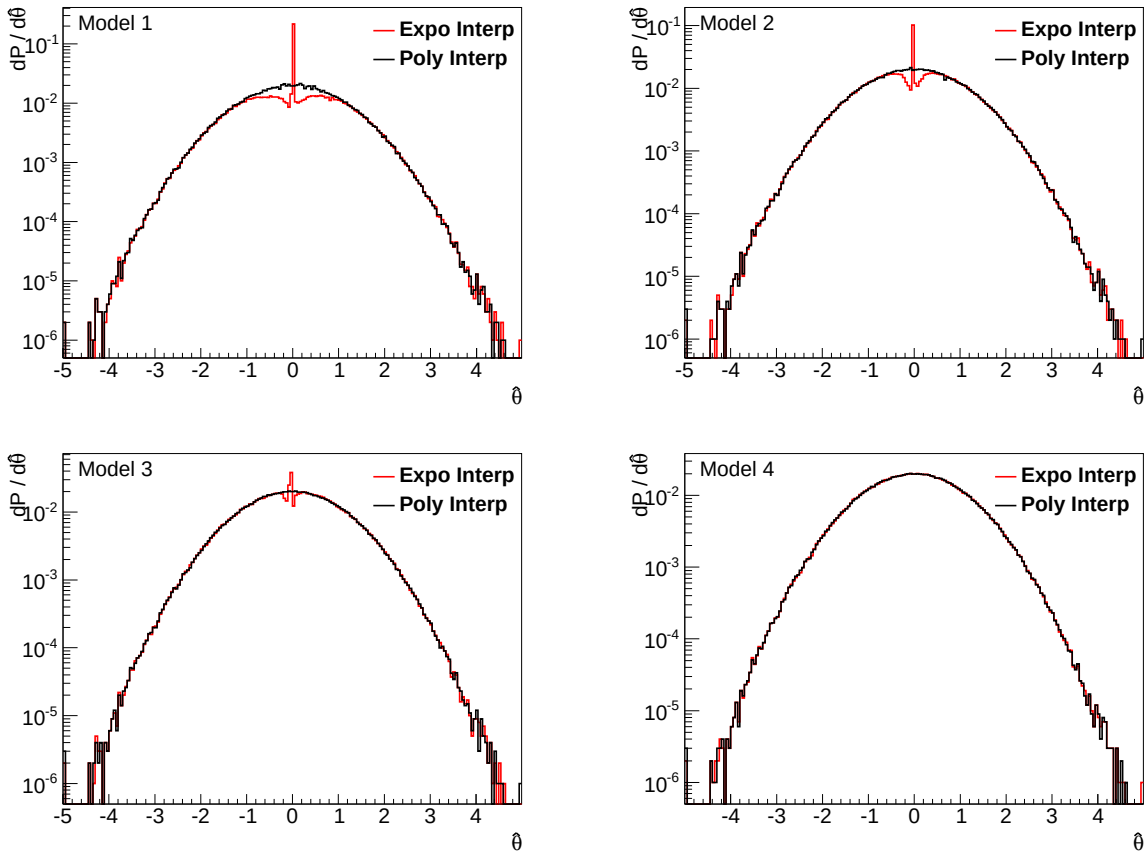


Figure 5.4: Sampling distributions of the nuisance parameter for the four models listed in Table 5.2 with and without the polymonial interpolation scheme.

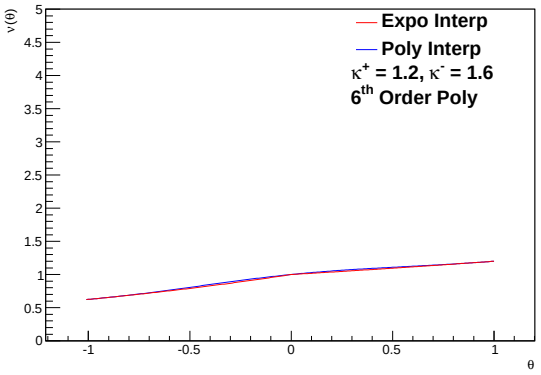
Model	N_{Sig}	N_{Bkg}	N_{Obs}	κ^{+}_{Bkg}	κ^{-}_{Bkg}
1	50	200	200	1.2	1.05
2	50	200	200	1.12	1.08
3	50	200	200	1.105	1.095
4	50	200	200	1.1	1.1

Table 5.2: Parameters of models used to test nuisance parameter interpolation.

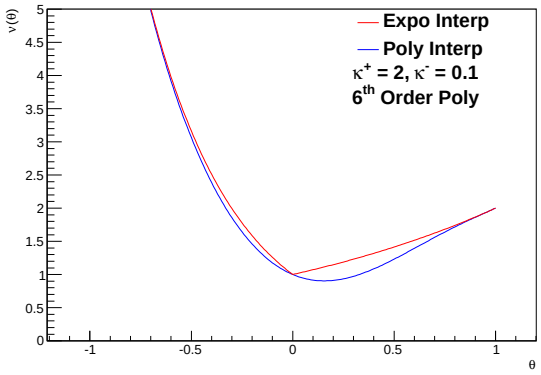
values of the boundaries $\nu(\pm\theta_b)$. For exponential extrapolation with very large asymmetries, the standard 6^{th} order polynomial can become negative. These are extreme cases, however. It has been observed that higher order polynomials can handle larger asymmetries without becoming negative, but this comes at the price of performance since the higher order terms must be computed many times for each likelihood evaluation. Figure 5.5 shows both a well behaved positive definite and an extreme case for $\nu(\theta)$. Figure 5.6 shows the same configurations for a 10^{th} order polynomial where it is now well behaved.

5.3.5 Jet Binned Uncertainties

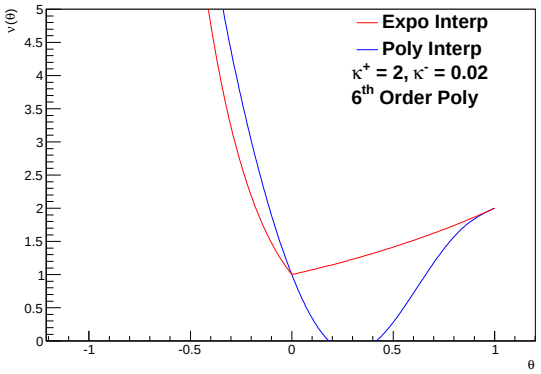
Several Higgs analyses separate channels by exclusive jet bins and require careful treatment of QCD scale systematics. Scale uncertainties on N-jet inclusive cross sections are taken as independent as opposed to the uncertainties on exclusive cross sections, resulting in anti-correlated migration-like uncertainties between jet bins. Further, these correlations need to be derived such that the log-normal treatment for normalization systematics still holds. This is addressed through the Stewart-Tackmann procedure [71]. Many Higgs analyses use exclusive 0- and 1-jet channels to target the gg→H production, and an inclusive 2-jet channel to target VBF Higgs production. The gg→H signal, having zero jets at leading order, is therefore subject to this procedure. Denote the three relevant inclusive production cross sections as σ_{tot}, $\sigma_{\geq 1}$, and $\sigma_{\geq 2}$, with associated relative uncertainties ε_{tot}, $\varepsilon_{\geq 1}$, and $\varepsilon_{\geq 2}$. Denote the parton level exclusive jet fractions as f_0 and f_1, and the inclusive 2-jet fraction $f_{\geq 2}$. The exclusive cross sections are therefore $\sigma_0 = \sigma_{tot} - \sigma_{\geq 1}$ and $\sigma_1 = \sigma_{\geq 1} - \sigma_{\geq 2}$. We desire to compute the exclusive cross sections as a function of the nuisance parameters θ_{tot},



(a) Interpolation for a typical asymmetry

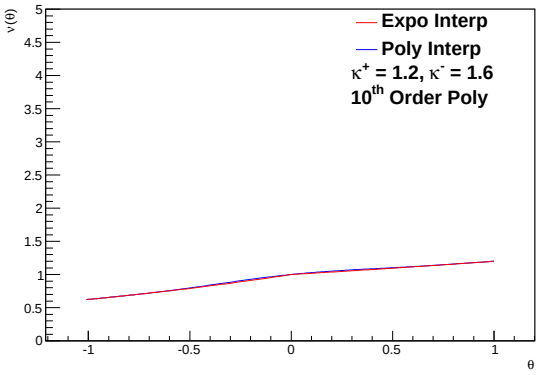


(b) Interpolation for a mild asymmetry

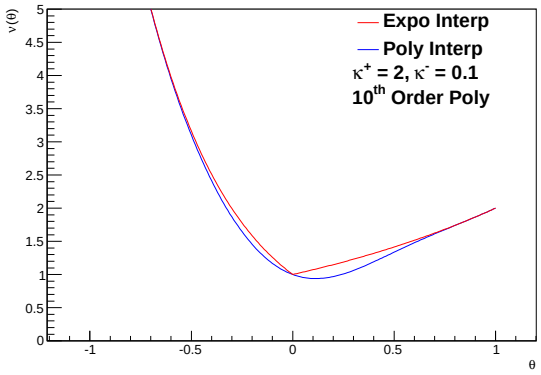


(c) Interpolation for an extreme asymmetry

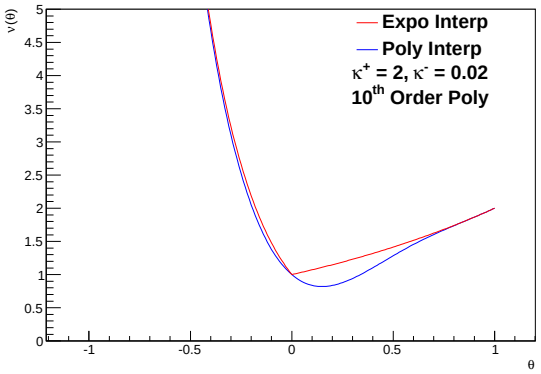
Figure 5.5: 6^{th} order polynomial response term interpolation for a typical (a), mild (b) and an extreme (c) asymmetry.



(a) Interpolation for a typical asymmetry



(b) Interpolation for a mild asymmetry



(c) Interpolation for an extreme asymmetry

Figure 5.6: 10^{th} order polynomial response term interpolation for a typical (a), mild (b) and an extreme (c) asymmetry.

$\theta_{\geq 1}$, and $\theta_{\geq 2}$. The 0-jet exclusive cross section can be derived as follows:

$$\begin{aligned}
\sigma_0(\theta_{\rm tot}, \theta_{\geq 1}) &\approx \sigma_{\rm tot}(1+\varepsilon_{\rm tot}\theta_{\rm tot}) - \sigma_{\geq 1}(1+\varepsilon_{\geq 1}\theta_{\geq 1}) \\
&= \sigma_{\rm tot} - \sigma_{\geq 1} + \sigma_{\rm tot}\varepsilon_{\rm tot}\theta_{\rm tot} - \sigma_{\geq 1}\varepsilon_{\geq 1}\theta_{\geq 1} \\
&= \sigma_0 + \tfrac{1}{f_0}\sigma_0\varepsilon_{\rm tot}\theta_{\rm tot} - \tfrac{f_1+f_{\geq 2}}{f_0}\sigma_0\varepsilon_{\geq 1}\theta_{\geq 1} \\
&\approx \sigma_0(1+\tfrac{1}{f_0}\varepsilon_{\rm tot}\theta_{\rm tot})(1-\tfrac{f_1+f_{\geq 2}}{f_0}\varepsilon_{\geq 1}\theta_{\geq 1}) \\
&\approx \sigma_0(1+\varepsilon_{\rm tot})^{\frac{1}{f_0}\theta_{\rm tot}}(1+\varepsilon_{\geq 1})^{-\frac{f_1+f_{\geq 2}}{f_0}\theta_{\geq 1}} \\
&\equiv \sigma_0 \left(\kappa_{\rm tot}^{\frac{1}{f_0}}\right)^{\theta_{\rm tot}} \left(\kappa_{\geq 1}^{-\frac{f_1+f_{\geq 2}}{f_0}}\right)^{\theta_{\geq 1}}
\end{aligned} \tag{5.39}$$

The 1-jet exclusive derivation follows similarly:

$$\begin{aligned}
\sigma_1(\theta_1, \theta_{\geq 2}) &\approx \sigma_{\geq 1}(1+\varepsilon_{\geq 1}\theta_{\geq 1}) - \sigma_{\geq 2}(1+\varepsilon_{\geq 2}\theta_{\geq 2}) \\
&= \sigma_{\geq 1} - \sigma_{\geq 2} + \sigma_{\geq 1}\varepsilon_{\geq 1}\theta_{\geq 1} - \sigma_{\geq 2}\varepsilon_{\geq 2}\theta_{\geq 2} \\
&= \sigma_1 + \tfrac{f_1+f_{\geq 2}}{f_1}\sigma_1\varepsilon_{\geq 1}\theta_{\geq 1} - \tfrac{f_{\geq 2}}{f_1}\sigma_1\varepsilon_{\geq 2}\theta_{\geq 2} \\
&\approx \sigma_1(1+\tfrac{f_1+f_{\geq 2}}{f_1}\varepsilon_{\geq 1}\theta_{\geq 1})(1-\tfrac{f_{\geq 2}}{f_1}\varepsilon_{\geq 2}\theta_{\geq 2}) \\
&\approx \sigma_1(1+\varepsilon_{\geq 1})^{\frac{f_1+f_{\geq 2}}{f_1}\theta_{\geq 1}}(1+\varepsilon_{\geq 2})^{-\frac{f_{\geq 2}}{f_1}\theta_{\geq 2}} \\
&\equiv \sigma_1 \left(\kappa_{\geq 1}^{\frac{f_1+f_{\geq 2}}{f_1}}\right)^{\theta_{\geq 1}} \left(\kappa_{\geq 2}^{-\frac{f_{\geq 2}}{f_1}}\right)^{\theta_{\geq 2}}
\end{aligned} \tag{5.40}$$

The 2-jet inclusive process is nearly trivial:

$$\begin{aligned}
\sigma_{\geq 2}(\theta_{\geq 2}) &= \sigma_{\geq 2}(1+\varepsilon_{\geq 2}\theta_{\geq 2}) \\
&\approx \sigma_{\geq 2}(1+\varepsilon_{\geq 2})^{\theta_{\geq 2}} \\
&\equiv \sigma_{\geq 2}(\kappa_{\geq 2})^{\theta_{\geq 2}}
\end{aligned} \tag{5.41}$$

5.3.6 Branching Ratio Uncertainties

Branching ratio uncertainties require a slight modification to insure that the property $\sum_i^{\rm decays} BR_i(\theta_i) = 1$ is true for all values of θ_i, which is broken in both exponential ($BR_i(\theta_i) = BR_{i,0}(1+\varepsilon_i)^{\theta_i}$) and linear treatments ($BR_i(\theta_i) = BR_{i,0}(1+\varepsilon_i\theta_i)$). This can be solved by a

simple normalization of each term to the sum of all:

$$BR_i(\theta_i) = \frac{BR_{i,0}(1+\varepsilon_i)^{\theta_i}}{\sum_j BR_{j,0}(1+\varepsilon_j)^{\theta_j}} \tag{5.42}$$

This has the consequence that, even outside of combination, branching ratio uncertainties from every channel have a small but non-zero impact on each individual channel.

5.4 Asymptotics

Under certain conditions, the distribution of a test statistic can be approximated analytically. The formula describing the distribution depends on the test statistic used. This section will outline the asymptotic approximations in the case of a single parameter of interest.

5.4.1 Approximate Sampling Distributions

As the sample size of a model increases, the sampling distribution of $\hat{\mu}$ under the μ' hypothesis approaches a normal distribution centered at μ' with width σ:

$$f(\hat{\mu}|\mu') \to \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\hat{\mu}-\mu')^2}{2\sigma^2}} \tag{5.43}$$

If the test statistic Q is monotonic in $\hat{\mu}$ there exists a one-to-one correspondence between Q and $\hat{\mu}$, such that $f(\hat{\mu}|\mu')$ can be transformed into the distribution of Q through a change of variables:

$$f(Q|\mu') = f(\hat{\mu}|\mu') \left| \frac{d\hat{\mu}}{dQ} \right| \tag{5.44}$$

Through the Wald approximation [72], $-2\ln\lambda(\mu)$ is parabolic around $\hat{\mu}$ up to higher

order terms. A similar statement can be made for $-2\ln\tilde{\lambda}(\mu)$:

$$\begin{aligned}
-2\ln\lambda(\mu) &= \left(\tfrac{\mu-\hat{\mu}}{\sigma}\right)^2 + \mathcal{O}(1/\sqrt{N}) \\
-2\ln\tilde{\lambda}(\mu) &= \begin{cases} -2\ln\frac{\mathcal{L}(\mu,\hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu},\hat{\theta})}, & \hat{\mu} > 0 \\ -2\ln\frac{\mathcal{L}(\mu,\hat{\theta}_\mu)}{\mathcal{L}(0,\hat{\theta}(0))}, & \hat{\mu} \leq 0 \end{cases} \\
&= \begin{cases} -2\ln\lambda(\mu), & \hat{\mu} > 0 \\ -2\ln\lambda(\mu) + 2\ln\lambda(0), & \hat{\mu} \leq 0 \end{cases} \\
&= \mathcal{O}(1/\sqrt{N}) + \begin{cases} \left(\frac{\mu-\hat{\mu}}{\sigma}\right)^2, & \hat{\mu} > 0 \\ \left(\frac{\mu}{\sigma}\right)^2 - \frac{2\mu\hat{\mu}}{\sigma^2}, & \hat{\mu} \leq 0 \end{cases}
\end{aligned} \tag{5.45}$$

For the uncapped test statistic $\tilde{r}_\mu$, the relevant substitutions are as follows:

$$\begin{aligned}
\hat{\mu} &= \begin{cases} \frac{1}{2}(\mu - \frac{\tilde{r}_\mu}{\mu/\sigma^2}), & \hat{\mu} \leq 0 \\ \mu - \sigma\sqrt{\tilde{r}_\mu}, & 0 < \hat{\mu} < \mu \\ \mu + \sigma\sqrt{-\tilde{r}_\mu}, & \hat{\mu} \geq \mu \end{cases} \\
\left|\frac{d\tilde{r}_\mu}{d\hat{\mu}}\right| &= \begin{cases} \frac{2\mu}{\sigma^2}, & \hat{\mu} \leq 0 \\ \frac{2\sqrt{\tilde{r}_\mu}}{\sigma}, & 0 < \hat{\mu} < \mu \\ \frac{2\sqrt{-\tilde{r}_\mu}}{\sigma}, & \hat{\mu} \geq \mu \end{cases}
\end{aligned} \tag{5.46}$$

The $\hat{\mu}$ conditionals $\hat{\mu} \leq 0$, $0 < \hat{\mu} < \mu$, and $\hat{\mu} \geq \mu$ correspond to test statistic conditionals $\tilde{r}_\mu \geq (\frac{\mu-\mu'}{\sigma})^2$, $0 < \tilde{r}_\mu < (\frac{\mu-\mu'}{\sigma})^2$, and $\tilde{r}_\mu < 0$, respectively. This along with the substitutions

above yield the approximate sampling distribution:

$$f(\tilde{r}_\mu|\mu') = \begin{cases} \frac{1}{2\sqrt{2\pi}(\mu/\sigma)} e^{-\frac{1}{2}\left(\frac{\tilde{r}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2}{2\mu/\sigma}\right)^2} & \tilde{r}_\mu \geq (\frac{\mu-\mu'}{\sigma})^2 \\ \frac{1}{2\sqrt{2\pi\tilde{r}_\mu}} e^{-\frac{1}{2}\left(\sqrt{\tilde{r}_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2} & 0 < \tilde{r}_\mu < (\frac{\mu-\mu'}{\sigma})^2 \\ \frac{1}{2\sqrt{-2\pi\tilde{r}_\mu}} e^{-\frac{1}{2}\left(-\sqrt{-\tilde{r}_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2} & \tilde{r}_\mu \leq 0 \end{cases} \tag{5.47}$$

Integrating gives the corresponding cumulative function $F(\tilde{r}_\mu|\mu')$:

$$F(\tilde{r}_\mu|\mu') = \begin{cases} \Phi(\frac{\tilde{r}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2}{2\mu/\sigma}), & \tilde{r}_\mu \geq (\frac{\mu}{\sigma})^2 \\ \Phi(\sqrt{\tilde{r}_\mu} - \frac{\mu-\mu'}{\sigma}), & 0 < \tilde{r}_\mu < (\frac{\mu}{\sigma})^2 \\ \Phi(-\sqrt{-\tilde{r}_\mu} - \frac{\mu-\mu'}{\sigma}), & \tilde{r}_\mu \leq 0 \end{cases} \tag{5.48}$$

$\Phi(x)$ is the cumulative distribution function of the standard normal distribution:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt \tag{5.49}$$

The p-value p_μ corresponds to the case $\mu = \mu'$, with $p_\mu = 1 - F(\tilde{r}_\mu|\mu)$, while p_b corresponds to $\mu' = 0$, with $p_b = F(\tilde{r}_\mu|0)$.

The approximate sampling distributions for the test statistic r_μ are similar, with the modification that the $0 < r_\mu < (\frac{\mu-\mu'}{\sigma})^2$ conditional extends to all $r_\mu > 0$:

$$f(r_\mu|\mu') = \begin{cases} \frac{1}{2\sqrt{2\pi r_\mu}} e^{-\frac{1}{2}\left(\sqrt{r_\mu} - \frac{\mu}{\sigma}\right)^2} & r_\mu > 0 \\ \frac{1}{2\sqrt{-2\pi r_\mu}} e^{-\frac{1}{2}\left(-\sqrt{-r_\mu} - \frac{\mu}{\sigma}\right)^2} & r_\mu \leq 0 \end{cases} \tag{5.50}$$

The cumulative function follows:

$$F(r_\mu|\mu') = \begin{cases} \Phi(\sqrt{r_\mu} - \frac{\mu-\mu'}{\sigma}), & r_\mu > 0 \\ \Phi(-\sqrt{-r_\mu} - \frac{\mu-\mu'}{\sigma}), & r_\mu \leq 0 \end{cases} \tag{5.51}$$

For r_0, the sampling distributions are nearly identical:

$$f(r_0|\mu') = \begin{cases} \frac{1}{2\sqrt{2\pi r_0}} e^{-\frac{1}{2}\left(\sqrt{r_0} - \frac{\mu-\mu'}{\sigma}\right)^2} & r_0 > 0 \\ \frac{1}{2\sqrt{-2\pi r_0}} e^{-\frac{1}{2}\left(-\sqrt{-r_0} - \frac{\mu-\mu'}{\sigma}\right)^2} & r_0 \leq 0 \end{cases} \tag{5.52}$$

Finally, the cumulative function:

$$F(r_0|\mu') = \begin{cases} \Phi(\sqrt{r_0} - \frac{\mu-\mu'}{\sigma}), & r_0 > 0 \\ \Phi(-\sqrt{-r_0} - \frac{\mu-\mu'}{\sigma}), & r_0 \leq 0 \end{cases} \tag{5.53}$$

The distributions and cumulative functions for the remaining test statistics follow.

- t_μ

$$\begin{aligned} f(t_\mu|\mu') &= \frac{1}{2\sqrt{2\pi t_\mu}} \left[e^{-\frac{1}{2}\left(\sqrt{t_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\sqrt{t_\mu} + \frac{\mu-\mu'}{\sigma}\right)^2} \right] \\ F(t_\mu|\mu') &= \Phi(\sqrt{t_\mu} - \tfrac{\mu-\mu'}{\sigma}) + \Phi(\sqrt{t_\mu} + \tfrac{\mu-\mu'}{\sigma}) - 1 \end{aligned} \tag{5.54}$$

- $\tilde{t}_\mu$

$$f(\tilde{t}_\mu|\mu') = \frac{1}{2\sqrt{2\pi\tilde{t}_\mu}} e^{-\frac{1}{2}\left(\sqrt{\tilde{t}_\mu} + \frac{\mu-\mu'}{\sigma}\right)^2} + \begin{cases} \frac{1}{2\sqrt{2\pi}(\mu/\sigma)} e^{-\frac{1}{2}\left(\frac{\tilde{t}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2}{2\mu/\sigma}\right)^2}, & \tilde{t}_\mu \geq (\frac{\mu}{\sigma})^2 \\ \frac{1}{2\sqrt{2\pi\tilde{t}_\mu}} e^{-\frac{1}{2}\left(\sqrt{\tilde{t}_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2}, & \tilde{t}_\mu < (\frac{\mu}{\sigma})^2 \end{cases}$$

$$F(\tilde{t}_\mu|\mu') = \Phi(\sqrt{\tilde{t}_\mu} + \tfrac{\mu-\mu'}{\sigma}) + \begin{cases} \Phi(\frac{\tilde{t}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2}{2\mu/\sigma}), & \tilde{t}_\mu \geq (\frac{\mu}{\sigma})^2 \\ \Phi(\sqrt{\tilde{t}_\mu} - \frac{\mu-\mu'}{\sigma}), & \tilde{t}_\mu < (\frac{\mu}{\sigma})^2 \end{cases} \tag{5.55}$$

- q_μ

$$\begin{aligned} f(q_\mu|\mu') &= \Phi(\tfrac{\mu'-\mu}{\sigma})\delta(q_\mu) + \frac{1}{2\sqrt{2\pi q_\mu}} e^{-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2} \\ F(q_\mu|\mu') &= \Phi(\sqrt{q_\mu} - \tfrac{\mu-\mu'}{\sigma}) \end{aligned} \tag{5.56}$$

- $\tilde{q}_\mu$

$$
\begin{aligned}
f(\tilde{q}_\mu|\mu') &= \Phi(\tfrac{\mu'-\mu}{\sigma})\delta(\tilde{q}_\mu) + \begin{cases} \frac{1}{2\sqrt{2\pi}(\mu/\sigma)} e^{-\frac{1}{2}\left(\frac{\tilde{q}_\mu-(\mu^2-2\mu\mu')/\sigma^2}{2\mu/\sigma}\right)^2}, & \tilde{q}_\mu \geq (\frac{\mu}{\sigma})^2 \\ \frac{1}{2\sqrt{2\pi\tilde{q}_\mu}} e^{-\frac{1}{2}\left(\sqrt{\tilde{q}_\mu}-\frac{\mu-\mu'}{\sigma}\right)^2}, & \tilde{q}_\mu < (\frac{\mu}{\sigma})^2 \end{cases} \\
F(\tilde{q}_\mu|\mu') &= \begin{cases} \Phi(\frac{\tilde{q}_\mu-(\mu^2-2\mu\mu')/\sigma^2}{2\mu/\sigma}), & \tilde{q}_\mu \geq (\frac{\mu}{\sigma})^2 \\ \Phi(\sqrt{\tilde{q}_\mu}-\frac{\mu-\mu'}{\sigma}), & \tilde{q}_\mu < (\frac{\mu}{\sigma})^2 \end{cases}
\end{aligned}
\tag{5.57}
$$

- q_0

$$
\begin{aligned}
f(q_0|\mu') &= (1-\Phi(\tfrac{\mu'}{\sigma}))\delta(q_\mu) + \tfrac{1}{2\sqrt{2\pi q_0}} e^{-\frac{1}{2}\left(\sqrt{q_0}-\frac{\mu'}{\sigma}\right)^2} \\
F(q_0|\mu') &= \Phi(\sqrt{q_0}-\tfrac{\mu'}{\sigma})
\end{aligned}
\tag{5.58}
$$

Finally, the sampling distribution for q_{RPL} follows a Gaussian distribution with mean $E[q] = \frac{1-2\mu'}{\sigma}$ and variance $V[q] = \frac{4}{\sigma^2}$. This gives the following cumulative function.

$$
F(q_{\text{RPL}}|\mu') = \Phi\left(\frac{q_{\text{RPL}} - (1-2\mu')/\sigma^2}{2/\sigma}\right) \tag{5.59}
$$

This leads to an important statement about the sampling distributions of test statistics that can be written as a function of the PLR: The shape of the distributions depend only on the tested μ, the hypothesized μ', and the variance of μ. Of these, the variance is the only model dependent quantity. As such, the variance merits exploration.

5.4.2 The Variance of μ

The variance of μ, namely σ^2, can be conceptualized from the distribution $f(\hat{\mu}|\mu')$. The p-value that is the tail probability of obtaining a result $\hat{\mu} > \mu$ given a hypothesized μ' can be converted into a significance Z, as shown in Figure 5.7. Z is the number of σ that μ is

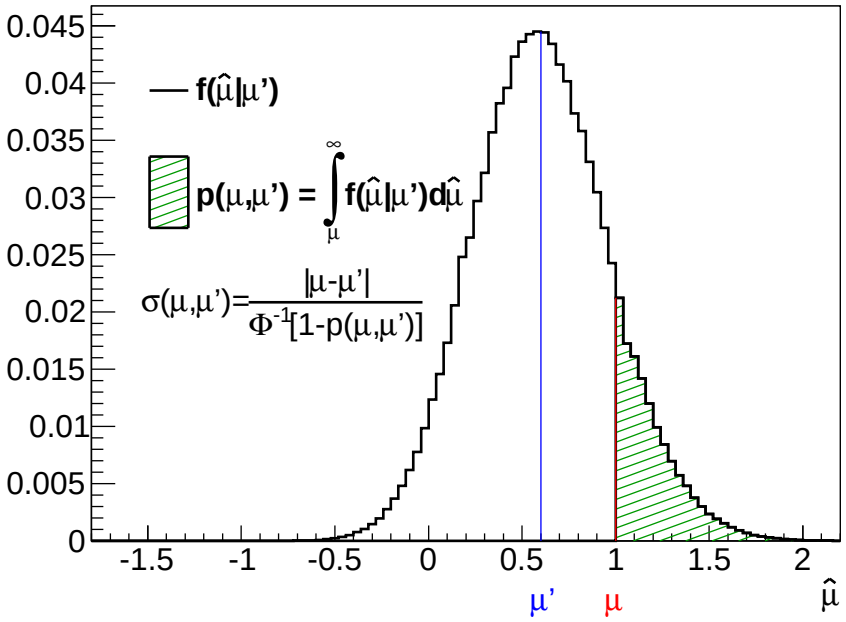


Figure 5.7: The sampling distribution $f(\hat{\mu}|\mu')$ can be used to derive $\sigma(\mu, \mu')$. Slight non-Gaussian behavior of the distribution causes dependence on both μ and μ'.

from μ', such that σ is recovered by simply dividing $\mu - \mu'$ by Z:

$$\begin{aligned} p(\mu, \mu') &= \int_{\mu}^{\infty} f(\hat{\mu}|\mu') d\hat{\mu} \\ Z(\mu, \mu') &= \Phi^{-1}(1 - p(\mu, \mu')) \\ \sigma(\mu, \mu') &= \frac{\mu - \mu'}{Z(\mu, \mu')} \end{aligned} \tag{5.60}$$

The dependence of σ on μ and μ' is shown explicitly here. If $f(\hat{\mu}|\mu')$ is a perfect Gaussian, $Z(\mu, \mu')$ is linear in $\mu - \mu'$ and so $\sigma(\mu, \mu')$ is constant. Unfortunately this is only the case for very idealistic models. Practically, $\sigma(\mu, \mu')$ must be estimated at each μ and μ'. The dependence of σ on μ and μ' depends on the size of the systematics in the model and their posterior correlation with μ. For small systematics and small Z, this dependence tends to be weak and linear. For models with large systematics, $f(\hat{\mu}|\mu')$ can become non-Gaussian very quickly and lead to non-linear behavior in σ.

Given a single dataset, the variance can be estimated with the test statistic following the

Model	N_{Sig}	N_{Bkg}	N_{Obs}	σ_{Sig}	σ_{Bkg}	σ_{Common}
1	50	200	200	-	-	-
2	50	200	200	10%	10%	10%
3	50	200	200	-	-	10%
4	50	200	200	-	10%	-
5	50	200	200	10%	-	-
6	50	200	250	-	-	10%

Table 5.3: Parameters of models used in tests to derive asymptotic bands. Model 6 is identical to Model 3 with the exception of the observed data, which has been changed to test the effects of the profiling on the results.

Wald approximation:

$$t_\mu \approx \left(\frac{\mu - \hat{\mu}}{\sigma}\right)^2 \implies \sigma \approx \frac{|\mu - \hat{\mu}|}{\sqrt{t_\mu}} \tag{5.61}$$

We are interested in the value of σ that characterizes the distribution of the test statistic. This corresponds to the median test statistic value, and can therefore be characterized by $t_{\mu,A_{\mu'}}$, which is the test statistic t_μ evaluated on the μ' Asimov dataset:

$$\sigma(\mu, \mu') \approx \frac{|\mu - \mu'|}{\sqrt{t_{\mu,A_{\mu'}}}} \tag{5.62}$$

Figure 5.8 shows examples for the six models outlined in Table 5.3. Each is a high statistics simple number counting experiment with single background and various degrees of systematics.

5.4.3 Exclusion Bands

A consequence of the μ and μ' dependence of σ is that the expected bands of an exclusion limit require a reformulation beyond what one obtains from assuming σ is constant. The bands one obtains from the exact procedure obtained with pseudo-experiments takes this dependence into account naturally. It is therefore required that the asymptotic procedure is formulated to follow the results obtained from pseudo-experiments as closely as possible. The aim of this section is to derive such an asymptotic procedure within the $\tilde{q}_\mu$ formalism.

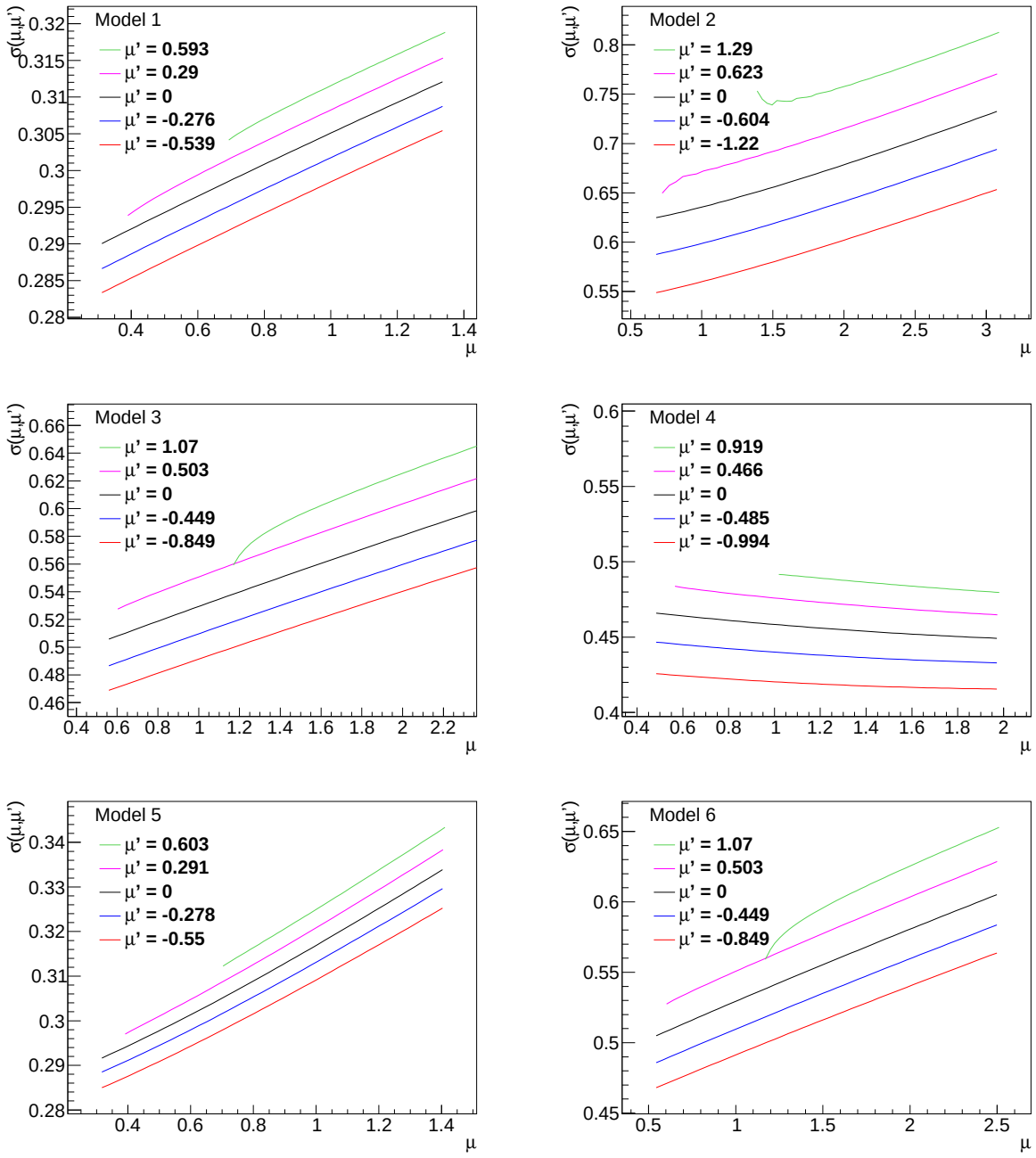


Figure 5.8: The value of $\sigma(\mu, \mu')$ as a function of μ for various values of μ'. Some instabilities (for example, the green curve in Model 2) result from values of $t_{\mu, A_{\mu'}}$ close to zero.

Note that for limits, the use of $\tilde{q}_\mu$ is equivalent to its corresponding uncapped $\tilde{r}_\mu$.

5.4.3.1 Toy Procedure

As a testbed for comparison, high statistics toy results are first used to compute the expected limit and bands for a single bin counting experiment in several configurations. The parameters for the models are outlined in Table 5.3. The sampling distributions $f(\tilde{q}_\mu|\mu)$ and $f(\tilde{q}_\mu|0)$ are computed for several points in μ and are shown in Figure 5.9. The two p-values $p_\mu = \int_{\tilde{q}_\mu}^{\infty} f(\tilde{q}_\mu^{'}|\mu)d\tilde{q}_\mu^{'}$ and $p_b = \int_{-\infty}^{\tilde{q}_\mu} f(\tilde{q}_\mu^{'}|0)d\tilde{q}_\mu^{'}$ are used to compute the ratio $\mathrm{CLs}(\mu;\tilde{q}_\mu) = \frac{p_\mu}{1-p_b}$. For some value μ, a calibrated value of $\tilde{q}_\mu$, denoted $\tilde{q}_\mu^{95}$, can be found that satisfies $\mathrm{CLs}(\mu;\tilde{q}_\mu^{95}) = 0.05$. $\tilde{q}_\mu^{95}$ is the value of the test statistic required for the upper limit to be μ on a given dataset. The 95% upper limit on μ for some given dataset is therefore found by scanning the test statistic until it crosses this calibrated value.

Given the $\tilde{q}_\mu^{95}$ curve, the expected limit and bands are obtained from $\mu' = 0$ pseudo-experiments. For each pseudo-experiment, the 95% confidence level upper limit μ_{up} is found by scanning $\tilde{q}_\mu$ until the crossing $\tilde{q}_{\mu_{up}} = \tilde{q}_\mu^{95}$ is found. From the ensemble one obtains the distribution $f(\mu_{up}|0)$. The N^{th} quantile of this distribution represents μ_{up+N}, the quantiles of the expected upper limit.

At this point it becomes important to distinguish between two key stages of this procedure when translating to asymptotics. The first is the procedure to obtain the $\tilde{q}_\mu^{95}$ curve. This will rely only on the validity of asymptotics and the correct approximation of σ to characterize $f(\tilde{q}_\mu|\mu)$ and $f(\tilde{q}_\mu|0)$. The second is the procedure to obtain the distribution $f(\mu_{up}|0)$ and μ_{up+N} given the $\tilde{q}_\mu^{95}$ curve. Estimating $\tilde{q}_\mu^{95}$ only requires solving the transcendental equation $\mathrm{CLs}(\mu;\tilde{q}_\mu^{95}) = 0.05$ given the analytical asymptotic formula for the p-values p_μ and p_b. Estimating μ_{up+N} involves finding a mapping from $f(\hat{\mu}|0)$ to $f(\mu_{up}|0)$ and is more involved.

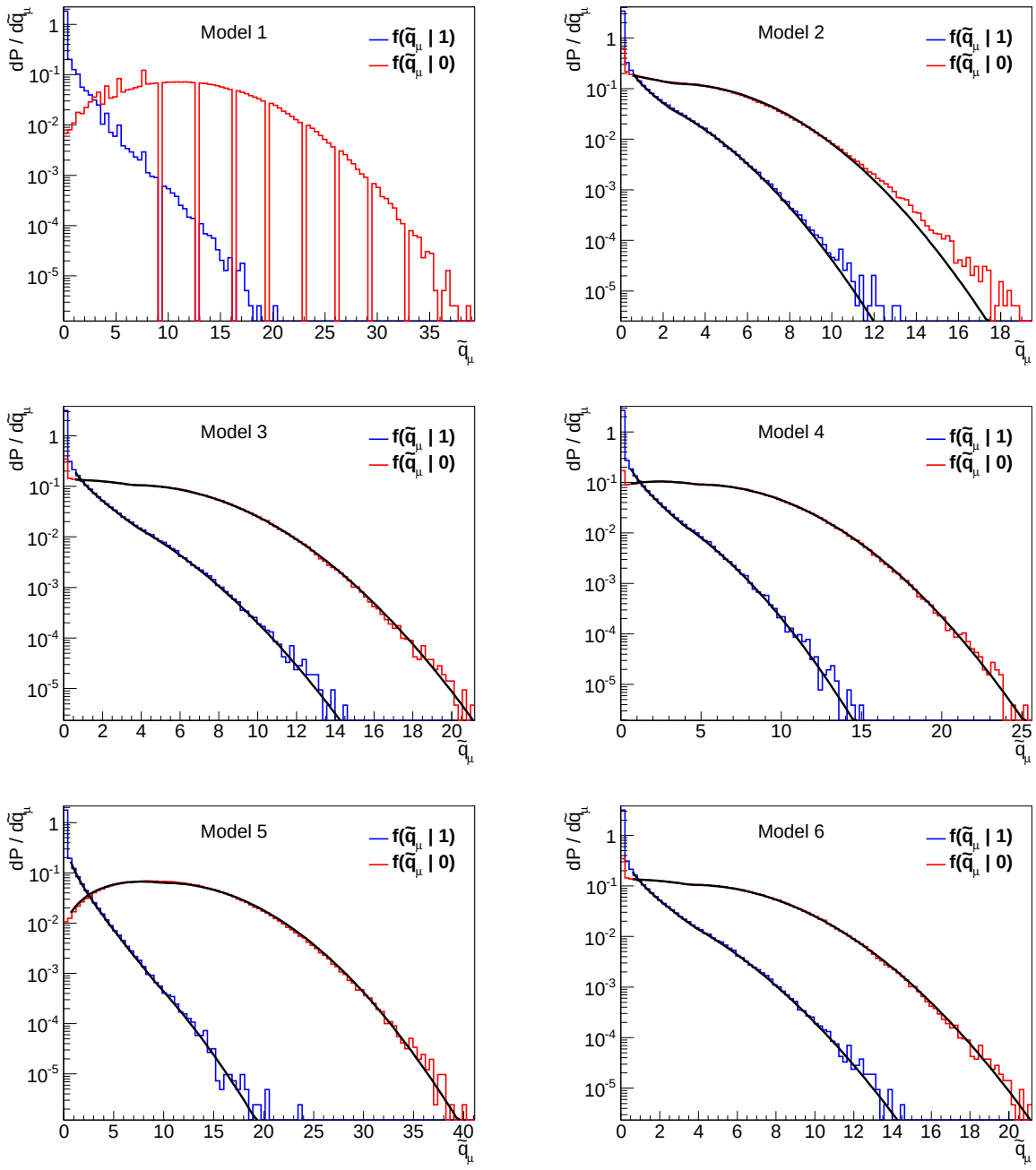


Figure 5.9: Sampling distributions of the test statistic $\tilde{q}_\mu$ for the six models shown in Table 5.3 with asymptotic overlays. Model 1 shows discreteness due to the absence of systematics, so the asymptotic overlay is left out. Note that Model 2 shows a deviation from asymptotics in the distribution under the $\mu' = 0$ hypothesis, likely due to the large systematics in the model.

5.4.3.2 Estimating μ_{up+N}

Consider the distribution $f(\hat{\mu}|0)$ derived from pseudo-experiments. Assuming that the Wald approximation holds and that σ is constant, toys with $\hat{\mu} = N\sigma$ are characteristic of the expected N^{th} quantile band, with $N = 0$ representing the median. In this case the value of the test statistic at the N^{th} quantile upper limit is $q_{\mu_{up+N}} \approx (\frac{\mu_{up+N} - N\sigma}{\sigma})^2$ (q_μ is used in this case to simplify the algebra). Recalling that CLs $= \frac{\mathrm{p}_\mu}{1-\mathrm{p}_\mathrm{b}}$, with $p_\mu = 1 - \Phi[\sqrt{q_\mu}]$ and $1 - p_b = \Phi[\frac{\mu}{\sigma} - \sqrt{q_\mu}]$, at μ_{up+N} the critical exclusion value is computed as so:

$$\begin{aligned} \alpha &= (1 - \Phi[\tfrac{\mu_{up+N} - N\sigma}{\sigma}])/\Phi[\tfrac{\mu_{up+N}}{\sigma} - \tfrac{\mu_{up+N} - N\sigma}{\sigma}] \\ &= (1 - \Phi[\tfrac{\mu_{up+N}}{\sigma} - N])/\Phi[N] \end{aligned} \tag{5.63}$$

Solving for μ_{up+N}, one obtains $\mu_{up+N} = \sigma\{\Phi^{-1}[1 - \alpha\Phi(N)] + N\}$. This is referred to as Method 1 and is the leading order approximation of μ_{up+N}. If σ is μ dependent, the terms in the denominator of α no longer cancel and the equation for α is transcendental in μ_{up+N}. Further, σ is also a function of μ_{up+N}, and in general this dependence cannot be quantified analytically.

A more rigorous approach for determining the expected N^{th} quantile of an upper limit, refered to as Method 2, requires the correct estimation of the σ terms in the equation for α. Starting with the distribution $f(\hat{\mu}|0)$ and again invoking the Wald approximation, there should exist a one-to-one monotonic mapping of $f(\hat{\mu}|0)$ onto $f(\mu_{up}|0)$. It follows that the N^{th} quantile of $f(\hat{\mu}|0)$ also maps to the N^{th} quantile of $f(\mu_{up}|0)$. The procedure for estimating μ_{up+N} can thus be thought of as two separate issues. The first is the correct estimation of $f(\hat{\mu}|0)$; in particular, it's N^{th} quantile. The second is the correct estimation of the mapping of $f(\hat{\mu}|0)$ onto $f(\mu_{up}|0)$. The estimation of the N^{th} quantiles of $f(\hat{\mu}|0)$ follows from section 5.4.2:

$$\begin{aligned} \sigma(\mu, 0) &= \tfrac{\mu}{N(\mu,0)} \\ \implies N(\mu, 0) &= \tfrac{\mu}{\sigma(\mu,0)} = \sqrt{-2\ln\lambda_{A_0}(\mu)} \end{aligned} \tag{5.64}$$

Denoting the expected N^{th} quantile of $f(\hat{\mu}|0)$ as μ_A^N, it's seen that, for some given N, μ_A^N is given by the solution to $\sqrt{-2\ln\lambda_{A_0}(\mu_A^N)} = N$. Table 5.4 shows the exact and estimated values of μ_A^N for each model. The agreement is generally good at the percent level.

Less straight forward is the estimation of the $f(\hat{\mu}|0) \to f(\mu_{up}|0)$ mapping. It is clear that $\hat{\mu} = \mu_A^N$ characterizes the value of μ for the pseudo-experiments that lie on the N^{th} quantile. More subtle are the values of the nuisance parameters that characterize the quantiles, denoted as θ_A^N. These correspond to the median values of nuisance parameters for pseudo-experiments having an upper limit of μ_{up+N}. The method for estimating θ_A^N will be explored through inspection.

Each model is fully specified by μ and θ, where θ is the nuisance parameter representing the systematic in the model. The distributions of the unconditional estimators of these parameters for each model, $\hat{\mu}$ and $\hat{\theta}$, are shown in Figures 5.10- 5.14. The median value of $\hat{\theta}$ within a small window of the quantiles of $\hat{\mu}$ is shown by the open circles for the $N = -3, -2, ..., +3$ quantiles of $\hat{\mu}$. $\theta_A^N \equiv \hat{\theta}(\mu_A^N)$ at the quantiles is shown by the open triangles, which follows closely the open circles. This shows that $\hat{\theta}(\mu_A^N)$ can provide an asymptotic estimate of the value of $\hat{\theta}$ that characterizes the quantiles of the upper limit.

Given both μ_A^N and θ_A^N, an Asimov dataset, denoted A^N, can be constructed that characterizes pseudo-experiments falling on the N^{th} quantile. Recall that in the exact toy procedure the upper limit of each pseudo-experiment is found from the solution to $\tilde{q}_{\mu_{up}} = \tilde{q}^{95}_{\mu_{up}}$. It follows that μ_{up+N} is also characterized by the solution to $\tilde{q}_{\mu_{up+N},A^N} = \tilde{q}^{95}_{\mu_{up+N}}$. This is the effective procedure to estimate the $f(\hat{\mu}|0) \to f(\mu_{up}|0)$ mapping.

The full procedure to compute the fully asymptotic μ_{up+N} can be summarized as follows:

- Construct a $\mu' = 0$ Asimov dataset, and with it solve $\sqrt{-2\ln\lambda_{A_0}(\mu_A^N)} = N$ to find μ_A^N.
- Simultaneously, extract $\theta_A^N \approx \hat{\theta}(\mu_A^N)$ from the maximum likelihood fit in the numerator of $\lambda_{A_0}(\mu_A^N)$.
- Construct the Asimov dataset A^N with the parameters μ_A^N and θ_A^N.

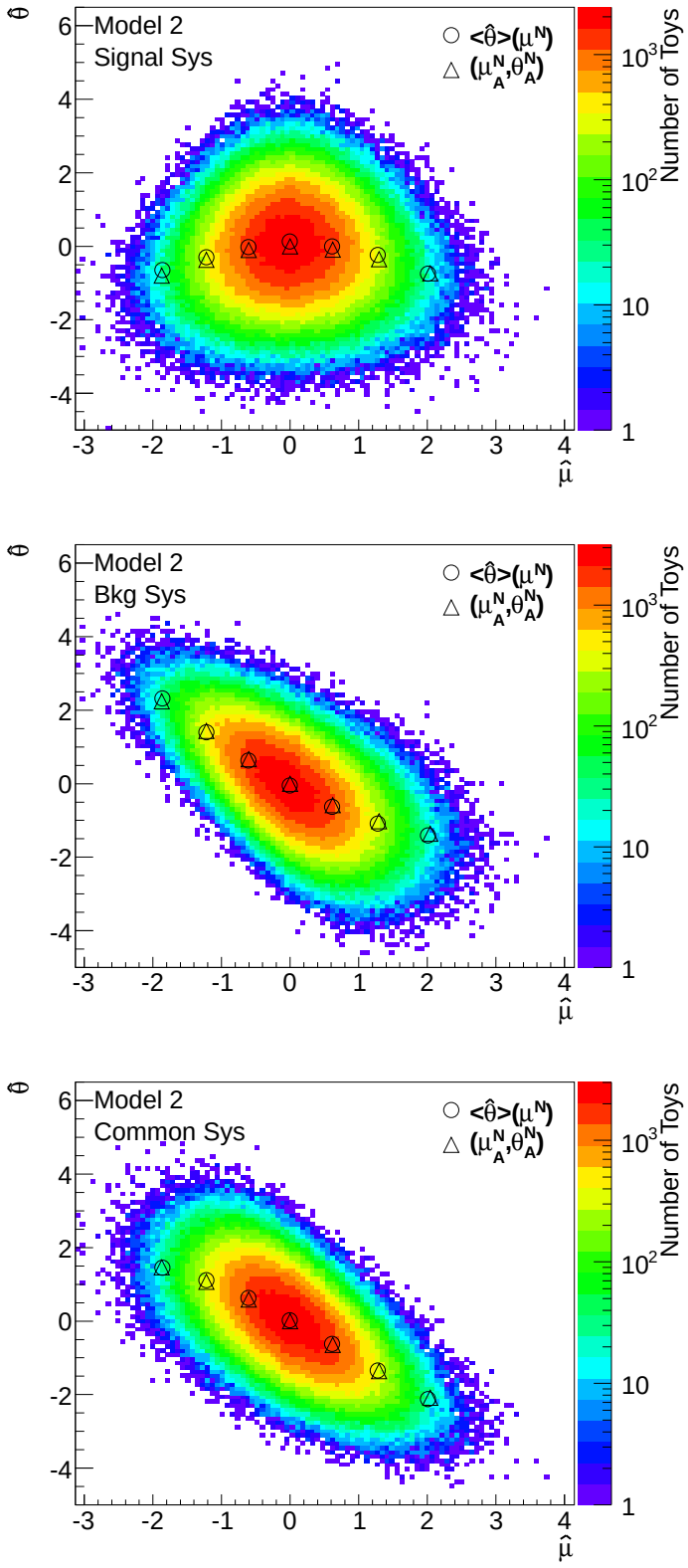


Figure 5.10: Two dimensional sampling distribution of $\hat{\mu}$ vs $\hat{\theta}$ for Model 2.

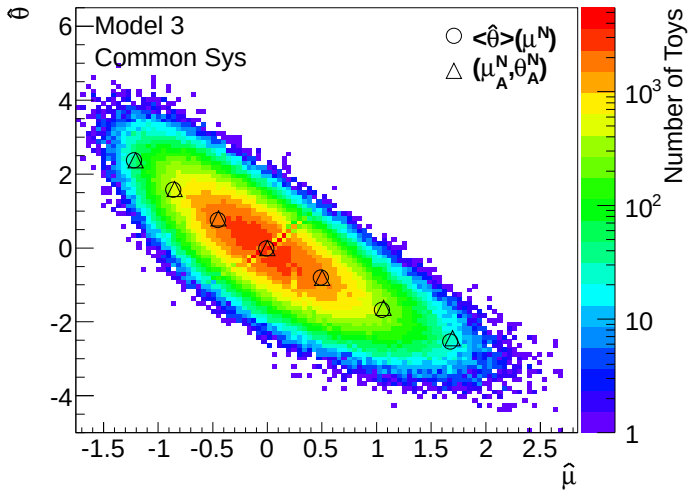


Figure 5.11: Two dimensional sampling distribution of $\hat{\mu}$ vs $\hat{\theta}$ for Model 3.

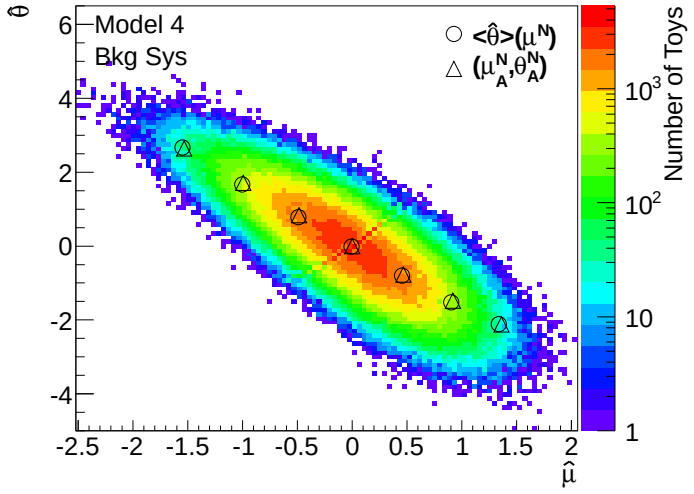


Figure 5.12: Two dimensional sampling distribution of $\hat{\mu}$ vs $\hat{\theta}$ for Model 4.

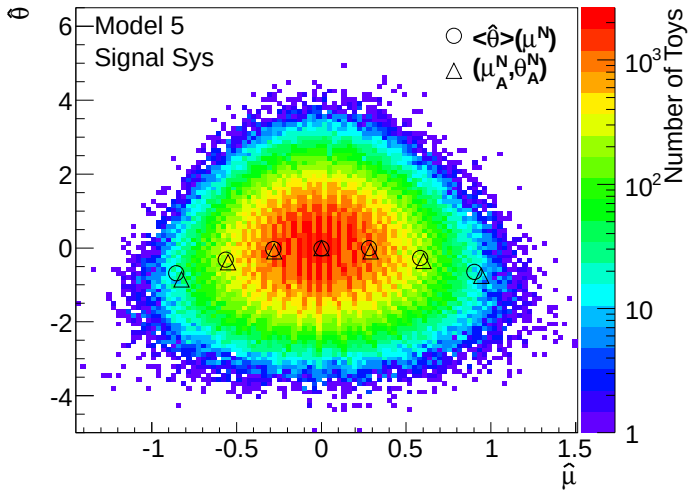


Figure 5.13: Two dimensional sampling distribution of $\hat{\mu}$ vs $\hat{\theta}$ for Model 5.

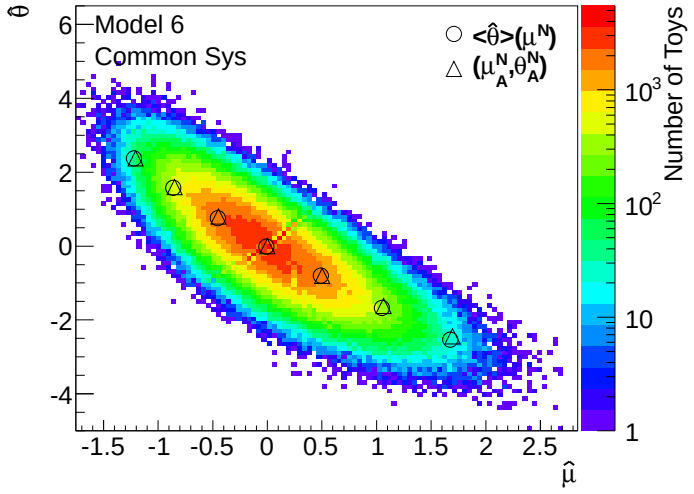


Figure 5.14: Two dimensional sampling distribution of $\hat{\mu}$ vs $\hat{\theta}$ for Model 6.

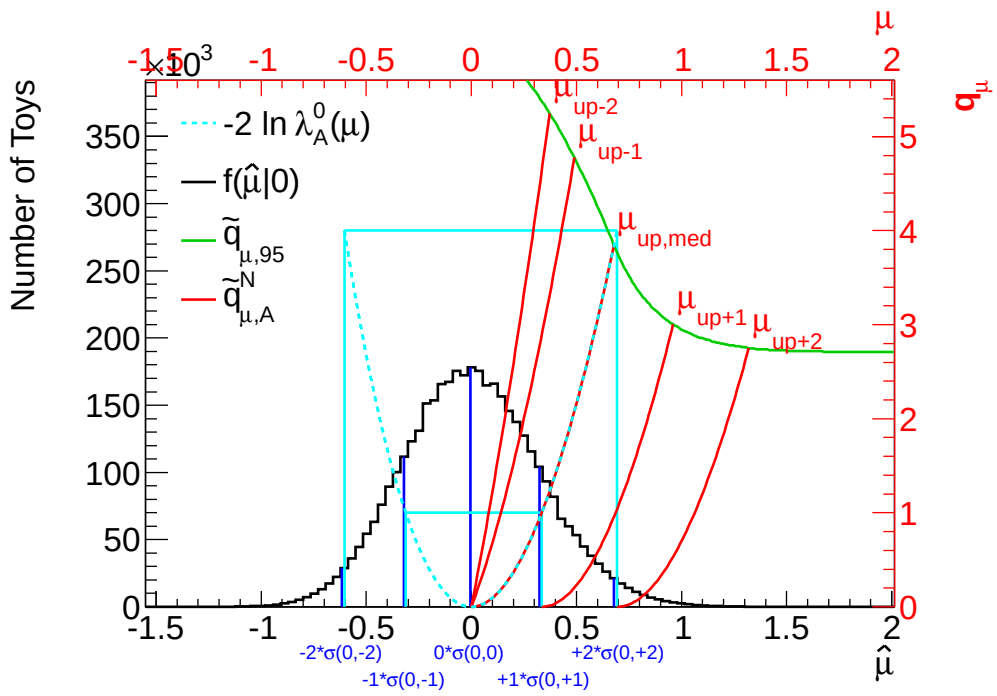


Figure 5.15: Visualization of the improved method for computing the asymptotic bands. In black is the distribution of $\hat{\mu}$ under the background-only hypothesis. The quantiles of this distribution are shown in blue. These characterize the value of $\hat{\mu}$ that will result in the corresponding expected limit quantile. The cyan curve shows that these values of $\hat{\mu}$ are well estimated asymptotically. In red are the curves of $\tilde{q}_\mu$ computed with Asimov datasets constructed with the asymptotically estimated $\hat{\mu}$ quantiles. The quantiles of the expected limit correspond to the crossing of these curves with the asymptotic $\tilde{q}^{95}_\mu$ function in green.

- With A^N, solve $\tilde{q}_{\mu_{up+N},A^N} = \tilde{q}^{95}_{\mu_{up+N}}$ to find μ_{up+N}. The $\tilde{q}^{95}_\mu$ curve can be obtained with the analytical asymptotic expression for CLs, which is computationally inexpensive.

The procedure is visualized in Figure 5.15. It can also be condensed into a simple transcendental equation for μ_{up}:

$$\begin{aligned} \mu_{up+N} &= N_1\sigma(\mu^N, 0) + N_2\sigma(\mu_{up+N}, \mu^N) \\ N_1 &= \sqrt{-2\ln\lambda_{A_0}(\mu_A^{N_1})} \\ N_2 &= \sqrt{\tilde{q}^{95}_{\mu_{up+N},A^N}} \end{aligned} \tag{5.65}$$

The results from Method 1, Method 2, and the exact toy procedure for all six models are shown in Table 5.5. It can be seen that Method 2 is in better agreement with the toy results

than Method 1. Similar tests have shown that deviations in Method 1 are exaggerated in models with large systematics that cause σ to become highly μ dependent, while Method 2 replicates the toy results.

Model 1							
	-3	-2	-1	0	1	2	3
μ^N	-0.82	-0.56	-0.29	0	0.28	0.58	0.88
μ_A^N	-0.79	-0.54	-0.28	0	0.29	0.59	0.91
Model 2							
	-3	-2	-1	0	1	2	3
μ^N	-1.86	-1.22	-0.61	-0.01	0.61	1.27	2.01
μ_A^N	-1.87	-1.22	-0.60	0	0.62	1.29	2.04
θ_s^N	-0.65	-0.30	-0.02	0.127	-0.01	-0.24	-0.74
$\theta_{A,s}^N$	-0.80	-0.37	-0.10	0	-0.09	-0.35	-0.74
θ_b^N	2.32	1.39	0.63	-0.05	-0.63	-1.09	-1.40
$\theta_{A,b}^N$	2.24	1.43	0.66	0	-0.59	-1.02	-1.36
θ_c^N	1.45	1.11	0.63	0.02	-0.63	-1.34	-2.12
$\theta_{A,c}^N$	1.46	1.07	0.59	0	-0.65	-1.36	-2.09
Model 3							
	-3	-2	-1	0	1	2	3
μ^N	-1.22	-0.86	-0.46	-0.01	0.49	1.05	1.67
μ_A^N	-1.21	-0.85	-0.45	0	0.50	1.07	1.70
θ_c^N	2.37	1.58	0.76	-0.01	-0.80	-1.67	-2.53
$\theta_{A,c}^N$	2.37	1.59	0.80	0	-0.81	-1.62	-2.44
Model 4							
	-3	-2	-1	0	1	2	3
μ^N	-1.55	-1.00	-0.49	-0.01	0.46	0.91	1.34
μ_A^N	-1.53	-1.00	-0.49	0	0.47	0.92	1.36
θ_b^N	2.66	1.67	0.79	-0.01	-0.80	-1.52	-2.11
$\theta_{A,b}^N$	2.64	1.71	0.83	0	-0.77	-1.48	-2.12
Model 5							
	-3	-2	-1	0	1	2	3
μ^N	-0.86	-0.56	-0.28	0	0.28	0.58	0.90
μ_A^N	-0.82	-0.55	-0.28	0	0.29	0.60	0.94
θ_s^N	-0.69	-0.32	-0.028	-0.01	-0.01	-0.27	-0.65
$\theta_{A,s}^N$	-0.85	-0.39	-0.097	0	-0.09	-0.35	-0.75
Model 6							
	-3	-2	-1	0	1	2	3
μ^N	-1.22	-0.86	-0.46	-0.01	0.49	1.05	1.67
μ_A^N	-1.21	-0.85	-0.45	0	0.50	1.07	1.70
θ_c^N	2.37	1.58	0.76	-0.01	-0.80	-1.67	-2.53
$\theta_{A,c}^N$	2.37	1.59	0.80	0	-0.81	-1.62	-2.44

Table 5.4: Estimated and exact values of parameters that characterize the N^{th} quantile expected upper limits on μ. The bias in θ_s^N is related to the asymptotic breaking shown in Figure 5.9

Model 1					
	Expected	$+2\sigma$	$+1\sigma$	-1σ	-2σ
Method 1	0.58	1.08	0.81	0.42	0.31
Method 2	0.58	1.11	0.82	0.42	0.32
Toys	0.58	1.10	0.81	0.42	0.31
Model 2					
	Expected	$+2\sigma$	$+1\sigma$	-1σ	-2σ
Method 1	1.27	2.36	1.76	0.91	0.68
Method 2	1.27	2.56	1.82	0.91	0.71
Toys	1.26	2.55	1.81	0.89	0.64
Model 3					
	Expected	$+2\sigma$	$+1\sigma$	-1σ	-2σ
Method 1	1.04	1.94	1.45	0.75	0.56
Method 2	1.04	2.14	1.51	0.73	0.54
Toys	1.04	2.13	1.51	0.72	0.53
Model 4					
	Expected	$+2\sigma$	$+1\sigma$	-1σ	-2σ
Method 1	0.90	1.68	1.25	0.65	0.48
Method 2	0.90	1.64	1.24	0.66	0.50
Toys	0.90	1.63	1.23	0.65	0.49
Model 5					
	Expected	$+2\sigma$	$+1\sigma$	-1σ	-2σ
Method 1	0.59	1.10	0.82	0.43	0.32
Method 2	0.59	1.17	0.84	0.43	0.33
Toys	0.59	1.15	0.84	0.42	0.31
Model 6					
	Expected	$+2\sigma$	$+1\sigma$	-1σ	-2σ
Method 1	1.01	1.89	1.41	0.73	0.54
Method 2	1.01	2.08	1.47	0.71	0.52
Toys	1.01	2.07	1.47	0.71	0.52

Table 5.5: Quantiles of expected upper limits on μ for the two asymptotic methods in comparison with the full toy result.

CHAPTER VI

Search in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ Channel

"Our feeblest contemplations of the Cosmos stir us - there is a tingling in the spine, a catch in the voice, a faint sensation, as if a distant memory, of falling from a height. We know we are approaching the greatest of mysteries."

- Carl Sagan

6.1 Overview

The $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay mode features a relatively large branching fraction throughout the $m_H > 110$ GeV search region. It is characterized by two high p_T oppositely charged leptons and large missing transverse energy. The spin-0 nature of the Higgs yields a dilepton system with low $m_{\ell\ell}$ due to the small opening angle between the lepton pair. The primary production modes that contribute to this channel are gluon-gluon fusion and vector boson fusion (see Section 2.7). Between these two, the $gg \to H$ mode has a production cross section around 10 times that of VBF at low m_H for $\sqrt{s} = 8$ TeV. The $gg \to H$ production mode has zero jets at leading order, but large radiative corrections due to the gluons in the initial state lead to a large fraction of events with one high p_T jet. In contrast to this, the VBF production topology gives two high p_T forward jets at leading order, with large longitudinal separation between them. The different signal and background topologies that lead to different background levels and compositions of the 0-, 1-, and $\geq$2-jet final states

make it beneficial for this analysis to have dedicated search channels each. Further, the $gg \to H \to WW^{(*)}$ and $q\bar{q} \to Hq\bar{q} \to WW^{(*)}q\bar{q}$ modes differ from a physics standpoint, in that $gg \to H \to WW^{(*)}$ mode depends on the coupling of the Higgs to both fermions and bosons, while $q\bar{q} \to Hq\bar{q} \to WW^{(*)}q\bar{q}$ depends only on the coupling to bosons. The 2-jet inclusive analysis is therefore considered to be a dedicated search for the VBF production mechanism. Signal contributions from the associated production mode are also included and contribute between 1-5% depending on jet multiplicity.

The search is further divided by final state lepton flavor combinations. This is primarily because of the large $Z/\gamma^* \to \ell\ell$ background in the same flavor ($ee, \mu\mu$, denoted SF) final state, which is small in the different flavor ($e\mu, \mu e$, denoted DF) state. The DF channel is separated into $e\mu$ and μe based on the leading lepton due to different electron and muon performance at low p_T as well as differing rates of backgrounds due to fake leptons.

The two neutrinos in the final state prevents full reconstruction of the invariant mass of the Higgs. Because of this, the analysis is largely a number counting experiment, although $m_T = \sqrt{(E_T^{\ell\ell} + \not{E}_T)^2 - (\mathbf{p}_{\mathrm{T}}^{\ell\ell} + \vec{\not{E}}_T)^2}$, the transverse mass, can be used to give some small measure of sensitivity to m_H as well as additional separation between signal and background.

The bulk of this chapter represents the analysis on the 8 TeV dataset. Details specific to the 7 TeV analysis are specified in Section 6.12.

6.2 Data and Monte Carlo Samples

Table 6.1 shows the Monte Carlo (MC) samples used in this analysis along with the cross section times branching ratio ($\sigma \times \mathcal{B}$) for each physics process. The branching ratio assumes and sums over leptonic decays of W/Z bosons, with the exception of the top backgrounds and VH signal samples, which are inclusive. Some cross sections include generator level selections, which are mentioned in the comments.

The ggF signal corresponds to $m_H = 125$ GeV. The ggF cross section is computed at next-to-next-leading-order (NNLO) in QCD scale [37, 35, 70, 64, 25, 65] using the MSTW2008

Process	Generator	$\sigma \cdot \mathcal{B}$ (pb)
ggF	POWHEG+PYTHIA8	0.441
VBF	POWHEG+PYTHIA8	$35 \cdot 10^{-3}$
WH/ZH	PYTHIA8	$127 \cdot 10^{-3}$
$q\bar{q}/g \to WW$	POWHEG+PYTHIA6	5.68
$q\bar{q}/g \to WW{+}2j$	Sherpa, no QCD vertices	0.039
$gg \to WW$	GG2WW+HERWIG	0.16
$t\bar{t}$	MC@NLO+HERWIG	238.1
tW/tb	MC@NLO+HERWIG	28
tqb	AcerMC+PYTHIA6	88
inclusive W	ALPGEN+HERWIG	$37 \cdot 10^3$
inclusive Z/γ^*	ALPGEN+HERWIG	$16 \cdot 10^3$
$Z/\gamma^*{+}2j$	Sherpa, no QCD vertices	1.178
$Z^{(*)}Z^{(*)} \to 4l$	POWHEG+PYTHIA8	0.73
$W(Z/\gamma^*)(m_{(Z/\gamma^*)} > 7\text{GeV})$	POWHEG+PYTHIA8	0.825
$W(Z/\gamma^*)(m_{(Z/\gamma^*)} < 7\text{GeV})$	MADGRAPH+PYTHIA6	11.0
$W\gamma$	ALPGEN+HERWIG	369

Table 6.1: Monte Carlo generators used to model the signal and background processes in which all of the W and Z decay channels are included in the corresponding product of the cross section (σ) and branching fraction ($\mathcal{B}$) at $\sqrt{s} = 8\,\text{TeV}$. Masses are given in units of GeV.

PDF set [58]. Next-to-leading-order (NLO) electroweak and next-to-next-leading-log (NNLL) QCD corrections are applied [24, 23]. The VBF cross section is computed at NNLO and includes NLO QCD and NLO EW corrections. The VH cross sections are computed up to NNLO in QCD scale and include NLO EW corrections. The branching fractions were calculated with PROFECY4F [29, 28] and the total width was calculated with HDECAY [19].

Several generators are used to simulate the hard scatter (HS), parton shower (PS), hadronisation, and underlying event (UE). PYTHIA6 or PYTHIA8 is used for the signal and some background processes for simulating PS, hadronisation, and UE. In cases that HERWIG is used for hadronisation and PS, JIMMY is used for UE. SHERPA is used for HS and PS for VBF-like background processes such as $qq' \to Zqq'$ and $qq' \to W^+W^-qq'$. Full simulation of the ATLAS detector with GEANT4 is used for almost all processes, with the exception of the $qq/gg \to WW$ and single top backgrounds, which uses fast simulation to increase Monte

Carlo statistics. The CT10 parton density function (PDF) set is used for POWHEG and MC@NLO, while CTEQ6L1 is used for ALPGEN, MADGRAPH, PYTHIA6, and PYTHIA8.

The Wγ and Wγ^* NLO K-factors are computed with MCFM. For Wγ, K = 1.15. For Wγ^*, K = 2.01, which is computed on the phase space $0.5 < m_{\ell\ell} < 7$ GeV, $p_T^{\text{lead}} > 25$ GeV, $p_T^{\text{sub-lead}} > 15$ GeV, and $|\eta^\ell| < 2.8$.

The data collected uses inclusive single electron and muon triggers with a 24 GeV p_T threshold and loose isolation. This is 90% efficient for electrons and 90% (65%) efficient for muons in the end-cap (barrel) based on tag-and-probe studies using Z events in data, and was measured as a function of p_T, η, and data-taking period. The total integrated luminosity is 20.7 fb^{-1} after offline data-quality selection.

6.3 Object Selection

This section gives an overview of the selection of physics objects used in the analysis.

6.3.1 Trigger

The analysis uses unprescaled single lepton triggers. The ee channel uses an "or" between the EF_e60_medium1 and EF_e24vhi_medium1 triggers. The e60 refers to a 60 GeV trigger threshold. The higher p_T trigger is used to recover efficiency loss at high p_T. The "vh" in the e24 trigger refers to the fact that it uses η and p_T dependent thresholds, and includes a hadronic leakage cut at L1. The "i" indicates that the lepton must be isolated. The $\mu\mu$ channel uses an "or" between the EF_mu36_tight and EF_mu24i_tight triggers. Similar to electrons, the "i" indicates that the muon must be isolated. The $e\mu$ channel uses an "or" between all four triggers.

The per-lepton efficiency is calculated from MC and corrected for based on tag-and-probe studies in data. The per-event scale factor can be written as a function of this per-lepton

scale factor.

$$SF_{\text{Event}} = \frac{1 - (1 - \epsilon_{MC}^{lead} \times SF^{lead}) \times (1 - \epsilon_{MC}^{sub} \times SF^{sub})}{1 - (1 - \epsilon_{MC}^{lead}) \times (1 - \epsilon_{MC}^{sub})} \tag{6.1}$$

ϵ_{MC}^{lead} and ϵ_{MC}^{sub} are the per-lepton trigger efficiencies from MC for the leading and subleading leptons, respectively, and SF refers to the per-lepton scale factor determined from tag-and-probe.

6.3.2 Electrons

Electrons are selected with $p_T > 15$ GeV and $|\eta_{\text{cluster}}| < 2.47$, excluding the barrel/endcap transition region $1.37 < |\eta_{\text{cluster}}| < 1.52$. Cuts are applied to the calorimeter shower shape, track quality, track-cluster matching, and transition radiation energy criteria according to the ATLAS Tight++ identification menu. A cut is applied to the transverse impact parameter divided by its uncertainty, $\frac{d_0}{\sigma(d_0)} < 3$. The longitudinal impact parameter projected onto the beam axis is required to satisfy $|z_0 \sin\theta| < 0.4$mm. The electron is required to be isolated both in the sum of calorimeter cluster E_T and the sum of track p_T relative to the electron's own p_T in a cone of $\Delta R < 0.3$ around the electron. The calorimeter isolation requirement is topoEtcone30$/p_T = \sum_{cell}(E_T^{cell})/p_T < 0.16$, where the effects of pileup are estimated and subtracted from the numerator. The track isolation is optimized separately for two electron p_T regions, and is required to satisfy pTcone30$/p_T = \sum_{track}(p_T^{track})/p_T < 0.12$ for $15 < p_T < 25$ GeV and $\sum_{track}(p_T^{track})/p_T < 0.16$ for $p_T > 25$ GeV. Tracks with $p_T > 400$ MeV are included in the sum for the 8 TeV analysis, while this is increased to 900 MeV in the 7 TeV analysis. The isolation and impact parameter selections are primarily optimized to reject the large W+jets background. p_T dependent efficiency corrections are applied to the Monte Carlo based on Data/MC comparisons with a tag-and-probe method.

6.3.3 Muons

The analysis uses muons satisfying $p_T > 15$ GeV and $|\eta| < 2.5$. A track is required to be reconstructed in both the inner detector and muon spectrometer. The final muon four-momentum is taken from a combination of the two using the STACO algorithm. Muon combined performance quality cuts are required on the inner detector track as follows. The number of hits in the inner most tracking layer (B-Layer) are required to be at least one. The number of hits in the pixel detector are required to be at minimum one. The number of hits in the SCT are required to be at least five. The number of pixel and SCT holes associated with the track should be less than three, where a hole indicates a position in the detector where the track should have left a hit. Denote the number of hits in the TRT as n_{TRT}^{hits} and the number of outliers as $n_{TRT}^{outliers}$, with $n \equiv n_{TRT}^{hits} + n_{TRT}^{outliers}$. A TRT outlier can either be a straw tube that has a signal that is not crossed by a nearby track or a sequence of hits that do not fit well when combined with SCT and Pixel measurements. For $0.1 < |\eta| < 1.9$, $n > 5$ and $n_{TRT}^{outliers} < 0.9n$ is required. For $|\eta| \leq 0.1$ or $|\eta| \geq 1.9$, if $n > 5$ then $n_{TRT}^{outliers} < 0.9n$ is required.

Muons/jet overlap is treated as described in section 6.3.4. Similar to electrons, selections are made to calorimeter and track isolation relative to the muon's p_T, as well as to d_0 significance and $z_0 \sin\theta$, while data-driven p_T dependent corrections for these are applied to the MC efficiency with tag-and-probe. These selections are targeted to reject non-prompt and W+jets background. The impact parameter requirements are $\frac{d_0}{\sigma(d_0)} < 3$ and $|z_0 \sin\theta| <$ 1.0mm. The calorimeter isolation criteria is EtConeCor30/$p_T < 0.014 p_T(\text{GeV}) - 0.15$ and EtConeCor30/$p_T < 0.20$. The calorimeter isolation variable is corrected for its dependence on the number of primary vertices in the event to account for high pileup [75]. The track isolation requirement is PtCone30/$p_T < 0.01 p_T(\text{GeV}) - 0.105$ and PtCone30/$p_T < 0.15$.

6.3.4 Jets

Jets are reconstructed using the Anti-k_T algorithm with a distance parameter R=0.4 [30] and are required to fall within $|\eta| < 4.5$. Jet cleaning is applied according to the criteria proposed by the jet performance group [1]. The p_T threshold of the jet is 25 GeV for $|\eta| < 2.5$ and 30 GeV for $2.5 < |\eta| < 4.5$. Jets overlapping muons are removed from the set of selected jets. To suppress pileup jets, the jet vertex fraction (JVF) [59] is required to satisfy $|JVF| > 0.5$, where JVF is defined as the fraction of the sum of track momentum in the jet associated with the primary vertex, where all tracks with $p_T > 400$ MeV are considered. This is defined to be -1 for jets falling outside tracking coverage, hence the absolute value effectively removes the selection where no tracks can be reconstructed. The asymmetric p_T threshold is designed to reduce the effect of pileup in the forward region. Using a $Z \to \mu\mu$ enriched data sample, the JVF cut is optimized to reject pileup jets while maintaining a high efficiency for non-pileup jets.

6.3.4.1 b-Tagging

The b-tagging algorithm used in the analysis is the MV1 algorithm with an 85% working point. MV1 is the output of a neural network that uses JetFitter+IP3D, IP3D, and SV1 as inputs. The details of these three inputs can be found in Section 4.4.3. This is used in both 1- and 2-jet analyses to suppress top events and to define the top control regions.

6.3.5 Missing Transverse Energy

The base for the missing transverse energy used in the analysis is $\not{E}_T^{\text{ref,final}}$. This is a prioritized sum of cells from final state reconstructed objects using their corresponding calibrations, and of cells not associated with reconstructed objects. The latter is called the Cell Out Energy Flow (COEF) missing transverse energy.

$$\not{E}_{x(y)}^{\text{ref,final}} = \not{E}_{x(y)}^{e} + \not{E}_{x(y)}^{\gamma} + \not{E}_{x(y)}^{\tau} + \not{E}_{x(y)}^{\text{jets}} + \not{E}_{x(y)}^{COEF} + \not{E}_{x(y)}^{\mu} \tag{6.2}$$

Cells within $|\eta| < 4.9$ are considered. The order of the sum in the equation represents the priority given to calibration, with highest priority starting from the left. Muons are considered out to $|\eta| < 2.7$. Corrections to the muon term are added for muons reconstructed as tracks in the inner detector but not in the MS. Low p_T tracks that do not make it to the calorimeter are also re-added.

In the 0- and 1-jet analyses, $\not{E}_{T\,\mathrm{rel}}$ is used in place of $\not{E}_T$.

$$\not{E}_{T\,\mathrm{rel}} = \begin{cases} \not{E}_T, & \Delta\phi \geq \pi/2 \\ \not{E}_T \sin\Delta\phi, & \Delta\phi < \pi/2 \end{cases} \tag{6.3}$$

$\Delta\phi$ in this case is the absolute value of the azimuthal angle between the $\not{E}_T$ direction and the nearest lepton or jet. This is designed to reduce fake $\not{E}_T$ induced by the calorimeter's intrinsic energy resolution.

In the 2-jet analysis, a third $\not{E}_T$ ($\not{E}_T{}^{\mathrm{STVF}}$) is introduced to better handle $\not{E}_T$ induced by pileup jets. The $\not{E}_T{}^{\mathrm{jets}}$ term only uses jets with $|JVF| > 0$. The Soft Term Vertex Fraction (STVF) weights the $\not{E}_T{}^{\mathrm{COEF}}$ term. STVF is the ratio of the sum of p_T of tracks coming from the first PV but unmatched to reconstructed physics objects to all tracks unmatched to reconstructed physics objects.

$$STVF = \sum_{\mathrm{PV\ tracks}} p_T / \sum_{\mathrm{tracks}} p_T \tag{6.4}$$

To qualify as coming from the first PV, tracks in the numerator must satisfy $|d_0| < 1.5$mm and $|z_0 \sin\theta| < 1.5$mm.

Lastly, the SF channels use track based missing transverse momentum, $\not{p}_T$, to further reject Z/γ^*. Tracks must satisfy $p_T > 500$ MeV, $|\eta| < 2.5$, $|d_0| < 1.5$mm, $|z_0 \sin\theta| < 1.5$mm, and must have at least one (six) hits in the Pixel (SCT) detector. Lepton tracks that fail the track selections but pass the criteria to be used in the standard $\not{E}_T$ are also used. All electrons passing the medium++ quality criteria with $|\eta| < 2.47$ are considered. All STACO

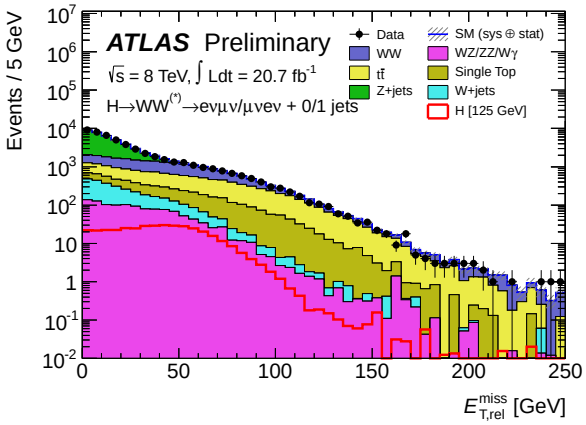
combined muons are also used. If the lepton is an electron, the cluster energy is used instead of its track momentum.

6.4 Event Preselection

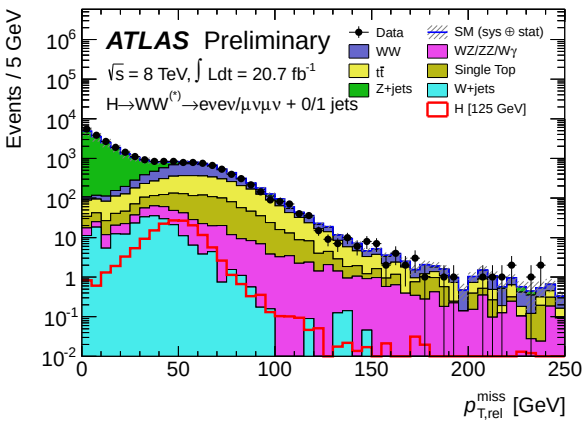
Before separation by jet multiplicity, a common preselection is applied to reduce the major backgrounds. Events are required to have a primary vertex consistent with the beamspot position, with at least three associated inner detector tracks with $p_T > 400$ MeV. Event cleaning is applied to reject cosmic rays, beam background, and calorimeter noise. Exactly two oppositely charged leptons (electrons or muons) are required that satisfy $p_T^{\text{subleading}} > 15$ GeV and $p_T^{\text{leading}} > 25$ GeV. The dilepton invariant mass is required to be at least 10 GeV for DF and 12 GeV for SF. For SF, $m_{\ell\ell}$ must also fall outside of a ± 15 GeV window around the Z-pole. In the 0- and 1-jet analysis $\not{E}_{T\,\text{rel}} > 25$ GeV is applied to DF, while for SF this is increased to 45 GeV to further supress Drell-Yan background. For the 2-jet analysis $\not{E}_T$ is used rather than $\not{E}_{T\,\text{rel}}$ due to the high jet multiplicity and is required to be at least 20 GeV in DF and 45 GeV in SF. In SF, $\not{E}_T{}^{\text{STVF}}$ is also required to be larger than 35 GeV in the 2-jet bin. For SF, f_{recoil} (see Section 6.9) should be greater than 0.05 in 0-jet, while $f_{\text{recoil}}^{\text{extended}}$ should be greater than 0.2 in 1-jet. Figure 6.1 shows the distributions for the different $\not{E}_T$ definitions in various jet multiplicities. Also shown is f_{recoil} in the 0-jet SF channel.

After all preselection cuts, the jet multiplicity is shown in Figure 6.2 for DF and SF separately. At this stage the DF channel is dominated by dileptonic top background, while SF has similar contributions from both top and $Z \to \ell\ell$.

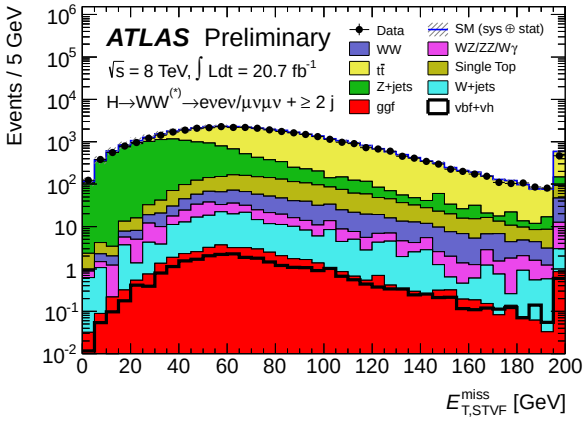
At this point, events are further divided by jet multiplicity. Different topological selections are applied to each multiplicity to target the suppression of different backgrounds and while retaining signal efficiency under dissimilar event topologies. Table 6.2 compactly summarizes the signal region event selection for all jet multiplicities and lepton flavor categories. These are elaborated on in the following sections.



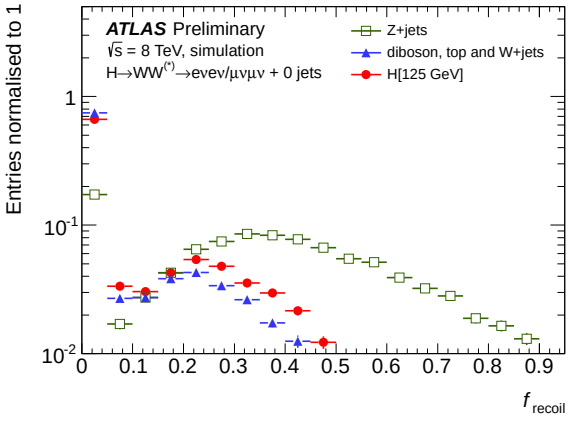
(a) $\not\!\!E_{T\,\mathrm{rel}}$ for $N_{\mathrm{jet}} \leq 1$, $e\mu$+μe



(b) $p_{\mathrm{T,rel}}^{\mathrm{miss}}$ for $N_{\mathrm{jet}} \leq 1$, ee+$\mu\mu$

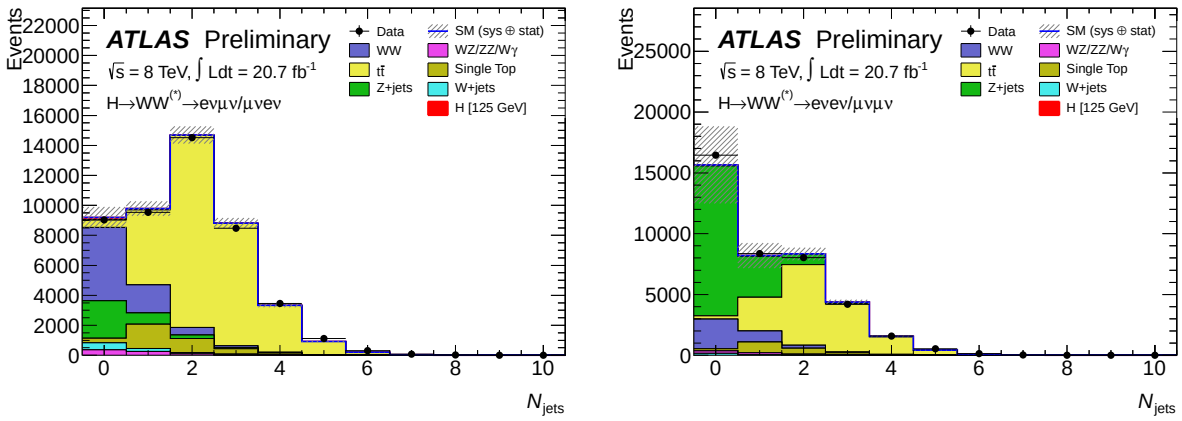


(c) $\not\!\!E_T{}^{\mathrm{STVF}}$ for $N_{\mathrm{jet}} \geq 2$, ee+$\mu\mu$



(d) f_{recoil} for N_{jet}=0, ee+$\mu\mu$

Figure 6.1: Missing transverse momentum distributions for events after pre-selection for (a) $\not\!\!E_{T\,\mathrm{rel}}$ for $N_{\mathrm{jet}} \leq 1$, (b) $p_{\mathrm{T,rel}}^{\mathrm{miss}}$ for $N_{\mathrm{jet}} \leq 1$, and (c) $\not\!\!E_T{}^{\mathrm{STVF}}$ for $N_{\mathrm{jet}} \geq 2$ modes. The plot in (b) is made after the requirement on $\not\!\!E_{T\,\mathrm{rel}}$ and the one in (c) after the requirement on $\not\!\!E_T$. The plot in (d) shows the f_{recoil} distribution in ee+$\mu\mu$ events passing the N_{jet}=0 selection after $m_{\ell\ell} < 50\,\mathrm{GeV}$ for simulated DY, non-DY and signal processes. The shaded area (too small to be visible in these figures) represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources. The signal is overlaid as a red curve in (a) and (b); in (c), the ggF signal is stacked at the bottom while the VBF signal is overlaid as a thick black line.



(a) N_{jet} distribution for $e\mu$+μe

(b) N_{jet} distribution for ee+$\mu\mu$

Figure 6.2: Jet multiplicity for events in 8 TeV data. The plots are shown for the (a) $e\mu$+μe and (b) ee+$\mu\mu$ channels after pre-selection and $\not\!\!E_{T\,\text{rel}} > 25$ GeV and > 45 GeV, respectively. The signal is too small to be seen, but falls mostly in the 0-jet bin. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

6.5 0-jet Analysis

Events with zero selected jets must satisfy further topological selections. The opening angle between the dilepton system and the $\not\!\!E_T$ direction, $\Delta\phi_{\ell\ell,\not\!\!E_T}$, must be greater than $\frac{\pi}{2}$. Further selections are $p_{T,\ell\ell} > 30$ GeV, $m_{\ell\ell} < 50$ GeV, and $\Delta\phi_{\ell\ell} < 1.8$ radians. The transverse mass is fit rather than cut on to obtain the final result, but a cut of $0.75 \times m_H < m_T < m_H$ is considered to give a representation of the event counts in the signal rich region of m_T. Table 6.3 shows the expected signal and background events along with the number observed in data after each cut in the SF and DF channels in this jet category.

Figure 6.3 shows the $m_{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ distributions in the DF channel before the $m_{\ell\ell}$ selections. Figure 6.4 shows the transverse mass for both DF and SF channels after all selections.

6.6 1-jet Analysis

In addition to the $m_{\ell\ell}$ and $\Delta\phi$ selections of the 0-jet analysis, the 1-jet analysis applies additional cuts to suppress the top and Z$\to\tau\tau$ backgrounds. Events where the jet is tagged

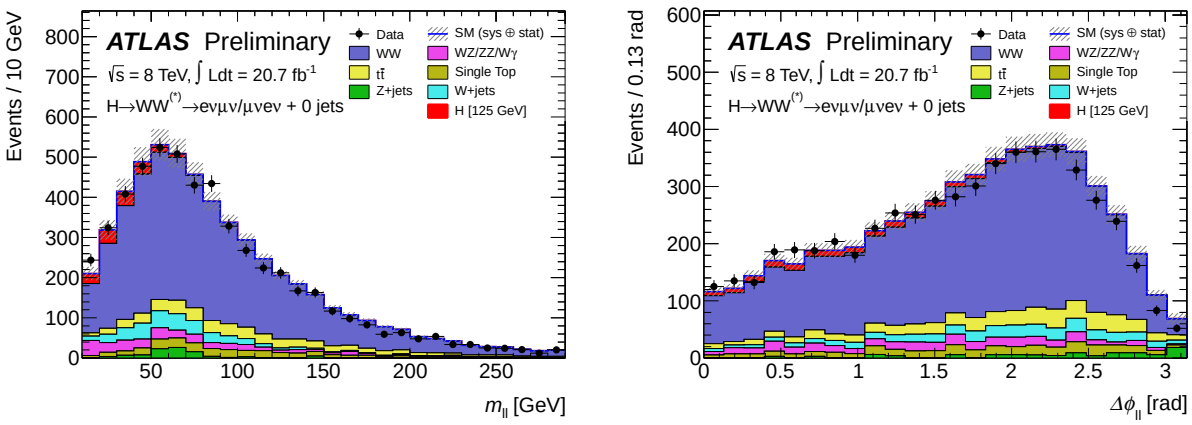


Figure 6.3: Kinematic distributions in the N_{jet}=0 channel: $m_{\ell\ell}$ (left) and $\Delta\phi_{\ell\ell}$ (right) after the cut on $p_{\text{T}}^{\ell\ell}$. The signal is added on top of the background. The WW and top backgrounds are scaled to use the normalisation derived from the corresponding control regions described in the text. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

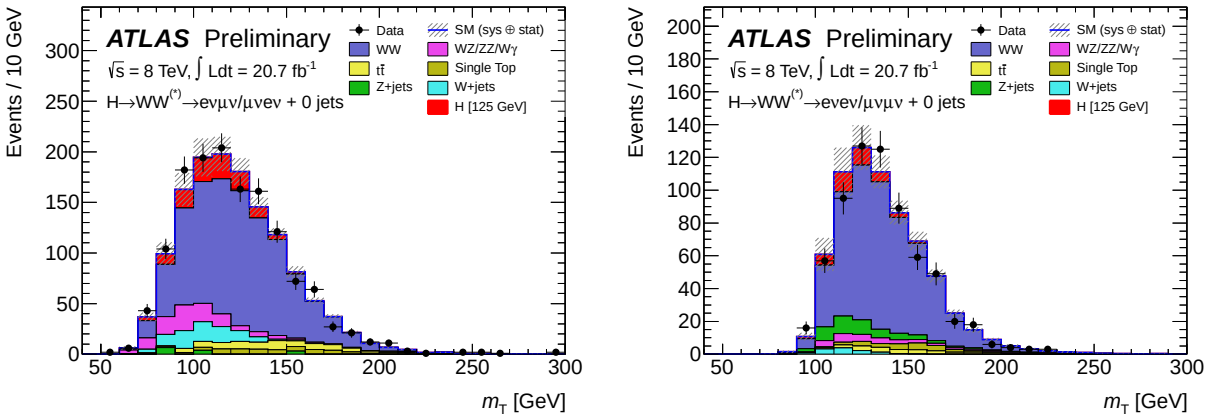


Figure 6.4: Distribution of the transverse mass, m_{T}, for 8 TeV data. The plots are shown for the $e\mu$+μe (left) and ee+$\mu\mu$ (right) channels in the N_{jet}=0 mode. The distributions are shown prior to splitting the samples into two $m_{\ell\ell}$ regions for the $e\mu$+μe channel. The visible signal is stacked at the top of the background. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

Category	N_{jet}=0	N_{jet}=1	$N_{\text{jet}}\geq$2
Pre-selection	Two isolated leptons ($\ell = e, \mu$) with opposite charge Leptons with ${p_T}^{\text{lead}} > 25$ and ${p_T}^{\text{sublead}} > 15$ $e\mu$+μe: $m_{\ell\ell} > 10$ ee+$\mu\mu$: $m_{\ell\ell} > 12$, $\mid m_{\ell\ell} - m_Z \mid > 15$		
Missing transverse momentum and hadronic recoil	$e\mu$+μe: $\not{E}_{T\,\text{rel}} > 25$ ee+$\mu\mu$: $\not{E}_{T\,\text{rel}} > 45$ ee+$\mu\mu$: $p_{\text{T,rel}}^{\text{miss}} > 45$ ee+$\mu\mu$: $f_{\text{recoil}} < 0.05$	$e\mu$+μe: $\not{E}_{T\,\text{rel}} > 25$ ee+$\mu\mu$: $\not{E}_{T\,\text{rel}} > 45$ ee+$\mu\mu$: $p_{\text{T,rel}}^{\text{miss}} > 45$ ee+$\mu\mu$: $f_{\text{recoil}} < 0.2$	$e\mu$+μe: $\not{E}_T > 20$ ee+$\mu\mu$: $\not{E}_T > 45$ ee+$\mu\mu$: ${\not{E}_T}^{\text{STVF}} > 35$ -
General selection	- $\Delta\phi_{\ell\ell,\not{E}_T} > \pi/2$ $p_{\text{T}}^{\ell\ell} > 30$	$N_{b\text{-jet}} = 0$ - $e\mu$+μe: $Z/\gamma^* \to \tau\tau$ veto	$N_{b\text{-jet}} = 0$ $p_{\text{T}}^{\text{tot}} < 45$ $e\mu$+μe: $Z/\gamma^* \to \tau\tau$ veto
VBF topology	- - - -	- - - -	$m_{jj} > 500$ $\Delta y_{jj} > 2.8$ No jets ($p_T > 20$) in rapidity gap Require both ℓ in rapidity gap
$H \to WW^{(*)} \to \ell\nu\ell\nu$ topology	$m_{\ell\ell} < 50$ $\Delta\phi_{\ell\ell} < 1.8$ $e\mu$+μe: split $m_{\ell\ell}$ Fit m_{T}	$m_{\ell\ell} < 50$ $\Delta\phi_{\ell\ell} < 1.8$ $e\mu$+μe: split $m_{\ell\ell}$ Fit m_{T}	$m_{\ell\ell} < 60$ $\Delta\phi_{\ell\ell} < 1.8$ - Fit m_{T}

Table 6.2: Selection listing for 8 TeV data. The criteria specific to $e\mu$+μe and ee+$\mu\mu$ are noted as such; otherwise, they apply to both. Pre-selection applies to all N_{jet} modes. The rapidity gap is the y range spanned by the two leading jets. The $m_{\ell\ell}$ split is at 30 GeV. The modifications for the 7 TeV analysis are given in Section 6.12 and are not listed here. Energies, masses, and momenta are in units of GeV.

as a b-quark are vetoed. The $\tau\tau$ invariant mass $m_{\tau\tau}$ is computed assuming the neutrinos from the leptonic tau are collinear with the τ decays [68]. If the fraction of energy carried by the visible decay products for the two taus, $x_{\tau 1}$ and $x_{\tau 2}$, are positive, then $m_{\tau\tau}$ must fall outside of a $\pm$25 GeV window around the Z-pole. Table 6.4 shows the expected signal and background events along with the number observed in data after each cut for the SF and DF channels in this jet category.

Figure 6.5 shows the $m_{\ell\ell}$ distribution after the $Z \to \tau\tau$ veto and the $\Delta\phi_{\ell\ell}$ distribution after the $m_{\ell\ell}$ cut in the DF channel. Figure 6.6 shows the transverse mass for both DF and SF channels after all selections.

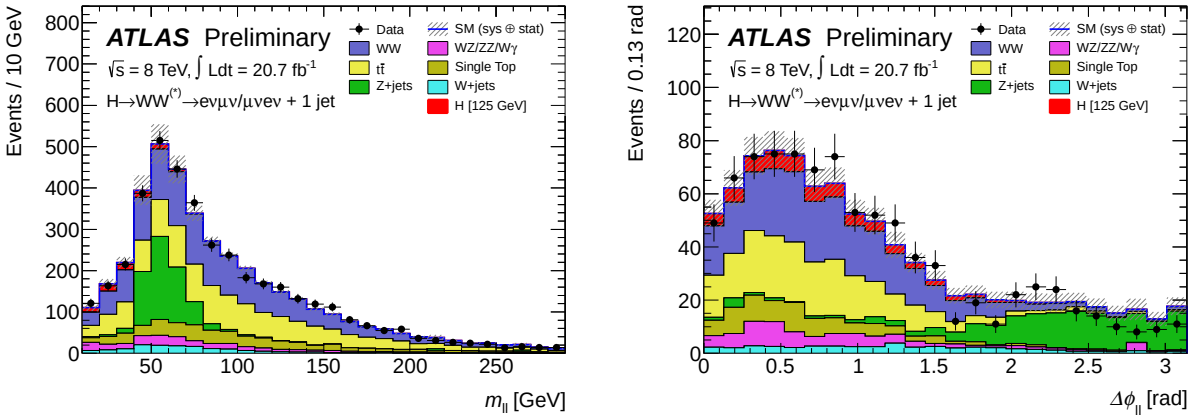


Figure 6.5: Kinematic distributions in the N_{jet}=1 channel: $m_{\ell\ell}$ after the $Z\to\tau\tau$ veto (left) and $\Delta\phi_{\ell\ell}$ after the cut on $m_{\ell\ell}$ (right), The signal is added on top of the background. The WW and top backgrounds are scaled to use the normalisation derived from the corresponding control regions described in the text. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

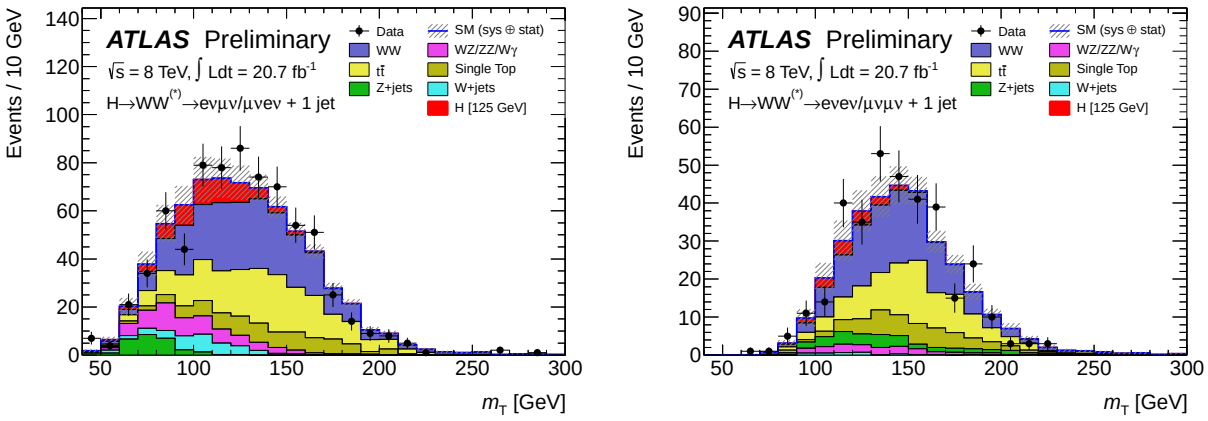


Figure 6.6: Distribution of the transverse mass, m_{T}, for 8 TeV data. The plots are shown for the $e\mu$+μe (left) and ee+$\mu\mu$ (right) channels in the N_{jet}=1 mode. More details are given in Figure 6.4

(a) $e\mu$+μe channel

Selection	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
N_{jet}=0	9024	9000 ± 40	172 ± 2	4900 ± 20	370 ± 10	510 ± 10	310 ± 10	2440 ± 30	470 ± 10
$\Delta\phi_{\ell\ell,\not{E}_T} > \frac{\pi}{2}$	8100	8120 ± 40	170 ± 2	4840 ± 20	360 ± 10	490 ± 10	310 ± 10	1690 ± 30	440 ± 10
$p_{\text{T}}^{\ell\ell} > 30$	5497	5490 ± 30	156 ± 2	4050 ± 20	290 ± 10	450 ± 10	280 ± 10	100 ± 10	320 ± 5
$m_{\ell\ell} < 50$	1453	1310 ± 10	124 ± 1	960 ± 10	110 ± 6	69 ± 3	46 ± 3	18 ± 7	100 ± 2
$\Delta\phi_{\ell\ell} < 1.8$	1399	1240 ± 10	119 ± 1	930 ± 10	107 ± 6	67 ± 3	44 ± 3	13 ± 7	88 ± 2

(b) ee+$\mu\mu$ channel

Selection	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
N_{jet}=0	16446	15600 ± 200	104 ± 1	2440 ± 10	190 ± 5	280 ± 6	175 ± 6	12300 ± 160	170 ± 10
$\Delta\phi_{\ell\ell,\not{E}_T} > \frac{\pi}{2}$	13697	12970 ± 140	103 ± 1	2430 ± 10	190 ± 5	280 ± 6	174 ± 6	9740 ± 140	160 ± 10
$p_{\text{T}}^{\ell\ell} > 30$	5670	5650 ± 70	99 ± 1	2300 ± 10	170 ± 5	260 ± 6	167 ± 5	2610 ± 70	134 ± 4
$m_{\ell\ell} < 50$	2314	2390 ± 20	84 ± 1	760 ± 10	64 ± 3	53 ± 3	42 ± 3	1410 ± 20	62 ± 3
$p_{\text{T,rel}}^{\text{miss}} > 45$	1032	993 ± 10	63 ± 1	650 ± 10	42 ± 2	47 ± 3	39 ± 3	200 ± 5	19 ± 2
$\Delta\phi_{\ell\ell} < 1.8$	1026	983 ± 10	63 ± 1	640 ± 10	41 ± 2	46 ± 3	39 ± 3	195 ± 5	18 ± 2
$f_{\text{recoil}} < 0.05$	671	647 ± 7	42 ± 1	520 ± 10	30 ± 2	19 ± 2	22 ± 2	49 ± 3	12 ± 1

Table 6.3: Selection table for N_{jet}=0 in 8 TeV data. The observed (N_{obs}) and expected (N_{exp}) yields for the signal (N_{sig}) and background (N_{bkg}) processes are shown for the (a) $e\mu$+μe and (b) ee+$\mu\mu$ channels. The composition of N_{bkg} is given on the right. The requirements are imposed sequentially from top to bottom. Energies, masses, and momenta are in units of GeV. The uncertainties shown are only those due to limited MC statistics.

(a) $e\mu$+μe channel

Selection	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
N_{jet}=1	9527	9460 ± 40	97 ± 1	1660 ± 10	270 ± 10	4980 ± 30	1600 ± 20	760 ± 20	195 ± 5
$N_{b\text{-jet}} = 0$	4320	4240 ± 30	85 ± 1	1460 ± 10	220 ± 10	1270 ± 10	460 ± 10	670 ± 10	160 ± 4
$Z \to \tau\tau$ veto	4138	4020 ± 30	84 ± 1	1420 ± 10	220 ± 10	1220 ± 10	440 ± 10	580 ± 10	155 ± 4
$m_{\ell\ell} < 50$	886	830 ± 10	63 ± 1	270 ± 4	69 ± 5	216 ± 6	80 ± 4	149 ± 5	46 ± 2
$\Delta\phi_{\ell\ell} < 1.8$	728	650 ± 10	59 ± 1	250 ± 4	60 ± 4	204 ± 6	76 ± 4	28 ± 3	34 ± 2

(b) ee+$\mu\mu$ channel

Selection	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
N_{jet}=1	8354	8120 ± 90	54 ± 1	820 ± 10	140 ± 10	2740 ± 20	890 ± 10	3470 ± 80	60 ± 10
$N_{b\text{-jet}} = 0$	5192	4800 ± 80	48 ± 1	720 ± 10	120 ± 10	720 ± 10	260 ± 10	2940 ± 70	40 ± 10
$m_{\ell\ell} < 50$	1773	1540 ± 20	38 ± 1	195 ± 4	35 ± 2	166 ± 5	65 ± 3	1060 ± 10	20 ± 2
$p_{\text{T,rel}}^{\text{miss}} > 45$	440	420 ± 10	21 ± 1	148 ± 3	21 ± 1	128 ± 5	52 ± 3	64 ± 4	5.1 ± 0.8
$\Delta\phi_{\ell\ell} < 1.8$	430	410 ± 10	20 ± 1	143 ± 3	20 ± 1	125 ± 5	51 ± 3	63 ± 4	4.5 ± 0.7
$f_{\text{recoil}} < 0.2$	346	320 ± 10	16 ± 1	128 ± 3	17 ± 1	97 ± 4	44 ± 3	25 ± 2	3.1 ± 0.6

Table 6.4: Selection table for N_{jet}=1 in 8 TeV data. More details are given in the caption of Table 6.3.

6.7 2-jet Analysis

The 2-jet selections are geared to select signal events from the VBF production mechanism. The selections were optimized treating the ggF process as a background to further enhance the VBF signal above all other physics processes. The event must contain at least two selected jets. A $Z \to \tau\tau$ and b-jet veto are made as in the 1-jet analysis. To further suppresses the top background, $p_{\mathrm{T}}^{\mathrm{tot}} = |\mathbf{p}_{\mathrm{T}}^{\ell\ell} + \mathbf{p}_{\mathrm{T}}^{jj} + \not\mathbf{E}_T|$ must be less than 45 GeV, where $\mathbf{p}_{\mathrm{T}}^{jj}$ only contains contributions from the tag jets. Further selections are made on the two jets with the highest p_T. The dijet invariant mass m_{jj} is required to be larger than 500 GeV. The rapidity gap between the jets must satisfy $|\Delta y_{jj}| > 2.8$. The top background will tend to have more soft gluon radiation than the VBF signal, which mostly falls into the central region of the detector. To suppress the top background events are therefore vetoed if they contain additional selected jets with $p_T > 20$ GeV between the two highest p_T jets. Events with leptons outside of the two tag jets in rapidity are also vetoed. $m_{\ell\ell} < 60$ GeV and $\Delta\phi < 1.8$ is finally required. Table 6.5 shows the expected signal and background events along with the number observed in data after each cut in the SF and DF channels in this jet category.

Figure 6.7 shows the $|\Delta y_{jj}|$ distribution after the $\mathbf{p}_{\mathrm{T}}^{\mathrm{tot}}$ selection as well as m_{jj} after the cut on $|\Delta y_{jj}|$ for both DF and SF channels. Figure 6.8 shows the $m_{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ distributions after the outside lepton veto. Finally, figure 6.9 shows the transverse mass distribution for DF and SF after all selections.

6.8 Control Regions

Background rich control regions (CRs) close in phase space to the signal region (SR) are used to normalize the main backgrounds. There are seven control regions used in total. There are 0- and 1-jet control regions for the WW background, 1- and 2-jet control regions for the top background, and 0-, 1-, and 2-jet control regions for $Z \to \tau\tau$. Nearly all of the

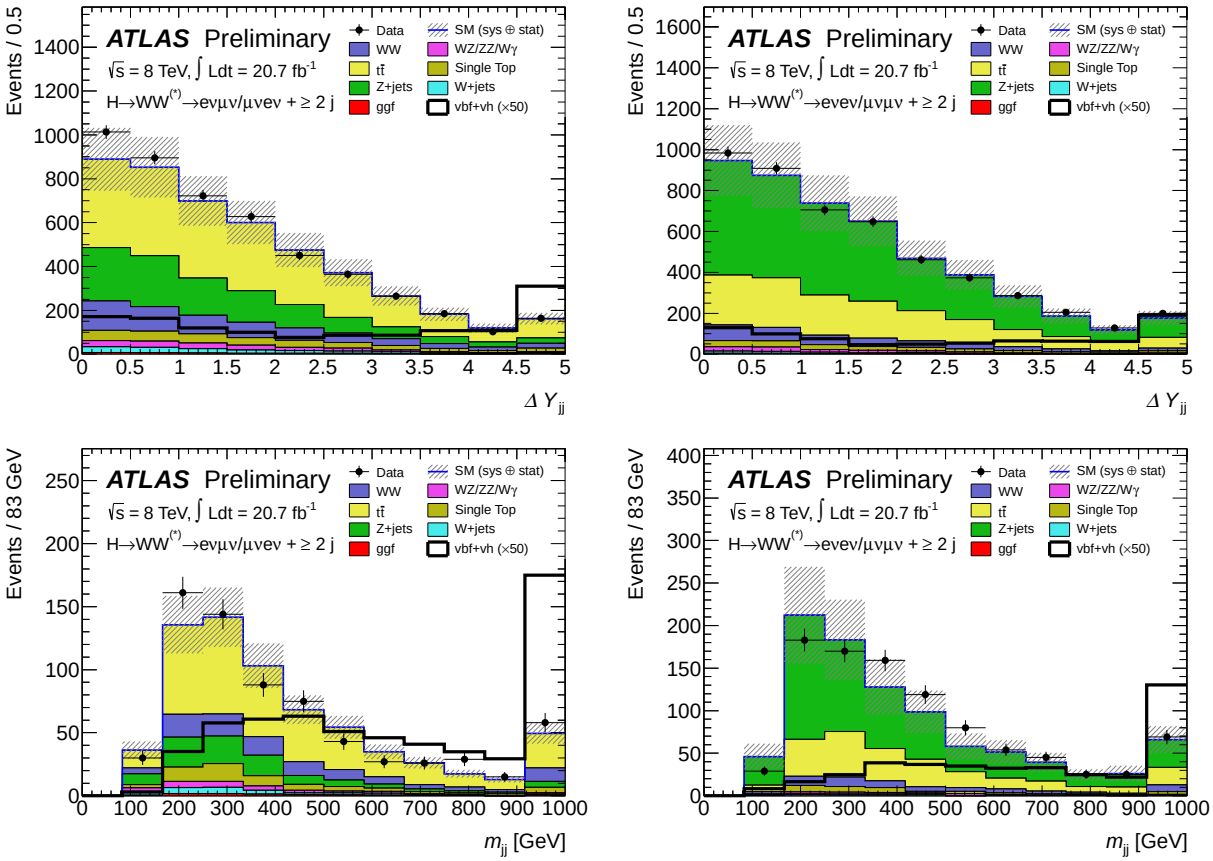


Figure 6.7: The Δy_{jj} and m_{jj} distributions after the $p_{\mathrm{T}}^{\mathrm{tot}} < 45$ GeV cut. $p_{\mathrm{T}}^{\mathrm{tot}}$ is defined as the total transverse momentum of all leptons, jets and missing E_T passing the selection. The m_{jj} distribution is shown after the $\Delta y_{jj} > 2.8$ cut. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

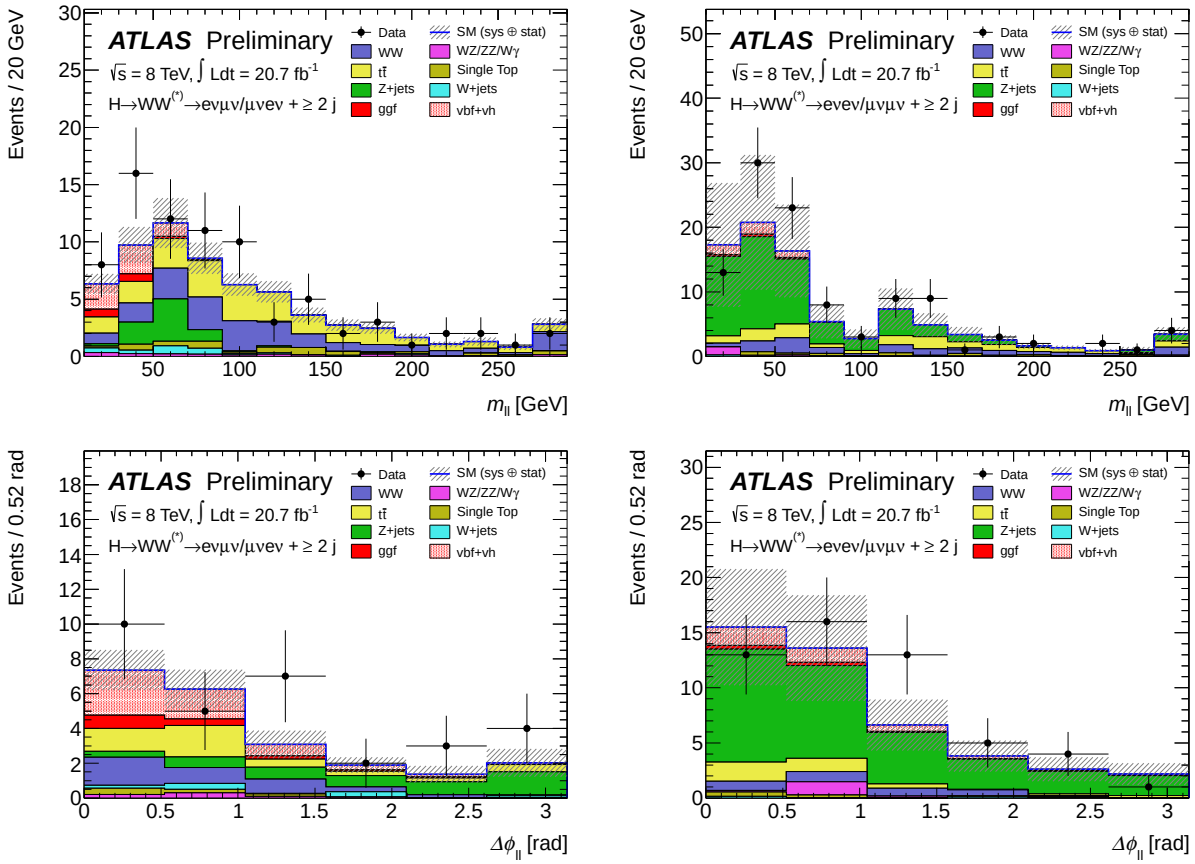


Figure 6.8: The $m_{\ell\ell}$ and the $\Delta\phi_{\ell\ell}$ distributions after the outside lepton veto cut, accepting events with leptons between the two tagging jets. $\Delta\phi_{\ell\ell}$ is shown after the $m_{\ell\ell} < 60$ GeV cut. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

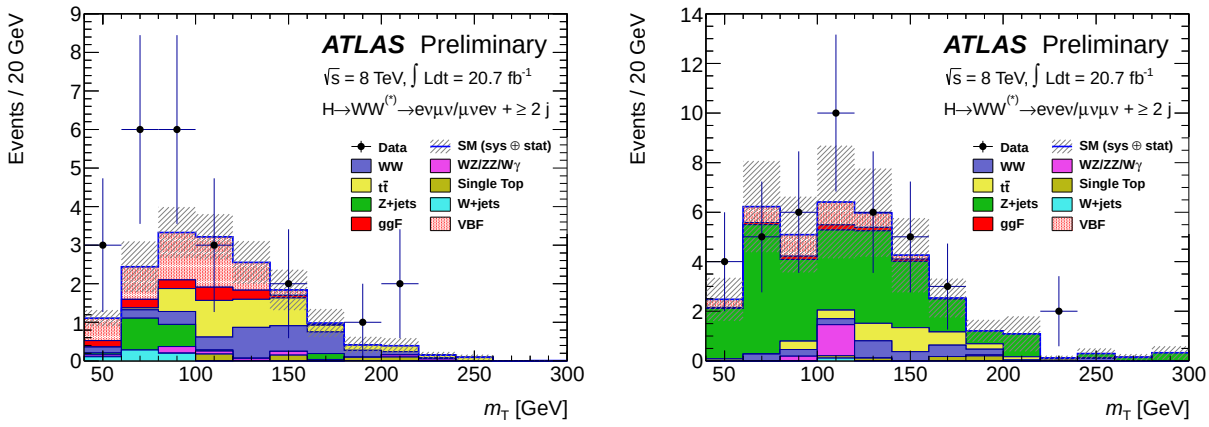


Figure 6.9: Distribution of the transverse mass, m_{T}, for 8 TeV data. The plots are shown for the $e\mu$+μe (left) and ee+$\mu\mu$ (right) channels in the $N_{\mathrm{jet}} \geq 2$ mode. More details are given in Figure 6.4

(a) $e\mu$+μe channel

Selection	N_{obs}	N_{bkg}	$N_{\text{sig,VBF}}$	$N_{\text{sig,ggF}}$	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
$N_{\text{jet}}\geq 2$	48723	47740 ± 80	43 ± 1	67 ± 1	940 ± 10	300 ± 20	41800 ± 70	2370 ± 20	1800 ± 30	440 ± 10
$N_{b\text{-jet}} = 0$	5852	5690 ± 30	31 ± 1	49 ± 1	690 ± 10	200 ± 10	2930 ± 20	350 ± 10	1300 ± 20	171 ± 5
$p_{\text{T}}^{\text{tot}} < 45$	4790	4620 ± 30	27 ± 1	41 ± 1	590 ± 10	160 ± 10	2320 ± 20	290 ± 10	1100 ± 20	126 ± 4
$Z\to\tau\tau$ veto	4007	3840 ± 30	25 ± 1	38 ± 1	540 ± 10	140 ± 10	2150 ± 20	260 ± 10	600 ± 20	108 ± 4
$\Delta y_{jj} > 2.8$	696	680 ± 10	12 ± 0.2	9.5 ± 0.3	100 ± 2	25 ± 3	380 ± 10	55 ± 3	95 ± 5	19 ± 2
$m_{jj} > 500$	198	170 ± 4	7.5 ± 0.1	2.9 ± 0.2	34 ± 1	5.6 ± 0.6	93 ± 3	11 ± 1	19 ± 2	4.4 ± 0.7
No jets in y gap	92	77 ± 2	6.3 ± 0.1	1.7 ± 0.2	25 ± 1	2.8 ± 0.4	30 ± 2	5.2 ± 0.8	9 ± 1	3.1 ± 0.6
Both ℓ in y gap	78	59 ± 2	6.1 ± 0.1	1.6 ± 0.1	19 ± 1	2.1 ± 0.3	22 ± 1	4.3 ± 0.7	7 ± 1	2.4 ± 0.5
$m_{\ell\ell} < 60$	31	16 ± 1	5.5 ± 0.1	1.5 ± 0.1	3.8 ± 0.4	0.7 ± 0.2	4.5 ± 0.7	0.7 ± 0.3	4.4 ± 0.8	1.0 ± 0.4
$\Delta\phi_{\ell\ell} < 1.8$	23	12 ± 1	5.1 ± 0.1	1.3 ± 0.1	3.5 ± 0.4	0.6 ± 0.2	3.7 ± 0.7	0.7 ± 0.3	1.9 ± 0.5	0.6 ± 0.3

(b) ee+$\mu\mu$ channel

Selection	N_{obs}	N_{bkg}	$N_{\text{sig,VBF}}$	$N_{\text{sig,ggF}}$	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
$N_{\text{jet}}\geq 2$	32877	32300 ± 100	26 ± 0.7	40 ± 1	540 ± 6	180 ± 10	24540 ± 60	1390 ± 20	5420 ± 90	190 ± 10
$N_{b\text{-jet}} = 0$	65388	6370 ± 80	19 ± 0.6	30 ± 1	390 ± 5	130 ± 10	1750 ± 20	200 ± 10	3810 ± 80	58 ± 4
$p_{\text{T}}^{\text{tot}} < 45$	4903	4830 ± 70	17 ± 0.5	24 ± 1	340 ± 4	92 ± 5	1370 ± 10	170 ± 10	2790 ± 70	43 ± 3
$\Delta y_{jj} > 2.8$	958	930 ± 30	8.1 ± 0.2	6.2 ± 0.3	61 ± 2	12 ± 1.3	252 ± 6	35 ± 2	560 ± 30	6 ± 1
$m_{jj} > 500$	298	245 ± 6	5.5 ± 0.1	2.1 ± 0.2	23 ± 1	4.1 ± 1.1	62 ± 3	9 ± 1	142 ± 5	1.4 ± 0.6
No jets in y gap	147	119 ± 4	4.7 ± 0.1	1.1 ± 0.1	17 ± 1	2.8 ± 1.1	19 ± 1	4.1 ± 0.7	74 ± 3	0.7 ± 0.4
Both ℓ in y gap	108	85 ± 3	4.5 ± 0.1	0.9 ± 0.1	12 ± 1	2.3 ± 1.1	14 ± 1	3.1 ± 0.6	51 ± 3	0.3 ± 0.3
$m_{\ell\ell} < 60$	52	40 ± 2	4.0 ± 0.1	0.8 ± 0.1	3.2 ± 0.3	1.6 ± 1.1	3.7 ± 0.6	0.8 ± 0.3	30 ± 2	0.1 ± 0.2
$\Delta\phi_{\ell\ell} < 1.8$	42	34 ± 2	3.7 ± 0.1	0.7 ± 0.1	2.8 ± 0.3	1.6 ± 1.1	3.3 ± 0.5	0.7 ± 0.3	25 ± 2	0.1 ± 0.2

Table 6.5: Selection table for $N_{\text{jet}}\geq 2$ in 8 TeV data. More details are given in the caption of Table 6.3. In this table, the $N_{\text{sig,ggF}}$ is included in N_{bkg}; the $N_{\text{sig},VH}$ is included in $N_{\text{sig,VBF}}$, but the contributions are negligible after the VBF-related criteria. The y gap is described in Table 6.2.

control regions mentioned here use only events from the $e/\mu + \mu/e$ channels, where both DF channels have been summed together. The exception is the 2-jet top CR, where both DF and SF are summed to increase statistics. Each control region is therefore a single bin. Associated with each control region is an independent normalization parameter for the target background. There is cross-talk between the backgrounds in the control regions such that the normalization parameters must be fit simultaneously. For example, the WW control region in 1-jet contains substantial top background. The top normalization in this region must therefore be simultanously determined using information from the 1-jet top control region.

Experimental and theoretical systematic uncertainties are considered on the extrapolation between signal and control regions, as well as on the extrapolation between the various control

regions. MC statistical uncertainties on the extrapolation are effectively taken into account through their implementation as described in Section 5.3.1. The relevant quantity to describe the extrapolation factor is the ratio of acceptances in each region. For a background B, this is $\alpha_B \equiv \frac{\varepsilon_B^{\mathrm{SR}}}{\varepsilon_B^{\mathrm{CR}}}$ for the CR→SR extrapolation and $\beta_B \equiv \frac{\varepsilon_B^{\mathrm{CR}_1}}{\varepsilon_B^{\mathrm{CR}_2}}$ for the CR_2 →CR_1 extrapolation. Theoretical uncertainties on the quantities α and β are typically applied to the expectation of the target background in the region to which the background is extrapolated. Experimental uncertainties are applied on the absolute acceptances in all regions. These two methods are equivalent and lead to the same effective uncertainty on the background, but the uncertainty on the ratio of acceptances is more straight forward to compute for sources like QCD scale, PDF uncertainties, or modeling uncertainties.

The WW CRs are defined similar to the SR. The m_T and $\Delta\phi_{\ell\ell}$ cuts are dropped. The region $50 < m_{\ell\ell} < 100$ GeV is used in 0-jet and $m_{\ell\ell} > 80$ GeV is used in 1-jet. A WW validation region with $m_{\ell\ell} > 100$ GeV is also used in 0-jet to validate the CR→SR WW extrapolation. This is described later in Section 6.8.1. The top CRs drop the selections on m_T, $\Delta\phi_{\ell\ell}$, and $m_{\ell\ell}$. For $Z \to \tau\tau$, all cuts are dropped after preselection with the exceptions $\Delta\phi_{\ell\ell} > 2.8$ and $m_{\ell\ell} < 80$ GeV.

Table 6.6 shows a summary of the yields in each control region. Figure 6.10 shows the transverse mass distribution in the WW CR for the 0- and 1-jet channels. Figure 6.11 shows the transverse mass for the top CR for the 1- and 2-jet channels.

6.8.1 Control Region Validation

The extrapolation from the WW CR to the SR in 0-jet is validated using a similarly defined region. A two category likelihood is used where the first is the standard 0-jet WW CR with $50 < m_{\ell\ell} < 100$ GeV, and the second is the $m_{\ell\ell} > 100$ GeV region. The method of computing systematics on the backgrounds and extrapolation is identical to the standard analysis. The likelihood can be parametrized such that the ratio of WW normalizations between the regions, $\alpha \equiv \frac{\mu_{WW1}}{\mu_{WW2}}$, is the parameter of interest and the WW normalization in

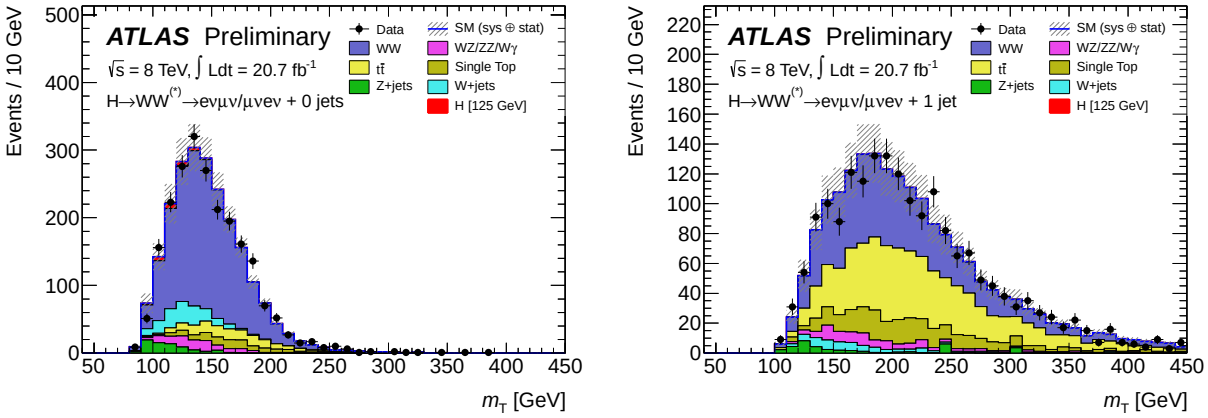


Figure 6.10: m_{T} distributions in the WW control region in the N_{jet}=0 (left) and N_{jet}=1 (right) analyses, before normalising the simulation to the rate in data. Only $e\mu$+μe channels and $\sqrt{s}$ = 8 TeV data are shown. The top backgrounds are scaled using the normalisation derived from the corresponding control regions described in the text. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

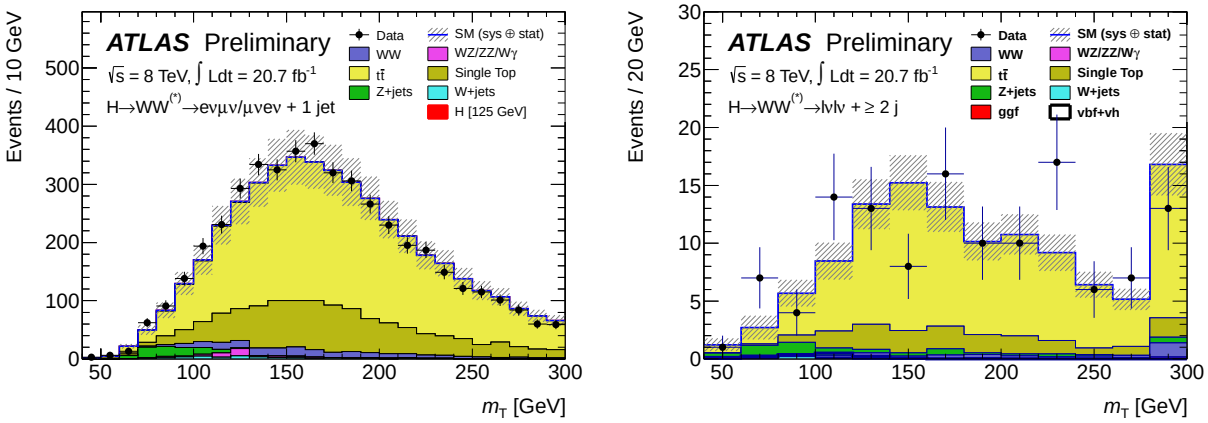


Figure 6.11: Distributions of m_{T} in N_{jet}=1 (left) and $N_{\mathrm{jet}}\geq$2 (right) top background control regions. The distributions are normalised to the data. The right-most bin in $N_{\mathrm{jet}}\geq$2 represents the overflow. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

Estimate	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
WW									
N_{jet}=0	2224	1970 ± 17	31 ± 0.7	1383 ± 9.3	100 ± 6.8	152 ± 4.4	107 ± 4.3	68 ± 10	160 ± 3.6
N_{jet}=1	1897	1893 ± 17	1.9 ± 0.3	752 ± 6.8	88 ± 5.5	717 ± 9.5	243 ± 6.7	37 ± 7.5	56 ± 2.5
$Z/\gamma^* \to \tau\tau$									
N_{jet}=0	1935	2251 ± 31	2.5 ± 0.2	61 ± 1.9	8.5 ± 1.1	4.5 ± 0.8	2.7 ± 0.6	2113 ± 31	61 ± 3.8
N_{jet}=1	2884	3226 ± 34	7.5 ± 0.3	117 ± 2.7	22 ± 3.1	570 ± 8.4	50 ± 3	2379 ± 32	88 ± 4.3
$N_{\text{jet}} \geq 2$	212	224 ± 7	0.6 ± 0.1	13 ± 1	4 ± 1	44 ± 3	5 ± 1	148 ± 6	9 ± 1
Top									
N_{jet}=1	4926	4781 ± 26	12 ± 0.5	184 ± 3.7	43 ± 9.5	3399 ± 20	1049 ± 13	72 ± 3.1	35 ± 2.2
$N_{\text{jet}} \geq 2$	126	201 ± 5	1.6 ± 0.1	6.4 ± 0.4	1.0 ± 0.3	157 ± 4	26 ± 2	9 ± 1	0.3 ± 0.4

Table 6.6: Control region yields for 8 TeV data. The observed (N_{obs}) and expected (N_{exp}) yields for the signal (N_{sig}) and background (N_{bkg}) processes are given. The composition of N_{bkg} is given on the right. For $N_{\text{jet}} \geq 2$, $N_{\text{sig,ggF}}$ is added to N_{bkg}. In general, no normalisation factors are applied with the following exception: the top and $Z/\gamma^* \to \tau\tau$ normalisation factors are applied for the corresponding estimates in the WW CRs. All uncertainties are statistical.

the validation region is a nuisance parameter. The test statistic $q = -2 \ln \frac{\mathcal{L}(\alpha=1)}{\mathcal{L}(\hat{\alpha})}$ is used to quantify compatibility. This is therefore testing the hypothesis that the two normalizations are equal to the hypothesis that they take on their independent unconditional values. q is asymptotically distributed as a χ^2 with one degree of freedom. The two sided probability for obtaining q larger than that that which is observed is $p = 2(1 - \Phi(\sqrt{q}))$. This corresponds to a statistical significance $Z = \Phi^{-1}(1-p) = \Phi^{-1}(2\Phi(\sqrt{q}) - 1)$. In the 8 TeV dataset, the test gives a statistical significance of 1.3 standard deviations.

This test is not able to be performed in the 1-jet channel because no validation region with alternate $m_{\ell\ell}$ criteria can be defined that would give sensitivity to the extrapolation validity.

6.9 Pacman

The Z/γ^* background in the SF channels is estimated using the Pacman method. After the $p_T^{\ell\ell}$ cut, the $Z/\gamma^* \to \ell\ell$ background that remains will be boosted in one direction, and balanced in the other by soft hadronic activity. Processes with real $\not{E}_T$ in the form of neutrinos, however, will tend to be balanced by real missing energy. The soft hadronic

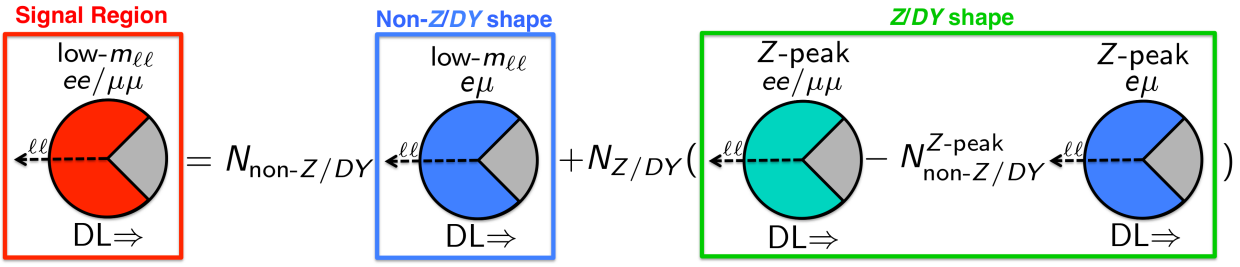


Figure 6.12: Scheme representing the Z/DY estimate procedure (the Pacman method).

activity in the quadrant opposite the dilepton axis can therefore be used to discriminate the Z/γ^* background from processes with real $\not{E}_T$. This method is represented in Figure 6.12. The variable to quantify this activity, f_{recoil}, is defined for the 0-jet channel as:

$$f_{\text{recoil}} = \frac{|\sum_{\text{jets}} |\text{JVF}| \times \tilde{\text{p}}_{\text{T}}^{\text{jet}}|}{p_T^{\ell\ell}} \tag{6.5}$$

The requirements $p_T > 10$ GeV, $|\eta| < 4.5$, and $\frac{3\pi}{4} < \Delta\phi(\ell\ell, \text{jet}) < \frac{5\pi}{4}$ are applied to the jets included in the numerator. The f_{recoil} distribution normalized to unit area is shown in Figure 6.13 for signal, DY, and the non-DY background for events in the 0-jet signal region. The Z/γ^* rejection versus signal efficiency is also shown.

The method can be extended to the 1-jet channel. In this case the dilepton system is already balanced by a hard jet for Z/γ^*, though the dilepton+jet system will still be balanced by the soft activity. The f_{recoil} variable can therefore be modified and redefined as $f_{\text{recoil}}^{\text{extended}}$. In addition to the p_T and $|\eta|$ selection, jets in the numerator must now pass $\frac{3\pi}{4} < \Delta\phi(\ell\ell j, \text{jet}) < \frac{5\pi}{4}$. That is, they must be in the quadrant opposite the dilepton+jet system. The denominator is also modified to be $p_T^{\ell\ell j}$. $f_{\text{recoil}}^{\text{extended}}$ is shown in Figure 6.14 for signal, DY, and non-DY backgrounds in the 1-jet signal region. Also shown is the Z/γ^* rejection versus signal efficiency for $f_{\text{recoil}}^{\text{extended}}$.

A cut of 0.05 is applied to f_{recoil} in 0-jet and 0.2 to $f_{\text{recoil}}^{\text{extended}}$ in 1-jet. Because these variables rely on soft hadronic activity, it is not expected that the MC will estimate the efficiency well. The data-driven method used to estimate the f_{recoil} and $f_{\text{recoil}}^{\text{extended}}$ efficiencies

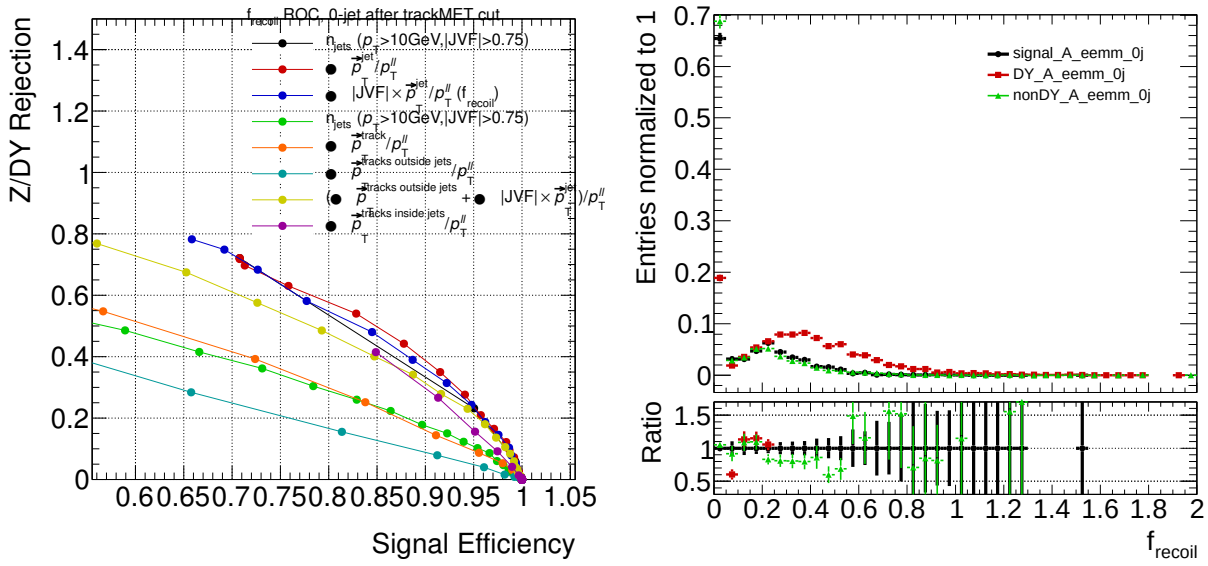


Figure 6.13: Left: Rejection of Z/DY background is plotted versus signal efficiency ("ROC" curves) for different hadronic recoil energy variables in the 0-jet signal region (after the $\Delta\phi_{\ell\ell}$<1.8 cut; therefore the full preselection has been applied, and $\not{E}_{T\,\mathrm{rel}}$>45 GeV, $p_{\mathrm{T,rel}}^{\mathrm{miss}}$>45 GeV, $p_{\mathrm{T}}^{\ell\ell}$>30 GeV, $m_{\ell\ell}$<50 GeV). Right: shape of f_{recoil} in the 0-jet signal region for Z/DY background, non-DY backgrounds (i.e. top, W+ jets, WW and other diboson backgrounds) and 125 GeV Higgs signal.

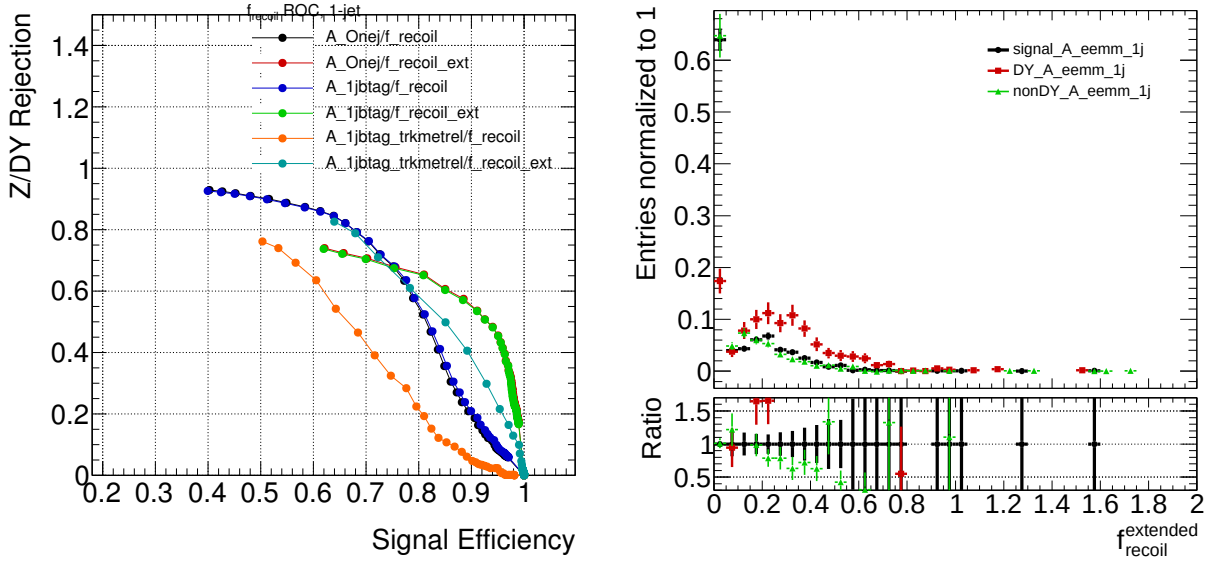


Figure 6.14: Left: Rejection of Z/DY background is plotted versus signal efficiency ("ROC" curves) for the standard and extended definitions in the 1-jet signal region. Right: shape of $f_{\mathrm{recoil}}^{\mathrm{extended}}$ in the 1-jet signal region for Z/DY background, non-DY backgrounds (i.e. top, W+ jets, WW and other diboson backgrounds) and 125 GeV Higgs signal.

N_{jet}	WW	top	$Z/\gamma^*\to\tau\tau$	$Z/\gamma^*\to\ell\ell$	$W+$jets	$WZ/ZZ/W\gamma$
$=0$ $e\mu$+μe	CR	CR	CR	MC	Data	MC+VR
$=0$ ee+$\mu\mu$	CR from $e\mu$+μe	CR from $e\mu$+μe	CR from $e\mu$+μe	Data	Data	MC+VR
$=1$ $e\mu$+μe	CR	CR	CR	MC	Data	MC+VR
$=1$ ee+$\mu\mu$	CR from $e\mu$+μe	CR from $e\mu$+μe	CR from $e\mu$+μe	Data	Data	MC+VR
≥ 2 $e\mu$+μe	MC	CR from $e\mu$+μe+ee+$\mu\mu$	CR	MC	Data	MC
≥ 2 ee+$\mu\mu$	MC	CR from $e\mu$+μe+ee+$\mu\mu$	CR from $e\mu$+μe	Data	Data	MC

Table 6.7: Background treatment summary. The estimation procedures for various background processes are given in four categories: normalised using a control region (CR); data-driven estimate (Data); normalised using the MC (MC); and normalised using the MC, but validated in a control region (MC+VR). The "from $e\mu$+μe" denotes that for the ee+$\mu\mu$ in the same N_{jet} mode, the $e\mu$+μe region is used instead. The "from $e\mu$+μe+ee+$\mu\mu$" denotes that the four regions are merged.

is described in Section 6.13.1.

6.10 Backgrounds

This section describes how the major backgrounds are estimated. WW and top ($t\bar{t}$ and single top, including tW, tb, and tqb) are the leading backgrounds in all jet bins. There are further contributions from $Z/\gamma^* \to \ell\ell$, $Z/\gamma^* \to \tau\tau$ with leptonic taus, W+jets with one fake lepton, and non-WW diboson, including $WZ^{(*)}$, $W\gamma$, $W\gamma^*$, and ZZ. Table 6.7 summarizes the major backgrounds and how they are estimated in each channel.

6.10.1 WW

In the 0- and 1-jet channels, the WW background is normalized using a control region close in phase space to the signal region, as described in Section 6.8. In the 2-jet channel, this background is taken from MC, and includes additional electroweak contributions from vector boson scattering.

6.10.2 Top

In the 1- and 2-jet channels, the top background is normalized using a control region close in phase space to the signal region, as described in Section 6.8. In the 0-jet channel, a

data-driven method is used. A 2-jet inclusive control region requiring at least one b-tagged jet at the preselection stage is used to estimate the probability for top events to lose a jet, $P_1^{\text{Btag,data}}$. Contributions due to non-top backgrounds are subtracted from the control sample. Probing jets in the control sample are defined as jets that are at least 0.1 away from the a tagged jet in ΔR. $P_1^{\text{Btag,data}}$ is therefore the ratio of the number of events with no probing jets to the total number of events. The corresponding quantity in MC is $P_1^{\text{Btag,MC}}$. The probability for a top event to end up in the 0-jet bin is $P_2^{\text{Estimated}} = \left(\frac{P_1^{\text{Btag,data}}}{P_1^{\text{Btag,MC}}}\right)^2 \times P_2^{MC}$. The scale factor $\frac{P_2^{\text{Estimated}}}{P_2^{MC}} = \left(\frac{P_1^{\text{Btag,data}}}{P_1^{\text{Btag,MC}}}\right)^2$ can therefore be used to correct the MC in the 0-jet bin.

6.10.3 $Z/\gamma^* \to \tau\tau$

In all channels, the $Z/\gamma^* \to \tau\tau$ background is normalized using a control sample close in phase space to the signal region, as described in Section 6.8

6.10.4 $Z/\gamma^* \to \ell\ell$

In the DF channels, the $Z/\gamma^* \to \ell\ell$ contribution is small and estimated from MC. In the SF channels, a data-driven method is used to estimate the background. For details on the method, see sections 6.9 and 6.13.1.

6.10.5 W+jets

A data-driven method is used to estimate the W+jets background in all regions. The lepton from the $W \to \ell\nu$ decay should be well modeled from MC, but the kinematics and fake rate of the second lepton is not. A W+jets enriched control sample used to estimate the probability for a jet to fake a lepton. Events in this sample must have one lepton which passes the standard lepton selections. The sample is split into two disjoint pass-fail samples. In the fail sample, a second lepton must fail the standard criteria, but pass a looser criteria. In the pass sample, a second lepton must pass the standard criteria. A fake factor f_ℓ is

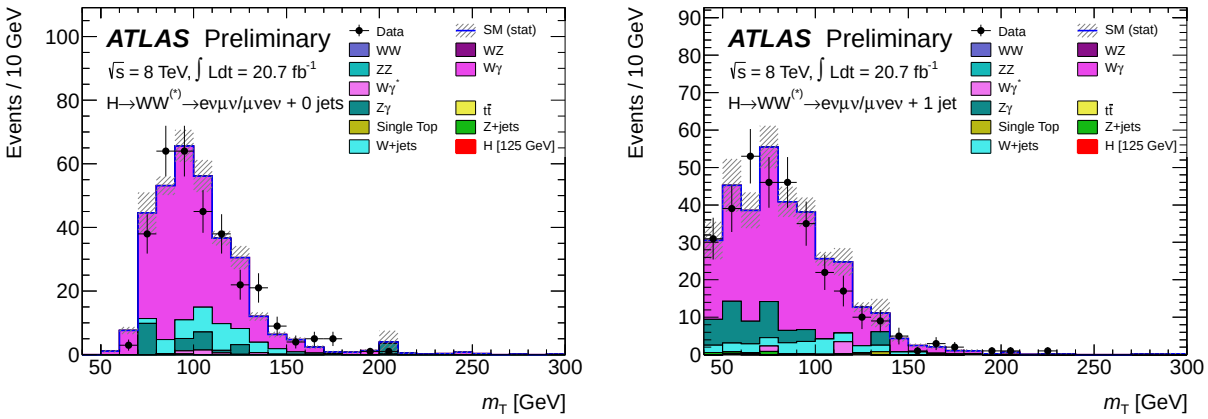


Figure 6.15: m_T distributions for the same-charge $W\gamma$ validation region: in the zero-jet (left) and one-jet (right) selection. The $e\mu$ and μe channels are combined. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

defined as the ratio of the pass to the fail sample and is computed as a function of η and p_T. The fail sample can also be defined for events in the standard event selection, in which one lepton passes and one fails the ID criteria, but also passes the looser selections. The W+jets contribution can be estimated with the events in this fail sample weighted by f_ℓ.

6.10.6 $WZ/ZZ/W\gamma/W\gamma^*$

Non-WW diboson backgrounds are estimated purely from MC. This is validated with a same-sign validation region, in which all standard selection criteria are applied, but the leptons have like signs. The validation region also receives contributions from W+jets, which can be estimated as in Section 6.10.5. The transverse mass distribution for this region is shown in Figure 6.15 separately for the DF and SF channels.

6.11 Systematics from Auxiliary Measurements

Systematics originating from external measurements can be divided into three classes. There are theory systematics on cross sections and the shapes of kinematic variables due to QCD scale, choice of parton density functions, and MC generator. There are experimental

systematics due to unknowns in object reconstruction and detector and collider performance, e.g. jet energy scale (JES), electron efficiency, and luminosity. Finally, there are uncertainties due to limited MC simulation statistics. The latter is covered in Section 5.3.1. The former are treated lognormally as discussed in Section 5.3. The sources considered are expanded upon here.

6.11.1 Theory Systematics

Systematics due to QCD scale inputs to theoretical calculations are computed by varying the renormalization and factorization scale up and down independently by a factor of two around its nominal value, with the maximum deviations from nominal taken as the $\pm 1\sigma$ uncertainty. This uncertainty is assumed to be uncorrelated among signal and background production processes. The QCD scale uncertainty on the ggF cross section is the dominant systematic on the final μ measurement. The uncertainty on the total and N-jet inclusive cross sections are all taken to be uncorrelated and amount to 8%, 20%, and 70% for the 0-, 1-, and 2-jet bins for $m_H = 125$ GeV. The uncertainty must be written within the exclusive jet bins and is treated according to Section 5.3.5. In the 2-jet bin, special consideration is made for the VBF selections. In particular, the central jet veto requires an additional uncertainty on the 3-jet inclusive cross section, which is treated as an extension to the 0- and 1-jet jet binned uncertainties. The QCD scale uncertainty on the VBF production cross section is small at around 1%. The scale uncertainties on production cross sections on the VH process as well as all backgrounds processes that are not estimated with data-driven methods are also considered.

The QCD scale uncertainties on kinematic acceptance for various processes are considered and computed as above. In cases where backgrounds are estimated from some auxiliary control region R_1 and extrapolated to a second region R_2, the uncertainty on the extrapolation coefficient $\frac{N_{MC}^{R_2}}{N_{MC}^{R_1}}$ are considered and implemented in the expected rate in R_2. The uncertainty on acceptance is assumed to be uncorrelated from the uncertainty on total production cross

section, and also between jet bins.

Systematics due to PDFs are computed as in [55, 56] for all processes to which they are applicable. These systematics are assumed to be correlated between processes with similar production modes, and grouped into the three categories gg, qg, and $q\bar{q}$. For example, both WZ and VBF Higgs processes are initiated through $q\bar{q}$ and so the uncertainty on the two are taken to be correlated. Uncertainties on acceptance due to PDFs are also computed and are taken to be uncorrelated with the uncertainty on the total cross section similar to the QCD scale treatment.

Systematics due to MC modeling, underlying event, and parton showering on acceptance are finally considered by comparing various MC generators and taking the envelope as the $\pm 1\sigma$ variation. These are non-negligible on the ggF Higgs signal and the extrapolation from WW and top control regions to the signal region.

6.11.2 Experimental Systematics

Experimental sources of systematics are treated as correlated between signal and background processes for each source. For each systematic, the source is varied by $\pm 1\sigma$ and the variation of the expected rate is computed. These are provided as the boundary values to the treatment outlined in Section 5.3.2. Systematics are considered from lepton energy scale (LES), resolution (LER), and efficiency, jet energy scale (JES) and resolution (JER), b-tagging efficiency, trigger efficiency, luminosity, and on the W+jets fake-factor.

The systematics due to LES and JES are propogated to the $\not{E}_T$ and $\not{E}_T{}^{\mathrm{STVF}}$ calculations. Further uncertainties on $\not{E}_T$ and $\not{E}_T{}^{\mathrm{STVF}}$ are considered from jets with $p_T < 20$ GeV and low energy calorimeter deposits not associated with reconstructed objects. The systematics on LES are further propogated to $\not{p}_T$. Further uncertianties are applied to $\not{p}_T$ scale and resolution due to sources not induced from charged leptons by comparing Z events in data and MC as a function of the total hard p_T of the event.

The JES systematics are divided into various experimental sources, each of which is

taken to be uncorrelated. These include the η intercalibration of jets from the central to the forward region, high-p_T jets, MC non-closure, topologies with close-by jets, quark/gluon composition and calorimeter response, b-jet energy scale, in-time and out-of-time pile-up, and in-situ jet energy corrections.

LER, LES, lepton and trigger efficiency uncertainties are estimated from $Z \to \ell\ell$, $J/\psi \to \ell\ell$, and $W \to \ell\nu$ decays. The electron efficiency uncertainty is between 2-5%, while the others are less than 1%.

The uncertainty on the W+jets fake factor is considered as a function of lepton p_T, and is around 40%.

The uncertainty on the total integrated luminosity is taken to be 3.6%.

6.11.3 Summary of Systematics

Table 6.8 shows the main systematics on the signal and background broken down by source. The extrapolation uncertainties on the control regions are shown in Table 6.9. These are futher broken down by source for the extrapolation of the WW background in Table 6.10.

Due to anticorrelation of many major systematics between the backgrounds, the uncertainty on the total background is lower than the individual backgrounds. Table 6.11 shows the expected events for the signal and total background along with the observed events, and the breakdown by each background source after a cut on the transverse mass. The uncertainty on the expected rates are also given.

6.12 7 TeV Analysis

The analysis of the 7 TeV data follows the 8 TeV analysis closely to allow for a robust combination of the two. Underlying event and pile-up is simulated with PYTHIA6 as opposed to PYTHIA8 at 8 TeV, though PYTHIA6 is still used for the WW background at 8 TeV. The JVF selections were loosened due to the lower pileup conditions in 2011. The SF channels were similarly optimized due to the lower Z/γ^* levels. Finally, the 2-jet analysis uses a cut

Source	Signal processes (%)			Background processes (%)		
	N_{jet}=0	N_{jet}=1	$N_{\text{jet}}\geq$2	N_{jet}=0	N_{jet}=1	$N_{\text{jet}}\geq$2
Theoretical uncertainties						
QCD scale for ggF signal for $N_{\text{jet}} \geq 0$	13	-	-	-	-	-
QCD scale for ggF signal for $N_{\text{jet}} \geq 1$	10	27	-	-	-	-
QCD scale for ggF signal for $N_{\text{jet}} \geq 2$	-	15	4	-	-	-
QCD scale for ggF signal for $N_{\text{jet}} \geq 3$	-	-	4	-	-	-
Parton shower and UE model (signal only)	3	10	5	-	-	-
PDF model	8	7	3	1	1	1
$H \to WW$ branching ratio	4	4	4	-	-	-
QCD scale (acceptance)	4	4	3	-	-	-
WW normalisation	-	-	-	1	2	4
Experimental uncertainties						
Jet energy scale and resolution	5	2	6	2	3	7
b-tagging efficiency	-	-	-	-	7	2
f_{recoil} efficiency	1	1	-	4	2	-

Table 6.8: Leading systematic uncertainties on the expected event yields for the 8 TeV analysis. The first four rows are calculated for inclusive N_{jet} modes and redistributed to exclusive ones. The QCD scale uncertainties on the inclusive ggF cross sections are anti-correlated between the exclusive N_{jet} modes. Sources contributing less than 4% to any column, and individual entries below 1%, are omitted.

Estimate	Stat. (%)	Theory (%)	Expt. (%)	Crosstalk (%)	Total (%)
WW					
N_{jet}=0	2.9	1.6	4.4	5.0	7.4
N_{jet}=1	6	5	4	36	37
Top					
N_{jet}=1	2	8	22	16	29
$N_{\text{jet}}\geq$2	10	15	29	19	39

Table 6.9: Total relative uncertainties on backgrounds that are normalised using control regions (CR). The statistical component (Stat.) is from the CR yields; the theoretical uncertainties (Theory) are from the α extrapolation parameter; the experimental (Expt.) uncertainties are given. The approximate uncertainties on the normalisation of other processes in the CR (Crosstalk) are given. The WW and top in N_{jet}=1 are anti-correlated due to the b-jet selection, so that the uncertainties partially cancel.

on $m_T < 150$ GeV rather than binning the distribution due to lower MC statistics.

The m_T distributions for the DF and SF channels in each 0- and 1-jet bins are shown in Figure 6.16. Table 6.12 shows the expected events by jet bin after an m_T cut.

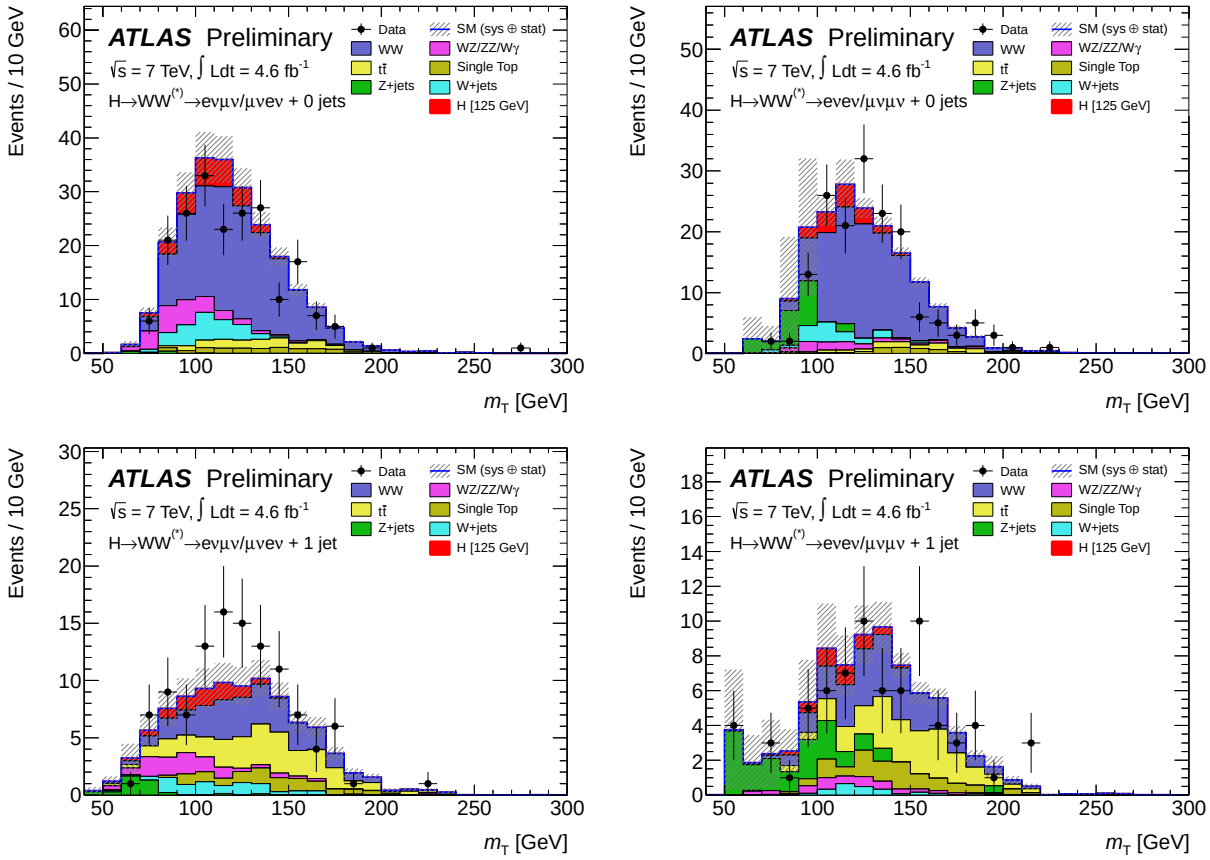


Figure 6.16: Distribution of transverse mass, m_{T}, for 7 TeV data. The plots are shown for the $e\mu$+μe (left) and ee+$\mu\mu$ (right) channels in the N_{jet}=0 (top) and N_{jet}=1 (bottom) modes. The visible signal is stacked on top of the background. The shaded area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources. Table 6.2 lists the selection order and Section 6.12 describes the modifications made with respect to the 8 TeV analysis.

Channel	Range (GeV)	QCD scale (%)	PS, UE (%)	PDF (%)	Modeling (%)
N_{jet}=0					
$e\mu$+μe	$10 < m_{\ell\ell} < 30$	0.9	0.2	1.5	−1.2
$e\mu$+μe	$30 \leq m_{\ell\ell} < 50$	0.9	0.8	1.1	−1.4
ee+$\mu\mu$	$12 < m_{\ell\ell} < 50$	1.0	0.3	1.1	1.7
N_{jet}=1					
$e\mu$+μe	$10 < m_{\ell\ell} < 30$	1.6	0.5	2.0	−5.1
$e\mu$+μe	$30 \leq m_{\ell\ell} < 50$	1.5	0.5	1.8	−5.0
ee+$\mu\mu$	$12 < m_{\ell\ell} < 50$	1.4	0.6	1.7	−3.1

Table 6.10: Uncertainties on the extrapolation parameters α for the WW background in the N_{jet}=0 and =1 channels. Uncertainties due to the QCD scale, PDF, parton shower (PS), underlying event (UE), and modeling of the NLO qq, $gq \to WW$ processes are given. Each source, represented by a column, is assumed to be uncorrelated, but for a given source the uncertainties are assumed to be fully correlated among all signal regions with N_{jet}=0 and =1. A relative sign between two entries in a column indicates anti-correlation between those signal regions for that source of uncertainty.

N_{jet}	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
= 0	831	739 ± 39	97 ± 20	551 ± 41	58 ± 8	23 ± 3	16 ± 2	30 ± 10	61 ± 21
= 1	309	261 ± 28	40 ± 13	108 ± 40	27 ± 6	68 ± 18	27 ± 10	12 ± 6	20 ± 5
≥ 2	55	36 ± 4	10.6 ± 1.4	4.1 ± 1.5	1.9 ± 0.4	4.6 ± 1.7	0.8 ± 0.4	22 ± 3	0.7 ± 0.2

Table 6.11: Summary selection table for 8 TeV data. The observed (N_{obs}) and the expected (N_{exp}) yields for the signal (N_{sig}) and background (N_{bkg}) processes are given in a window of m_{T}. The composition of N_{bkg} is given on the right. The $e\mu$+μe and ee+$\mu\mu$ channels are combined. The N_{sig} sums the ggF and VBF contributions. The selection modifications with respect to Table 6.2 are discussed in Section 6.12. The uncertainty on N_{bkg} accounts for the correlations among the sources.

N_{jet}	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
= 0	154	161 ± 11	25 ± 5	113 ± 10	12 ± 2	5 ± 1	4 ± 1	6 ± 2	21 ± 5
= 1	62	47 ± 6	7 ± 2	16 ± 6	5 ± 1	10 ± 3	6 ± 2	5 ± 2	5 ± 1
≥ 2	2	4.6 ± 0.8	1.4 ± 0.2	0.7 ± 0.2	-	0.7 ± 0.5	0.1 ± 0.1	2.4 ± 0.6	0.3 ± 0.1

Table 6.12: Summary selection table for 7 TeV data for events in a window of m_{T}. The uncertainty on N_{bkg} accounts for the correlations among the sources. More details are given in the caption of Table 6.11.

6.13 Statistical Treatment

The core philosophy of the statistical model in the analysis is that major backgrounds are normalized through dedicated control regions, which are described in section 6.8. Back-

grounds without control regions are taken either from MC simulation, or estimated with a data-driven method as described in section 6.10. The full transverse mass distribution after all selections is used to extract the signal. The shape is taken from template histograms for each signal and background process. In the DF 0- and 1-jet channels, the signal region in each channel is further divided by $10 < m_{\ell\ell} < 30$ GeV and $30 < m_{\ell\ell} < 50$ GeV to take advantage of different signal and background compositions and systematics. This is not done in the SF 0- and 1-jet channels due to poor Z/γ^* MC statistics. For the 0- and 1-jet channels, in order to minimize the effects of limited MC statistics and to avoid empty bins that could cause technical issues, the m_T distribution is first remapped such that the total nominal background is flat in each category. Five, three, and four bins are used in the 0-, 1-, and 2-jet channels, respectively. The number of bins are chosen to maximize sensitivity while allowing for a robust remapping given limited MC statistics. In the 2-jet channel the binning is manually selected to have boundaries at 50, 80, and 130 GeV, with all underflow and overflow included in the first and last bins, respectively. The full likelihood can be generally written as

$$\mathcal{L}(\mu,\theta) = \prod_{i}^{\text{signal regions}} \mathcal{L}_i^{SR}(\mu,\theta) \times \prod_{i}^{\text{control regions}} P(N_i^{CR}|\lambda_i(\mu,\theta)) \times \prod_{i}^{\text{systematics}} \mathcal{A}(\tilde{\theta}_i|\theta_i)$$

μ is a strength parameter multiplying the total signal such that $\mu = 0$ corresponds to the background only likelihood and $\mu = 1$ to the nominal signal + background likelihood. Each control region enters as a single Poisson, written explicitely in the formula, to avoid over constraints on the nuisance parameters.

Systematics follow the procedures outlined in section 5.3. Normalization systematics smaller than 0.5% are neglected to increase the computational performance of the model. Similarly, shape systematics with a maximum deviation less than 1% in all bins are neglected. Shape systematics are further examined to veto unphysical variations that arise simply due to limited MC statistics. MC statistical uncertainties derived from the nominal distributions are applied to the total background in each bin according to section 5.3.1.

Upper limits using CLs, p-values, and 68% confidence intervals are computed using asymptotic formulae described in Section 5.4.

6.13.1 Statistical Treatment of Pacman

Special consideration must be made to accommodate the Pacman method. The primary motivation is to estimate from data the f_{recoil} efficiency for Z/γ^* (ε^{DY}) and non-DY backgrounds (ε^{NDY}), and the normalization of Z/γ^* in the signal region using events that fail the f_{recoil} cut. The signal efficiency (ε^{S}) is nearly identical to the non-DY backgrounds, and the difference taken from MC is applied as both a systematic and a correction. The regions used to estimate these quantities have all cuts applied except m_T, $m_{\ell\ell}$, and $\Delta\phi$. The Z/γ^* efficiency can be constrained using SF events within a $\pm$15 GeV window around the Z-pole. The non-DY efficiency is constrained mostly from the DF SR. This is possible due to the very low Z/γ^* background in the DF SR and because the f_{recoil} shape is similar between signal and background. The Pacman-relevant components of the likelihood are written below.

$$
\begin{aligned}
&\mathcal{L}_{\text{pacman}} = \\
&P(N_{\text{pass}}^{DF\ SR}|\lambda_{\text{NDY}}^{DF\ SR}\varepsilon^{\text{NDY}})\times \\
&P(N_{\text{fail}}^{DF\ SR}|\lambda_{\text{NDY}}^{DF\ SR}(1-\varepsilon^{\text{NDY}}))\times \\
&P(N_{\text{pass}}^{SF\ SR}|\lambda_{\text{NDY}}^{SF\ SR}\varepsilon^{\text{NDY}} + \lambda_{\text{S}}^{SF\ SR}\varepsilon^{\text{S}}(\varepsilon^{\text{NDY}}) + \mu_{\text{DY}}^{SF\ SR}\lambda_{\text{DY}}^{SF\ SR}\varepsilon^{\text{DY}})\times \\
&P(N_{\text{fail}}^{SF\ SR}|\lambda_{\text{NDY}}^{SF\ SR}(1-\varepsilon^{\text{NDY}}) + \lambda_{\text{S}}^{SF\ SR}(1-\varepsilon^{\text{S}}(\varepsilon^{\text{NDY}})) + \mu_{\text{DY}}^{SF\ SR}\lambda_{\text{DY}}^{SF\ SR}(1-\varepsilon^{\text{DY}}))\times \\
&P(N_{\text{pass}}^{SF\ Z}|\lambda_{\text{NDY}}^{SF\ Z}\varepsilon^{\text{NDY}} + \mu_{\text{DY}}^{SF\ Z}\lambda_{\text{DY}}^{SF\ Z}\varepsilon^{\text{DY}})\times \\
&P(N_{\text{fail}}^{SF\ Z}|\lambda_{\text{NDY}}^{SF\ Z}(1-\varepsilon^{\text{NDY}}) + \mu_{\text{DY}}^{SF\ Z}\lambda_{\text{DY}}^{SF\ Z}(1-\varepsilon^{\text{DY}}))
\end{aligned}
\tag{6.6}
$$

The signal region terms show only a single Poisson as a representation, though in the full likelihood the SR pass regions are the binned m_T distributions. There are systematics applied to the extrapolation of ε^{NDY} from the DF SR to the SF SR, and to the extrapolation of ε^{DY} from the Z-peak to the SF SR. Finally, since ε^{S} is not identical to ε^{NDY}, a correction

and systematic are applied to the signal by writing ε^{S} as a function of $\varepsilon^{\mathrm{NDY}}$.

$$\begin{aligned}\varepsilon^{\mathrm{S}}(\varepsilon^{\mathrm{NDY}},\theta) &= \varepsilon^{\mathrm{NDY}} \times C_S(\varepsilon^{\mathrm{NDY}}) \times \kappa_S(\varepsilon^{\mathrm{NDY}})^{\theta} \\ &= \varepsilon^{\mathrm{NDY}} \times \left(1-(1-f_0)\frac{\varepsilon^{\mathrm{NDY}}(1-\varepsilon^{\mathrm{NDY}})}{\varepsilon_0^{\mathrm{NDY}}(1-\varepsilon_0^{\mathrm{NDY}})}\right) \times \left(1+\Delta\frac{\varepsilon^{\mathrm{NDY}}(1-\varepsilon^{\mathrm{NDY}})}{\varepsilon_0^{\mathrm{NDY}}(1-\varepsilon_0^{\mathrm{NDY}})}\right)^{\theta}\end{aligned} \tag{6.7}$$

$C_S(\varepsilon^{\mathrm{NDY}})$ is the correction term to correct for the difference between the non-DY background and signal f_{recoil} efficiency. $\kappa_S(\varepsilon^{\mathrm{NDY}})^{\theta}$ is a modified lognormal response term for the nuisance parameter θ and relative systematic Δ. $\varepsilon_0^{\mathrm{NDY}}$ is the nominal value of $\varepsilon^{\mathrm{NDY}}$ from data. f_0 is the ratio of the signal to the non-DY f_{recoil} efficiency from MC. The parametrization is chosen such that C_S and κ_S satisfy the boundary conditions $C_S(1) = \kappa_S(1) = C_S(0) = \kappa_S(0) = 1$, with $C_S(\varepsilon_0^{\mathrm{NDY}}) = f_0$ and $\kappa_S(\varepsilon_0^{\mathrm{NDY}}) = 1+\Delta$.

6.14 Results

The combined 7+8 TeV results are as follows. Figure 6.17 shows the value of p_0 and the 95% upper limit on the signal strength μ. At $m_H = 125$ GeV, the observed (expected) p_0 is 8×10^{-5} (1×10^{-4}), corresponding to 3.8 (3.7) standard deviations. The observed (expected) excluded mass range is $m_H > 133$ ($m_H > 119$) GeV. Note that the search range only extends to $m_H = 150$ GeV, so exclusion statements cannot be made about masses larger than this. Figure 6.18 shows the value of $\hat{\mu}$ versus m_H and the contour of the likelihood for μ versus m_H.

At $m_H = 125$ GeV, the measured value of μ is $\hat{\mu} = 1.01 \pm 0.21$ (stat.) $\pm$ 0.19 (theo. sys.) $\pm$ 0.12 (expt. sys.) $\pm$ 0.04 (lumi.) $= 1.01 \pm 0.31$ (tot.). The breakdown of the uncertainty on μ by source is shown in Table 6.13.

To address the statistical significance of the VBF production mechanism, the signal expectation can be reparametrized such that $\alpha = \frac{\mu_{\mathrm{VBF+VH}}}{\mu_{\mathrm{ggF+ttH}}}$ is the parameter of interest and μ_{ggF} is a nuisance parameter which is profiled. In this way the test statistic r_0 is a statement mostly about the VBF process, with a small VH contribution that must be included due to

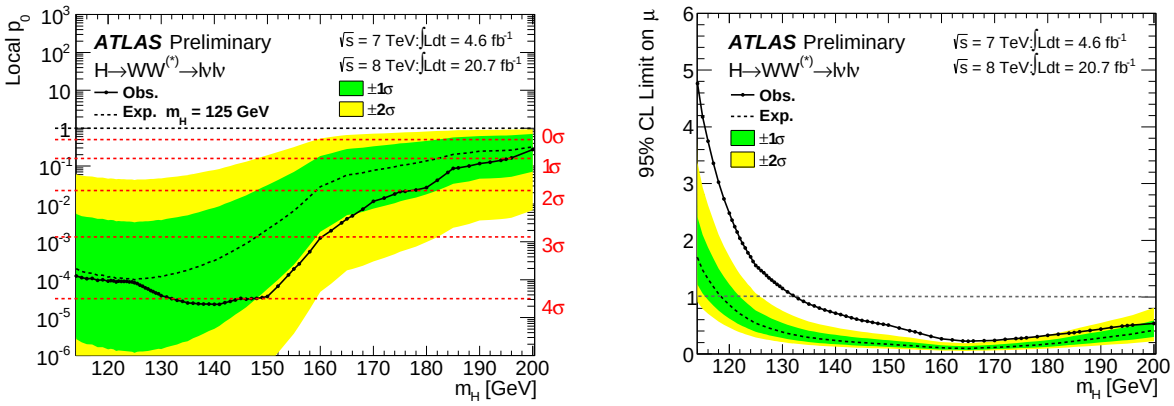


Figure 6.17: Results for p_0 (left) and 95% CL upper limit (right) using combined 7 TeV and 8 TeV data. The p_0 is the given probability for the background-only scenario as a function of m_H. The expected 95% CL upper limit is computed in the absence of a signal. The upper limit is on the cross section normalised to the SM cross section. For both figures, the smaller green bands represent $\pm 1\sigma$ uncertainties on the expected values, and the larger yellow bands represent $\pm 2\sigma$ uncertainties.

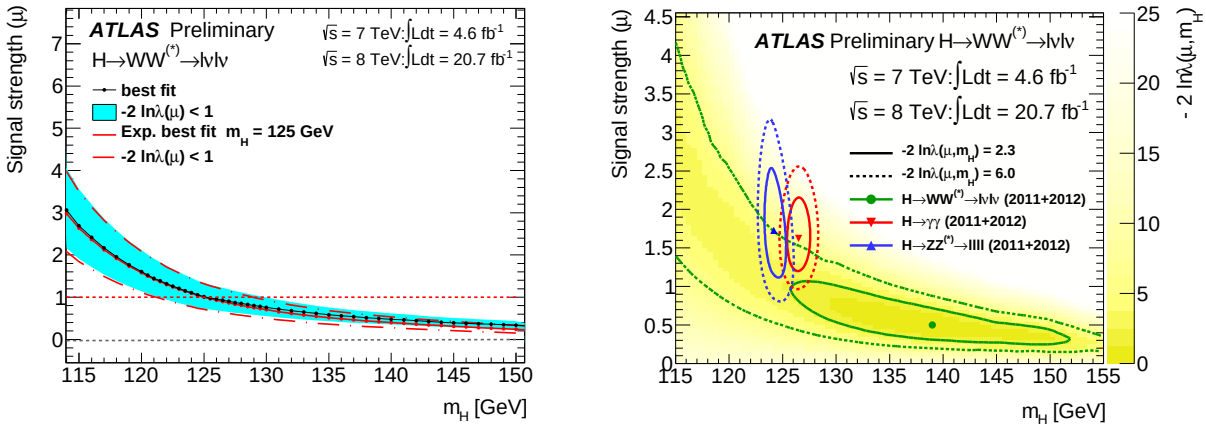


Figure 6.18: Signal strength parameter μ vs. m_H: (left) fitted μ value for the given m_H and (right) two-dimensional likelihood contours of $-2\ln\lambda(\mu, m_H)$ in the best-fit signal strength.

Category	Source	Uncertainty, up (%)	Uncertainty, down (%)
Statistical	Observed data	+21	−21
Theoretical	Signal yield ($\sigma \cdot \mathcal{B}$)	+12	−9
Theoretical	WW normalisation	+12	−12
Experimental	Objects and DY estimation	+9	−8
Theoretical	Signal acceptance	+9	−7
Experimental	MC statistics	+7	−7
Experimental	W+jets fake factor	+5	−5
Theoretical	Backgrounds, excluding WW	+5	−4
Luminosity	Integrated luminosity	+4	−4
Total		+32	−29

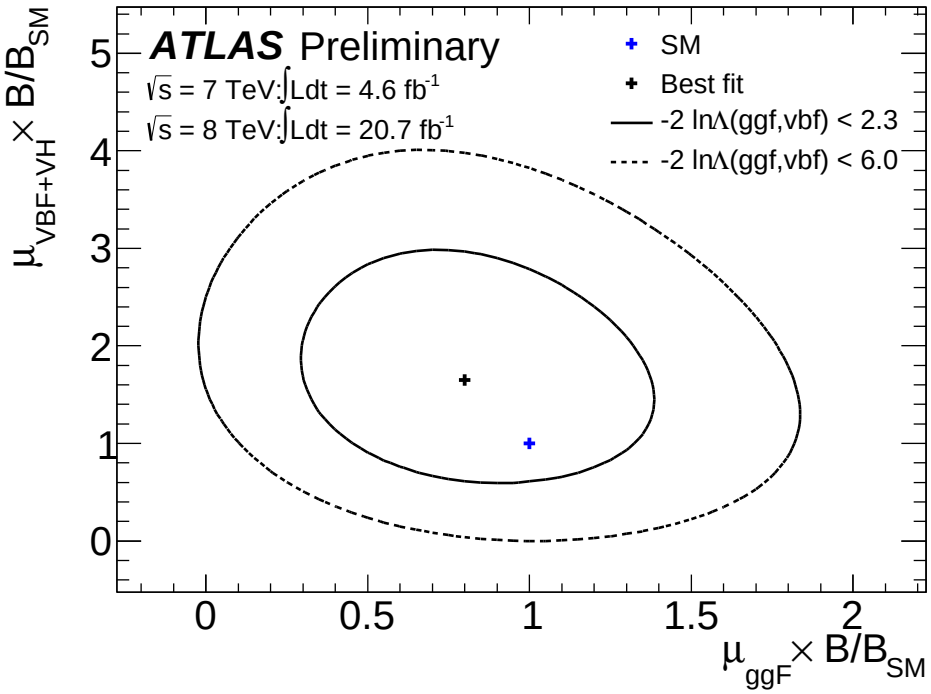
Table 6.13: Leading uncertainties on the signal strength μ for the combined 7 and 8 TeV analysis.

being similarly mediated through vector bosons. Figure 6.19 shows the delta log likelihood versus α for the full 0+1+2-jet analysis and the contour of $\mu_{\mathrm{VBF+VH}} \times \frac{B_{WW}}{B_{WW}^{SM}}$ versus $\mu_{\mathrm{ggF+ttH}} \times \frac{B_{WW}}{B_{WW}^{SM}}$. The statistical significance of the VBF process is asymptotically $\sqrt{\Delta \ln \mathcal{L}(\alpha = 0)}$, which corresponds to an observed (expected) significance of 2.5 (1.6) standard deviations.

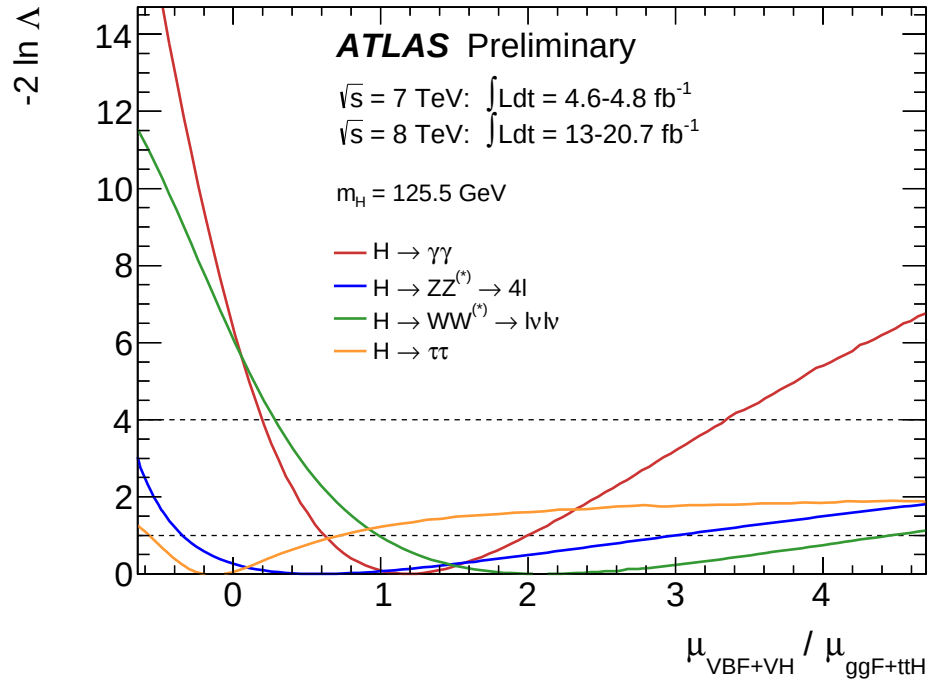
Figure 6.20 shows the background subtracted transverse mass distribution with the expected signal at 125 GeV overlayed. The individual 7 and 8 TeV results are summarized in Table 6.14. The measured value of the total inclusive $pp \to H$ production cross section at 8 TeV is shown in equation 6.8. The theoretical uncertainty on the total inclusive cross section and branching ratio is factorized out of this number since it is included in the expected SM Higgs production cross section shown below it.

$$\begin{aligned}
(\sigma \cdot \mathcal{B})_{\text{obs, 8 TeV}} &= 6.0 \pm 1.1 \text{ (stat.)} \pm 0.8 \text{ (theo. syst.)} \pm 0.7 \text{ (expt. syst.)} \pm 0.3 \text{ (lumi) pb} \\
&= 6.0 \pm 1.6 \text{ pb} \\
(\sigma \cdot \mathcal{B})_{\text{exp, 8 TeV}} &= 4.8 \pm 0.6 \text{ (cross section)} \pm 0.2 \text{ (branching ratio) pb} \\
&= 4.8 \pm 0.7 \text{ pb}
\end{aligned} \tag{6.8}$$

The compatibility between the 7 and 8 TeV results was tested using the same method described in Section 6.8.1, with the modification that the parameter of interest is defined as



(a)



(b)

Figure 6.19: (a) Likelihood contours for separate ggF and VBF signal strength parameters and (b) the likelihood curves for the ratio of the ggF/VBF strength parameters. The $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis uses the combined 7 and 8 TeV data.

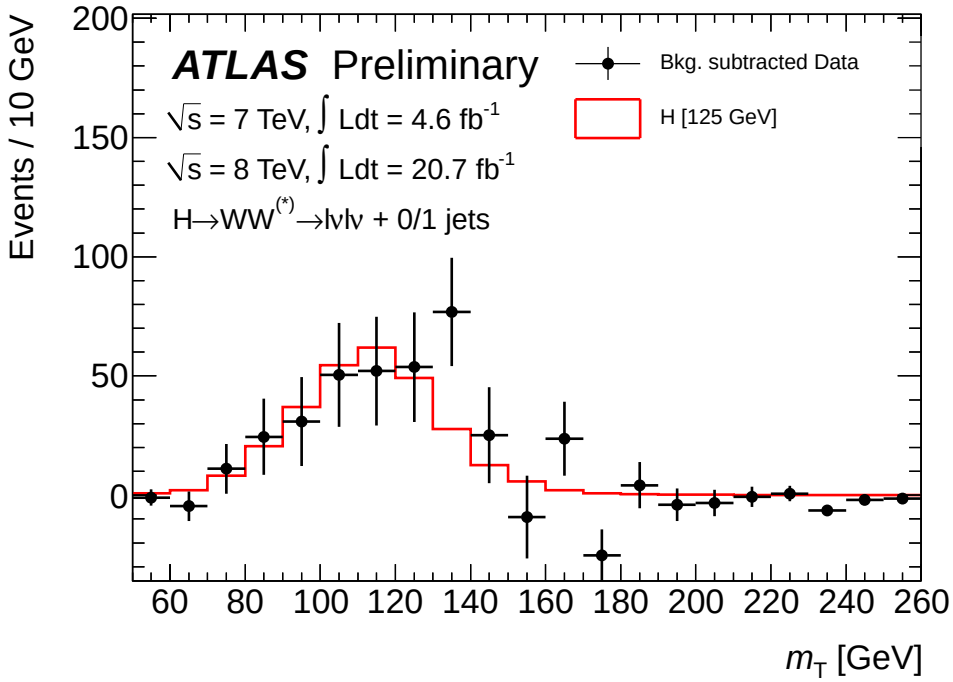


Figure 6.20: Background-subtracted m_{T} distribution for $N_{\mathrm{jet}} \leq 1$ in 7 and 8 TeV data. The signal is overlaid. The error bars represent the statistical uncertainties of the data and the subtracted background; it does not include the systematic uncertainties of the latter.

	8 TeV		7 TeV	
	Obs.	Exp.	Obs.	Exp.
p_0	1×10^{-5}	2×10^{-4}	0.5	0.04
Significance	4.3	3.5	0	1.8
$\hat{\mu}$	1.26±0.35	1±0.33	0±0.6	1±0.6

Table 6.14: Summary of individual 7 and 8 TeV results for a 125 GeV Higgs. The expected is computed assuming a signal strength $\mu = 1$.

$\alpha = \frac{\mu_{2012}}{\mu_{2011}}$. The two results are found to be compatible at the 1.8σ level.

CHAPTER VII

Higgs Combination

"If I have seen further than others, it is by standing on the shoulders of giants."

- Isaac Newton

This section gives details of the combination of search channels and measurements of the properties of the new boson. Several parametrizations of the likelihood are explored that are motivated by possible Beyond Standard Model (BSM) scenarios. Section 7.1 gives an overview of the search channels used througout the chapter, while Section 7.2 describes the procedure used for combination as well as the results for the alternative parametrizations employed to explore the Lagrangian structure of the Higgs.

7.1 Overview of Channels

The Higgs searches in ATLAS are divided by decay mode. The five currently accessible modes at low m_H are $H \to WW^{(*)} \to \ell\nu\ell\nu$, $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to \ell\ell\ell\ell$, $H \to b\bar{b}$, and $H \to \tau^+\tau^-$. Within each subchannel are targeted searches for each production process, namely gluon-gluon fusion, vector boson fusion, associated production, and production in association with top quark pairs. Table 7.1 shows a brief summary of the categories and integrated luminosity used for the SM search channels.

Higgs Boson Decay	Subsequent Decay	Sub-Channels	$\int L\,\mathrm{d}t$ [fb^{-1}]
		2011 $\sqrt{s}$ =7 TeV	
$H \to ZZ^{(*)}$	4ℓ	$\{4e, 2e2\mu, 2\mu2e, 4\mu\}$	4.6
$H \to \gamma\gamma$	–	10 categories $\{p_{\mathrm{Tt}} \otimes \eta_\gamma \otimes \text{conversion}\} \oplus \{\text{2-jet VBF}\}$	4.8
$H \to WW^{(*)}$	$\ell\nu\ell\nu$	$\{ee, e\mu, \mu e, \mu\mu\} \otimes \{\text{0-jet, 1-jet, 2-jet VBF}\}$	4.7
$H \to \tau\tau$	$\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$	$\{e\mu\} \otimes \{\text{0-jet}\} \oplus \{\ell\ell\} \otimes \{\text{1-jet, 2-jet, } p_{\mathrm{T},\tau\tau} > 100\,\mathrm{GeV}, VH\}$	4.6
	$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	$\{e, \mu\} \otimes \{\text{0-jet, 1-jet, } p_{\mathrm{T},\tau\tau} > 100\,\mathrm{GeV}, \text{2-jet}\}$	4.6
	$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$	$\{\text{1-jet, 2-jet}\}$	4.6
$VH \to Vbb$	$Z \to \nu\nu$	$E_{\mathrm{T}}^{\mathrm{miss}} \in \{120-160, 160-200, \geq 200\ \text{GeV}\} \otimes \{\text{2-jet, 3-jet}\}$	4.6
	$W \to \ell\nu$	$p_{\mathrm{T}}^{W} \in \{< 50, 50-100, 100-150, 150-200, \geq 200\ \text{GeV}\}$	4.7
	$Z \to \ell\ell$	$p_{\mathrm{T}}^{Z} \in \{< 50, 50-100, 100-150, 150-200, \geq 200\ \text{GeV}\}$	4.7
		2012 $\sqrt{s}$ =8 TeV	
$H \to ZZ^{(*)}$	4ℓ	$\{4e, 2e2\mu, 2\mu2e, 4\mu, 2-\text{jet VBF}, \ell-\text{tag}\}$	20.7
$H \to \gamma\gamma$	–	14 categories $\{p_{\mathrm{Tt}} \otimes \eta_\gamma \otimes \text{conversion}\} \oplus \{\text{2-jet VBF}\} \oplus \{\ell\text{-tag, } E_{\mathrm{T}}^{\mathrm{miss}}\text{-tag, 2-jet VH}\}$	20.7
$H \to WW^{(*)}$	$\ell\nu\ell\nu$	$\{ee, e\mu, \mu e, \mu\mu\} \otimes \{\text{0-jet, 1-jet, 2-jet VBF}\}$	20.7
$H \to \tau\tau$	$\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$	$\{\ell\ell\} \otimes \{\text{1-jet, 2-jet, } p_{\mathrm{T},\tau\tau} > 100\,\mathrm{GeV}, VH\}$	13
	$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	$\{e, \mu\} \otimes \{\text{0-jet, 1-jet, } p_{\mathrm{T},\tau\tau} > 100\,\mathrm{GeV}, \text{2-jet}\}$	13
	$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$	$\{\text{1-jet, 2-jet}\}$	13
$VH \to Vbb$	$Z \to \nu\nu$	$E_{\mathrm{T}}^{\mathrm{miss}} \in \{120-160, 160-200, \geq 200\ \text{GeV}\} \otimes \{\text{2-jet, 3-jet}\}$	13
	$W \to \ell\nu$	$p_{\mathrm{T}}^{W} \in \{< 50, 50-100, 100-150, 150-200, \geq 200\ \text{GeV}\}$	13
	$Z \to \ell\ell$	$p_{\mathrm{T}}^{Z} \in \{< 50, 50-100, 100-150, 150-200, \geq 200\ \text{GeV}\}$	13

Table 7.1: Summary of the individual channels entering the combined results presented here. In channels sensitive to associated production of the Higgs boson, V indicates a W or Z boson. The symbols $\otimes$ and $\oplus$ represent direct products and sums over sets of selection requirements, respectively.

7.1.1 $H \to WW^{(*)} \to \ell\nu\ell\nu$

The $H \to WW^{(*)} \to \ell\nu\ell\nu$ channel [14] is discussed in detail in Chapter VI, though a brief summary follows. The $H \to WW^{(*)} \to \ell\nu\ell\nu$ search channel is characterized by large missing energy and two high p_T isolated leptons with low dilepton mass due to the spin-0 nature of the Higgs. The search is divided into 0- and 1-jet exclusive analyses with a 2-jet inclusive analysis targeted for VBF production. The final state lepton combinations are divided into four categories: ee, $\mu\mu$, $e\mu$, and μe, where the latter two are distinguished by the flavor of the lepton with highest p_T. Due to the neutrinos in the final state, the invariant mass of the hypothetical Higgs cannot be fully reconstructed, therefore the transverse mass distribution

is fit.

7.1.2 $H \to \gamma\gamma$

The $H \to \gamma\gamma$ channel [13] is able to fully reconstruct the hypothesized Higgs mass through the diphoton invariant mass $m_{\gamma\gamma}$. At least two high p_T photons are required with $p_T^{\text{leading}} > 40$ GeV and $p_T^{\text{sub-leading}} > 30$ GeV. Events are divided into various categories depending on the event topology, as shown in Figure 7.1.

In each category, the $m_{\gamma\gamma}$ spectrum is fit to a continuous function. The background model varies from category to category, as described in Ref. [5], between a fourth order Bernstein polynomial, an exponential of a second-order polynomial, or a single exponential. The parameters of the background model are fit from the data. The signal is modeled with a Crystal Ball [44, 61, 69] plus Gaussian function. The choice of both signal and background functions is validated with MC. The inclusive distribution of $m_{\gamma\gamma}$ along with the fit to both background-only and signal+background is shown in Figure 7.2.

7.1.3 $H \to ZZ^{(*)} \to \ell\ell\ell\ell$

The $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ channel [12] is characterized by four high p_T isolated leptons. At $m_H = 125$ GeV, one pair of leptons should be consistent with the Z-pole mass with the second pair coming from the off-shell Z^*. Because there are no neutrinos in the final state, the hypothetical Higgs mass m_H can be fully reconstructed. This channel suffers from a relatively low expected rate, however, due to the small branching fractions of both $H \to ZZ^{(*)}$ and the leptonic decays of each Z. Even so, an even smaller background and peaked signal yields a large signal to background ratio. The distribution of $m_{4\ell}$ for this channel is shown in Figure 7.3.

The analysis is separated into ggF-, VBF-, and VH-like categories as follows. Events with at least two high-p_T jets, where jets must be at least 25 (30) GeV for $|\eta| < 2.5$ ($2.5 \leq |\eta| < 4.5$), are considered for the VBF and VH categories. If the two highest p_T jets are separated

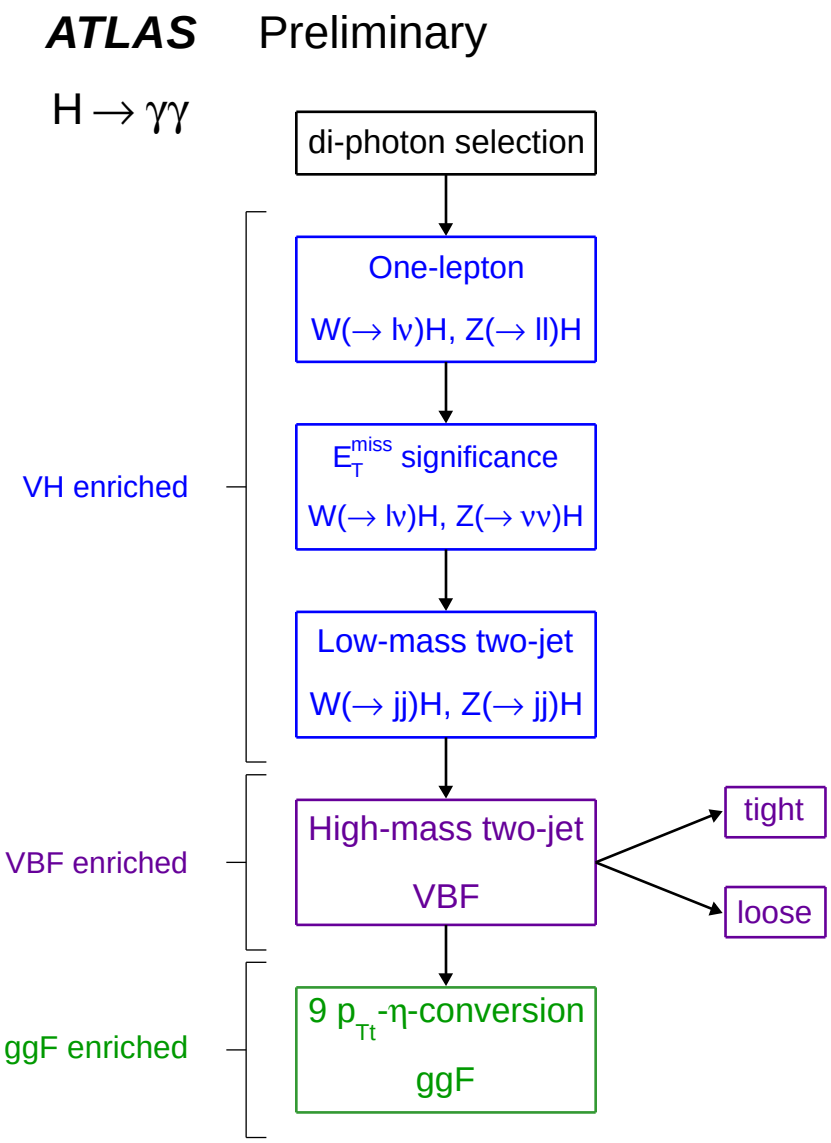


Figure 7.1: Flow-chart of the event categorization in the $H \to \gamma\gamma$ channel, giving the order of selection of the different categories.

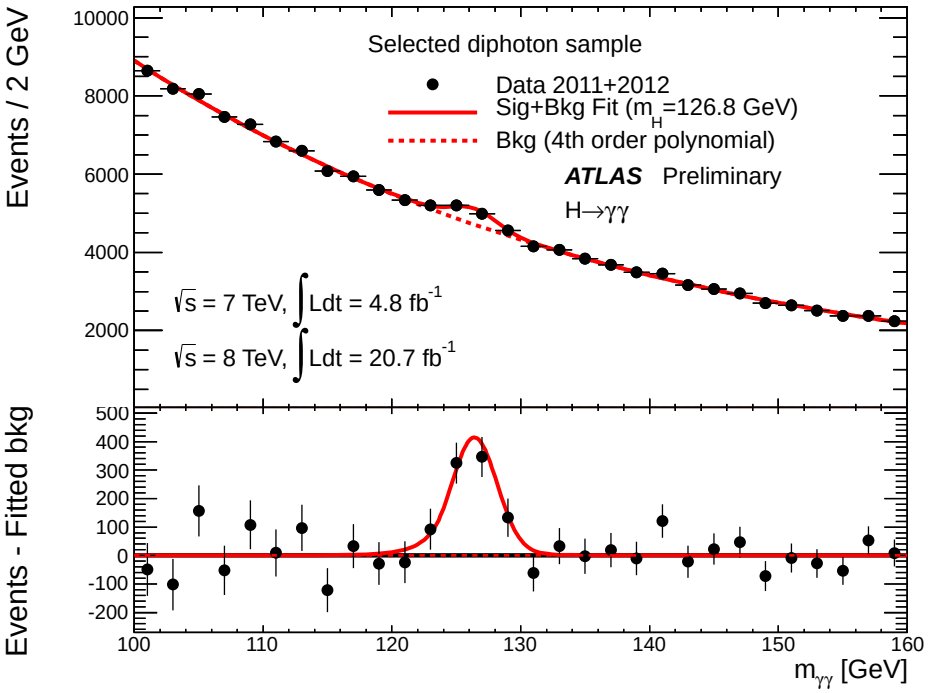


Figure 7.2: Distribution of $m_{\gamma\gamma}$ in the $H \to \gamma\gamma$ channel. This includes the maximum likelihood fits to both background-only and signal+background.

by at least three units of pseudo-rapidity and have an invariant mass $m_{jj} > 350$ GeV, the event is put into the VBF category, otherwise it is placed in the VH category. If the event does not fall into either of these, it is placed in the ggF category. Further, if any event has additional leptons that pass the lepton selection criteria beyond the standard four are placed in the VH category. For the signal, the four-lepton invariant mass $m_{4\ell}$ is parametrized using a Keys PDF [34] to model the signal shape and B-Splines (see Appendix A) to parametrize the expected rate and systematic uncertainties as a function of m_H. The background and data are modeled with binned template histograms.

7.1.4 $H \to b\bar{b}$

The $H \to b\bar{b}$ decay mode [7, 8] is accessible in the associated production process VH, $V = W, Z$. Three channels are considered in total. One mode is the case that the vector boson is a W which decays into a lepton and neutrino. The other two are for $V = Z$, where the Z-boson

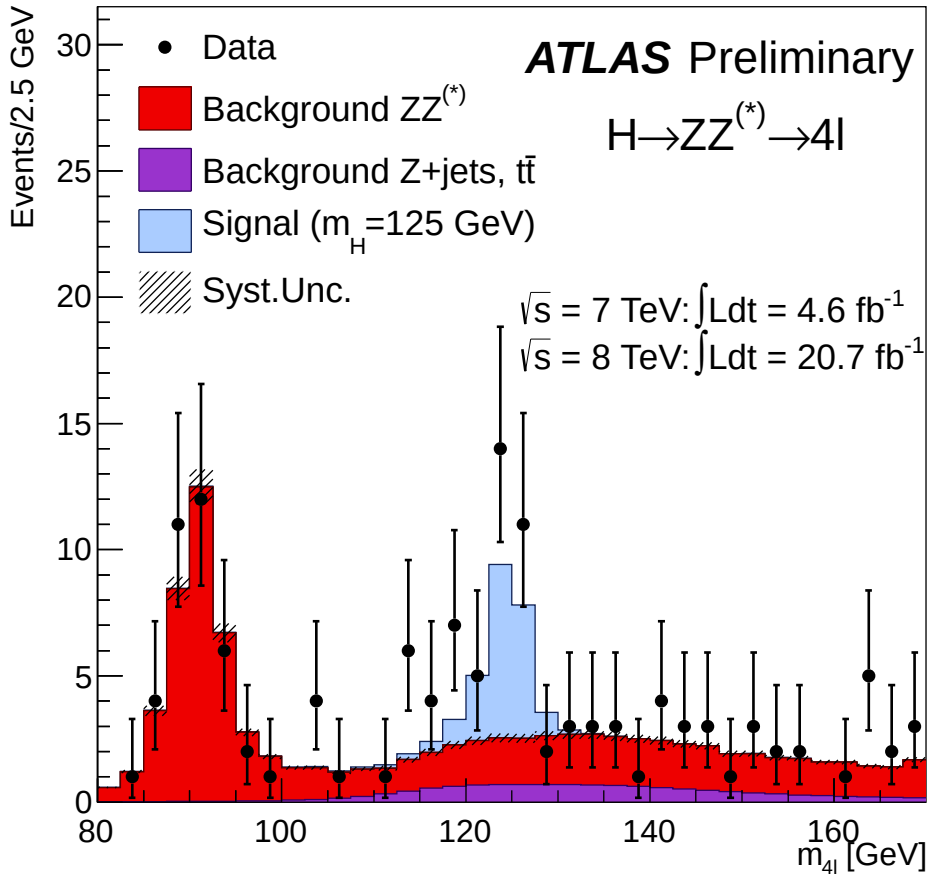


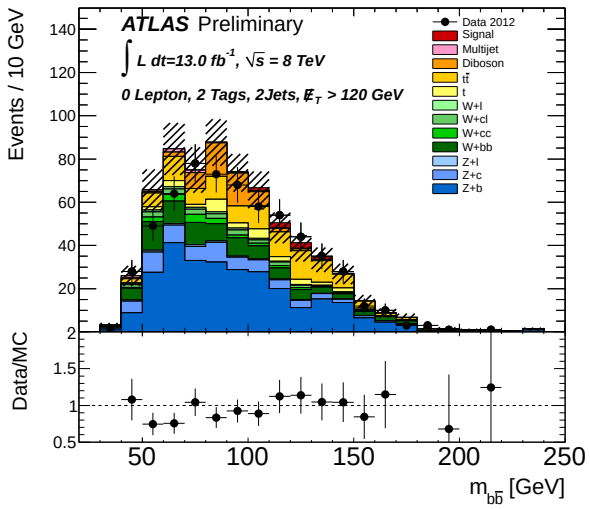
Figure 7.3: Four-lepton invariant mass distribution in the $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ channel.

decays into either lepton pairs or invisibly into two neutrinos. The analysis is separated into different categories depending on p_T^V, the p_T of the boson. In the 0-lepton mode, three p_T^V categories are used for each 2- and 3-jet channels, yielding six 0-lepton categories. For the 1- and 2-lepton modes, five p_T^V categories are used, yielding a total of 16 final categories. The final discriminant is the invariant mass of the two b-jets m_{bb}, which is modeled with a binned likelihood. The distribution for each lepton category is shown in Figure 7.4.

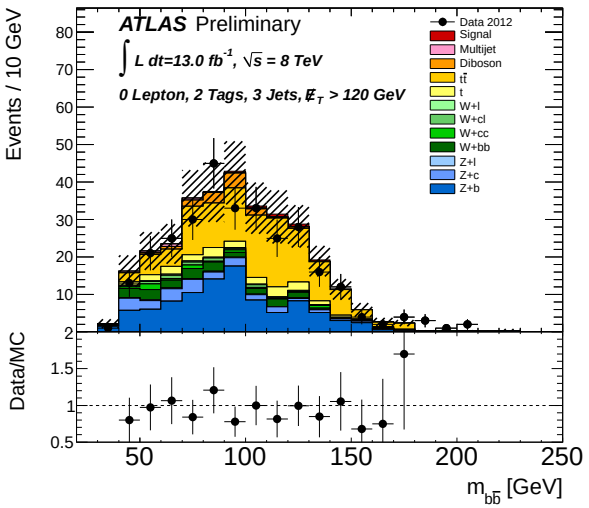
7.1.5 $H \to \tau^+\tau^-$

The $H \to \tau^+\tau^-$analysis [6] divides the search into three main categories based on the decay mode of the hypothetical τ leptons: $\tau_{\text{lep}}\tau_{\text{lep}}$, $\tau_{\text{lep}}\tau_{\text{had}}$, $\tau_{\text{had}}\tau_{\text{had}}$. Each of these final states is further subdivided into several search channels based on the flavor of the leptonic tau decay, the jet multiplicity, and the p_T of the reconstructed $\tau\tau$ system. Object preselection in the channel is as follows. Electrons are required to pass $E_T > 15$ GeV and fall within $|\eta| < 2.47$, excluding the transition region $1.37 < |\eta| < 1.52$. Muons must satisfy $p_T > 10$ GeV and $|\eta| < 2.5$. Track and calorimeter isolation is used for both electrons and muons. Hadronic tau decays are characterized by either one or three associated charged hadrons. To reconstruct hadronic taus, either one or three tracks with $p_T > 1$ GeV are required to fall within $\Delta R < 0.2$ of the tau candidate, and the total charge of the tau candidate decay products must be ± 1. A BDT is used based on tracking and calorimeter information to discriminate between real hadronic taus and fake jets. Jets themselves are reconstructed with the anti-k_T algorithm with a distance parameter of $R = 0.4$, and are required to have $p_T > 25$ GeV and fall within $|\eta| < 4.5$.

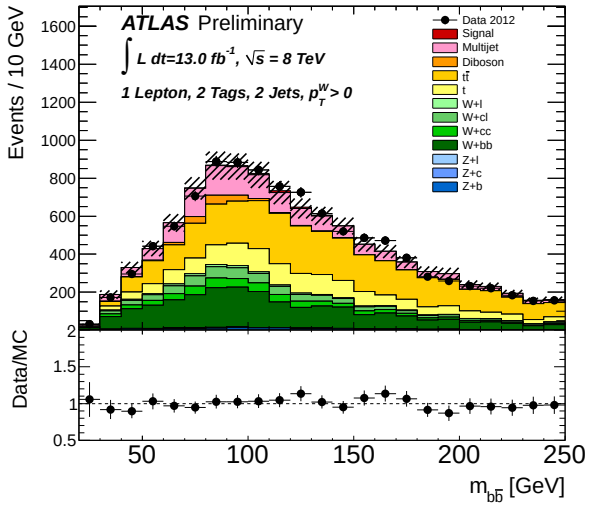
The categorization of the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel is summarized in Table 7.2. There is a 1-jet, 2-jet VBF, 2-jet VH, and a boosted channel in the 8 TeV analysis. Each of these are further divided by lepton flavor, being ee, $e\mu$, and $\mu\mu$. For the 7 TeV analyses there is also a 0-jet $e\mu$ category, yielding a total of 12 and 13 categories for 8 TeV and 7 TeV, respectively. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel there are a total of 8 categories for each 7 and 8 TeV analyses. For each



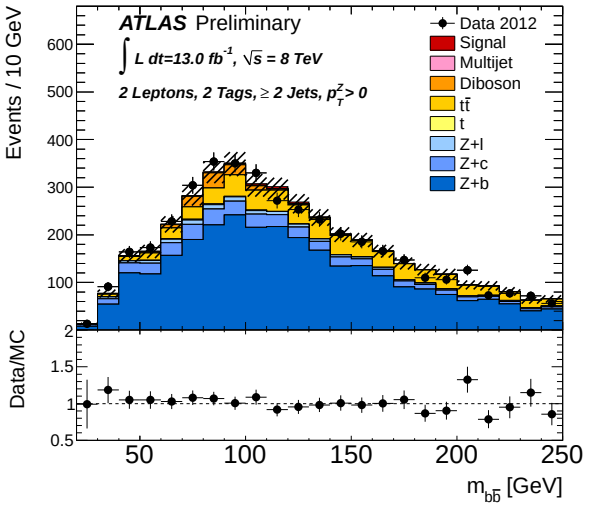
(a) 0-leptons, 2-jets



(b) 0-leptons, 3-jets



(c) 1-lepton



(d) 2-leptons

Figure 7.4: $H \to b\bar{b}$ invariant mass for each category integrated over p_T^V

2-jet VBF	Boosted	2-jet VH	1-jet
Pre-selection: exactly two leptons with opposite charges			
$30\,\text{GeV} < m_{\ell\ell} < 75\,\text{GeV}$ ($30\,\text{GeV} < m_{\ell\ell} < 100\,\text{GeV}$) for same-flavor (different-flavor) leptons, and $p_{T,\ell 1} + p_{T,\ell 2} > 35\,\text{GeV}$			
At least one jet with $p_T > 40\,\text{GeV}$ ($\lvert JVF\rvert > 0.5$ if $\lvert\eta_{\text{jet}}\rvert < 2.4$)			
$\not{E}_T > 40\,\text{GeV}$ ($\not{E}_T > 20\,\text{GeV}$) for same-flavor (different-flavor) leptons			
$H_{\text{T}}^{\text{miss}} > 40\,\text{GeV}$ for same-flavor leptons			
$0.1 < x_{1,2} < 1$			
$0.5 < \Delta\phi_{\ell\ell} < 2.5$			
$p_{T,j2} > 25\,\text{GeV}$ (JVF)	excluding 2-jet VBF	$p_{T,j2} > 25\,\text{GeV}$ (JVF)	excluding 2-jet VBF, Boosted and 2-jet VH
$\Delta\eta_{jj} > 3.0$	$p_{T,\tau\tau} > 100\,\text{GeV}$	excluding Boosted	$m_{\tau\tau j} > 225\,\text{GeV}$
$m_{jj} > 400\,\text{GeV}$	b-tagged jet veto	$\Delta\eta_{jj} < 2.0$	b-tagged jet veto
b-tagged jet veto	–	$30\,\text{GeV} < m_{jj} < 160\,\text{GeV}$	–
Lepton centrality and CJV		b-tagged jet veto	
0-jet (7 TeV only)			
Pre-selection: exactly two leptons with opposite charges			
Different-flavor leptons with $30\,\text{GeV} < m_{\ell\ell} < 100\,\text{GeV}$ and $p_{T,\ell 1} + p_{T,\ell 2} > 35\,\text{GeV}$			
$\Delta\phi_{\ell\ell} > 2.5$			
b-tagged jet veto			

Table 7.2: The categorization of the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ analysis. The JVF cut is $\lvert JVF\rvert > 0.75$ for 7 TeV data, the lepton centrality is not applied for 7 TeV analysis, and the 0-jet category is not used for 8 TeV data analysis.

electron and muon lepton flavor, there are 0-, 1-, and 2-jet categories along with a boosted category similar to the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel. These are outlined in more detail in Table 7.3. Finally, the $\tau_{\text{had}}\tau_{\text{had}}$ has a boosted 1-jet and 2-jet VBF category as outlined in Table 7.4. The final discriminant used in the likelihood is the $\tau\tau$ invariant mass, which is reconstructed using the Missing Mass Calculator (MMC) [40] technique, except in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel for 7 TeV data, which uses collinear approximation.

7.2 Combination

Given a set of N likelihoods $\mathcal{L}_i(\mu_i, \theta_i)$ from individual channels with a set of parameters of interest μ_i and nuisance parameters θ_i, a combined likelihood can be formed by a simple multiplication of the data portion of $\mathcal{L}_i$, namely $\mathcal{L}_i^{\text{data}}$, along with a multiplication of all unique auxiliary constraints $\mathcal{A}_{ij}$. This assumes the data portion of the individual likelihoods

7 TeV		8 TeV	
VBF Category	Boosted Category	VBF Category	Boosted Category
▷ ${p_T}^{\tau_{\text{had-vis}}}$ >30 GeV	–	▷ ${p_T}^{\tau_{\text{had-vis}}}$ >30 GeV	▷ ${p_T}^{\tau_{\text{had-vis}}}$ >30 GeV
▷ $\not\!\!E_T$ >20 GeV	▷ $\not\!\!E_T$ >20 GeV	▷ $\not\!\!E_T$ >20 GeV	▷ $\not\!\!E_T$ >20 GeV
▷ ≥ 2 jets	▷ $p_{\text{T}}^{\text{H}} > 100$ GeV	▷ ≥ 2 jets	▷ $p_{\text{T}}^{\text{H}} > 100$ GeV
▷ ${p_T}^{j1}, {p_T}^{j2} > 40$ GeV	▷ $0 < x_1 < 1$	▷ ${p_T}^{j1} > 40, {p_T}^{j2}$ >30 GeV	▷ $0 < x_1 < 1$
▷ $\Delta\eta_{jj} > 3.0$	▷ $0.2 < x_2 < 1.2$	▷ $\Delta\eta_{jj} > 3.0$	▷ $0.2 < x_2 < 1.2$
▷ $m_{jj} > 500$ GeV	▷ Fails VBF	▷ $m_{jj} > 500$ GeV	▷ Fails VBF
▷ centrality req.	–	▷ centrality req.	–
▷ $\eta_{j1} \times \eta_{j2} < 0$	–	▷ $\eta_{j1} \times \eta_{j2} < 0$	–
▷ ${p_T}^{\text{Total}} < 40$ GeV	–	▷ ${p_T}^{\text{Total}} < 30$ GeV	–
–	–	▷ ${p_T}^{\ell}$ >26 GeV	–
• m_{T} <50 GeV	• m_{T} <50 GeV	• m_{T} <50 GeV	• m_{T} <50 GeV
• $\Delta(\Delta R) < 0.8$	• $\Delta(\Delta R) < 0.8$	• $\Delta(\Delta R) < 0.8$	• $\Delta(\Delta R) < 0.8$
• $\sum \Delta\phi < 3.5$	• $\sum \Delta\phi < 1.6$	• $\sum \Delta\phi < 2.8$	–
–	–	• b-tagged jet veto	• b-tagged jet veto
1 Jet Category	0 Jet Category	1 Jet Category	0 Jet Category
▷ ≥ 1 jet, ${p_T}$ >25 GeV	▷ 0 jets ${p_T}$ >25 GeV	▷ ≥ 1 jet, ${p_T}$ >30 GeV	▷ 0 jets ${p_T}$ >30 GeV
▷ $\not\!\!E_T$ >20 GeV	▷ $\not\!\!E_T$ >20 GeV	▷ $\not\!\!E_T$ >20 GeV	▷ $\not\!\!E_T$ >20 GeV
▷ Fails VBF, Boosted	▷ Fails Boosted	▷ Fails VBF, Boosted	▷ Fails Boosted
• m_{T} <50 GeV	• m_{T} <30 GeV	• m_{T} <50 GeV	• m_{T} <30 GeV
• $\Delta(\Delta R) < 0.6$	• $\Delta(\Delta R) < 0.5$	• $\Delta(\Delta R) < 0.6$	• $\Delta(\Delta R) < 0.5$
• $\sum \Delta\phi < 3.5$	• $\sum \Delta\phi < 3.5$	• $\sum \Delta\phi < 3.5$	• $\sum \Delta\phi < 3.5$
–	• ${p_T}^{\ell} - {p_T}^{\tau} < 0$	–	• ${p_T}^{\ell} - {p_T}^{\tau} < 0$

Table 7.3: Event requirements applied in the different categories of the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ analysis. Requirements marked with a triangle (▷) are categorization requirements, meaning that if an event fails that requirement it is still considered for the remaining categories. Requirements marked with a bullet (•) are only applied to events passing all categorization requirements in a category; events failing such requirements are discarded.

are independent.

$$\begin{aligned} \mathcal{L}_i(\mu_i, \theta_i) &= \mathcal{L}_i^{\text{data}}(\mu_i, \theta_i) \times \prod_j^{M_i} \mathcal{A}_{ij}(\tilde{\theta}_{ij}|\theta_{ij}) \\ \mathcal{L}_{\text{combined}}(\mu, \theta) &= \left(\prod_i^{N} \mathcal{L}_i^{\text{data}}(\mu_i, \theta_i)\right)\left(\prod_j^{M_{\text{tot}}} \mathcal{A}_j(\tilde{\theta}_j|\theta_j)\right) \end{aligned} \tag{7.1}$$

The combined set of nuisance parameters $\theta = \theta_1 \cup \theta_2 \cup ... \cup \theta_{M_{\text{tot}}}$ is the set of unique parameters out of all individual sets θ_i. Similarly, $\mu = \mu_1 \cup \mu_2 \cup ... \cup \mu_{N_{\text{tot}}}$ is the set of unique parameters of interests of all individual sets μ_i. Correlation of parameters between channels is handled through this non-duplication of auxiliary constraints and unique parameters.

Cut	Description
Preselection	No muons or electrons in the event
	Exactly 2 medium τ_{had} candidates matched with the trigger objects
	At least 1 of the τ_{had} candidates identified as tight
	Both τ_{had} candidates are from the same primary vertex
	Leading $\tau_{\text{had-vis}}$ $p_T > 40$ GeV and sub-leading $\tau_{\text{had-vis}}$ $p_T > 25$ GeV, $\lvert\eta\rvert < 2.5$
	τ_{had} candidates have opposite charge and 1- or 3-tracks
	$0.8 < \Delta R(\tau_1, \tau_2) < 2.8$
	$\Delta\eta(\tau, \tau) < 1.5$
	if $\not{E}_T$ vector is not pointing in between the two taus, $\min\{\Delta\phi(\not{E}_T, \tau_1), \Delta\phi(\not{E}_T, \tau_2)\} < 0.2\pi$
VBF	At least two tagging jets, j_1, j_2, leading tagging jet with $p_T > 50$ GeV
	$\eta_{j1} \times \eta_{j2} < 0$, $\Delta\eta_{jj} > 2.6$ and invariant mass $m_{jj} > 350$ GeV
	$\min(\eta_{j1}, \eta_{j2}) < \eta_{\tau 1}, \eta_{\tau 2} < \max(\eta_{j1}, \eta_{j2})$
	$\not{E}_T > 20$ GeV
Boosted	Fails VBF
	At least one tagging jet with $p_T > 70(50)$ GeV in the 8(7) TeV dataset
	$\Delta R(\tau_1, \tau_2) < 1.9$
	$\not{E}_T > 20$ GeV
	if $\not{E}_T$ vector is not pointing in between the two taus, $\min\{\Delta\phi(\not{E}_T, \tau_1), \Delta\phi(\not{E}_T, \tau_2)\} < 0.1\pi$.

Table 7.4: Summary of the event selection and categories for the $H \to \tau_{\text{had}}\tau_{\text{had}}$ channel.

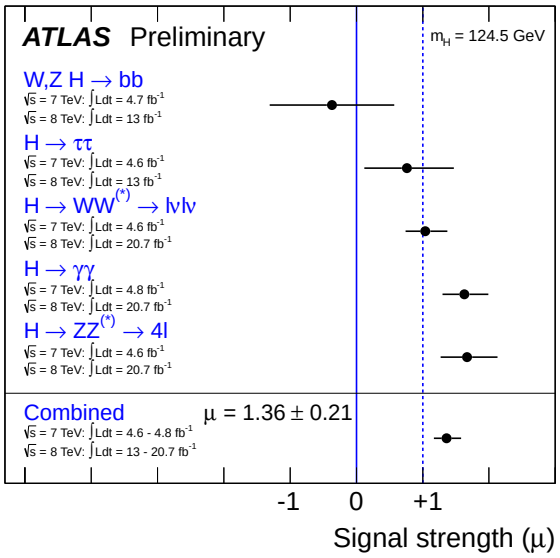
Higgs Decay Mode	$\hat{\mu}$ (m_H=125.5 GeV)
$VH \to Vbb$	-0.4 ± 1.0
$H \to \tau\tau$	0.8 ± 0.7
$H \to WW^{(*)}$	1.0 ± 0.3
$H \to \gamma\gamma$	1.6 ± 0.3
$H \to ZZ^{(*)}$	1.5 ± 0.4
Combined	1.30 ± 0.20

Table 7.5: Summary of the best-fit values and uncertainties for the signal strength μ for the individual channels and their combination at a Higgs boson mass of 125.5 GeV.

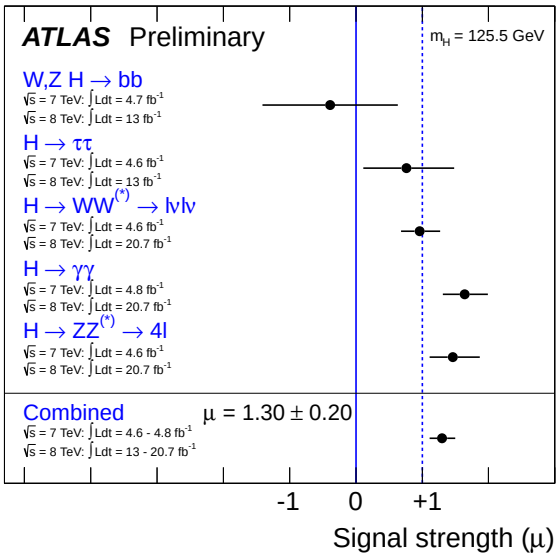
The measured global signal strengths in each individual channel as well as the combined is shown in Table 7.5 for $m_H = 125.5$ GeV. Figure 7.5 shows these for the masses 124.5, 125.5, and 126.5 GeV. Finally, the combined p_0 versus m_H can be seen in Figure 7.6.

7.2.1 Hypothesis Testing

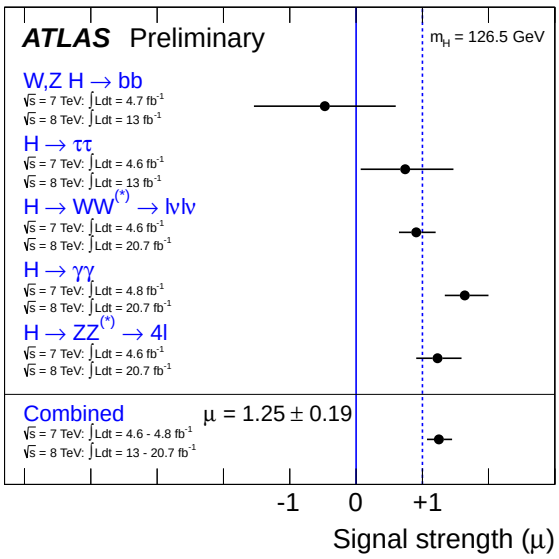
With the exception of the spin analyses, hypothesis testing is performed using the PLR as described in Chapter V. The test statistic used is $t_\mu^n = -2\ln\Lambda(\mu)$, where $\mu = \{\mu_1, ..., \mu_n\}$ are n parameters of interest. t_μ^n is asymptotically distributed as a χ^2 with n degrees of



(a) $m_H = 124.5$ GeV



(b) $m_H = 125.5$ GeV



(c) $m_H = 126.5$ GeV

Figure 7.5: $\hat{\mu}$ for the individual and combined channels for the three masses (a) 124.5, (b) 125.5, and (c) 126.5 GeV

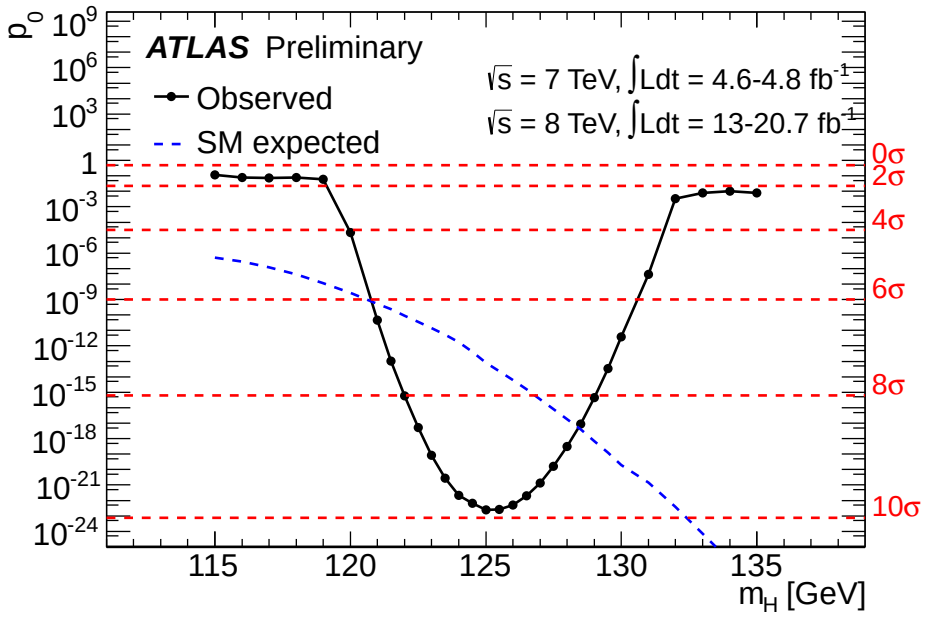


Figure 7.6: Expected and observed p_0 for the combined likelihood versus m_H.

freedom. For $n = 1$, the 68% and 95% confidence intervals are defined by $-2\ln\Lambda(\mu) < 1$ and $-2\ln\Lambda(\mu) < 4$, respectively, while for $n = 2$ the thresholds are 2.3 and 6.0.

7.2.2 Higgs Boson Mass

The two high resolution channels $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ are used to measure the mass [10]. The remaining channels have a low resolution and contribute negligibly to the mass sensitivity. The hypothesized mass m_H is the parameter of interest in the combined likelihood between the two channels. The signal strengths in each channel are treated as uncorrelated, though within each individual channel the production and decay modes use a common μ factor. The curve of $-2\ln\Lambda(m_H)$ for the two channels along with the combination of the two is shown in Figure 7.7.

A summary of the masses measured in the individual channels and in combination is shown in Table 7.6. The uncertainty is systematically dominated. In the $H \to \gamma\gamma$ channel, e/γ mass scale systematics dominate the measurement due to the two photons in the final state. The $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ channel includes approximately equal contributions from

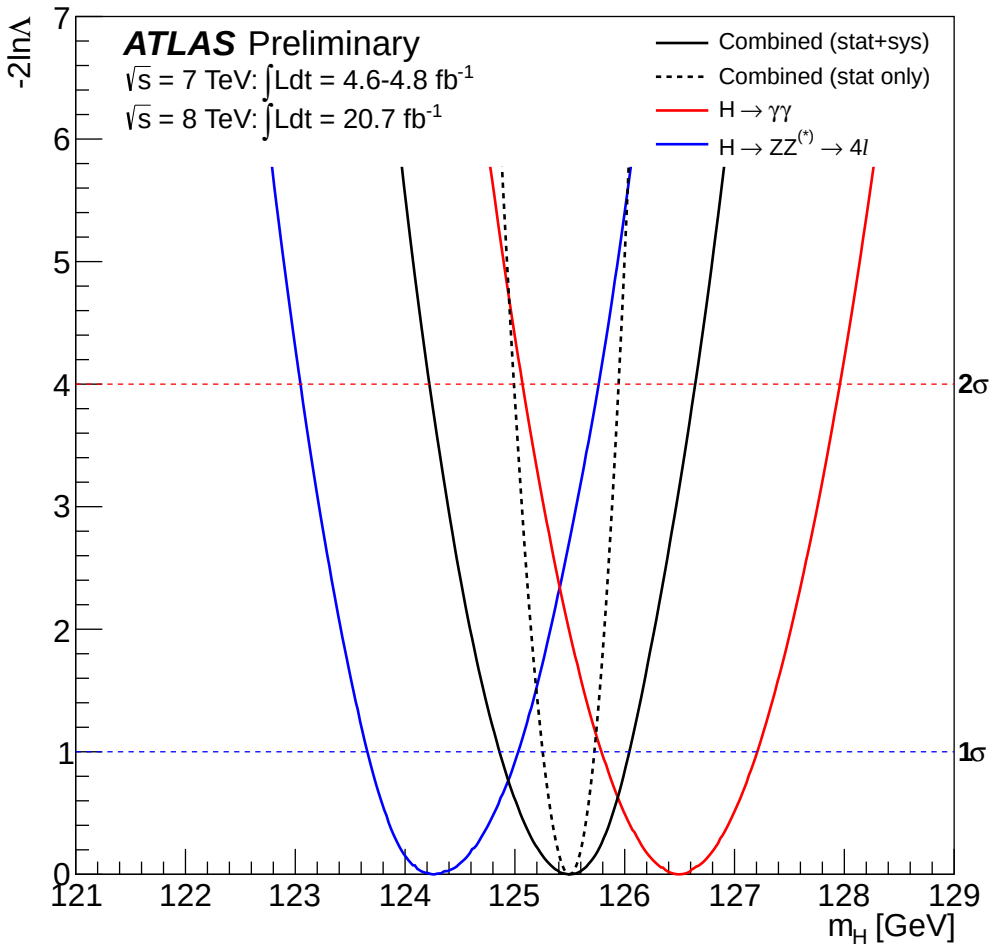


Figure 7.7: Likelihood versus tested mass for the $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ channels along with the combined curve in black.

Channel	Measured Mass (GeV)
$H \to ZZ^{(*)} \to \ell\ell\ell\ell$	$124.3^{+0.6}_{-0.5}$ (stat) $^{+0.5}_{-0.3}$ (sys)
$H \to \gamma\gamma$	126.8±0.2 (stat) ±0.7 (sys)
Combined	125.5±0.2 (stat) $^{+0.5}_{-0.6}$ (sys)

Table 7.6: Measured mass for both $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ and $H \to \gamma\gamma$ channels along with the combined mass measurement.

both electron and muon scale systematics, as the final state can include both electrons and muons. The two channels are therefore only partially correlated.

The discrepancy in the mass can be quantified by defining a parameter of interest $\Delta m \equiv m_{4\ell} - m_{\gamma\gamma}$. The curve $-2\ln\Lambda(\Delta m)$ versus Δm is shown in Figure 7.8. The value of $\sqrt{-2\ln\Lambda(\Delta m = 0)}$ asymptotically gives the significance of the discrepancy. This is observed to be 2.4 standard deviations, corresponding to a p-value 0.015. This was also tested with rectangular constraints on the mass scale systematics in place of the standard Gaussian constraints, which gives a p-value of 0.08, corresponding to 1.8 standard deviations. Further information can be extracted from the 2D contour of $-2\ln\Lambda(m_{4\ell}, m_{\gamma\gamma})$, shown in Figure 7.9.

7.2.3 Production and Decay

For the remaining property measurements [9], three assumptions are made about the boson. It is assumed that the particle is a CP even scalar, which is well motivated from the results in Ref. [16, 12]. The resonance is assumed to correspond to the same particle in each channel. Finally, the width is assumed to be narrow. In this way the production times decay can be factorized: $(\sigma \times \mathcal{B})(ii \to H \to jj) = \sigma_{ii} \cdot \frac{\Gamma_{ff}}{\Gamma_H}$.

It is useful to parameterize the number of signal events n^k_{Signal} in each channel k such that a parameter of interest μ scales each production and decay.

$$n^k_{\text{Signal}} = \left(\sum \mu_i \sigma_{i,SM} \times A^k_{if} \times \varepsilon^k_{if}\right) \times \mu_f \mathcal{B}_{f,SM} \times \mathcal{L}^k \tag{7.2}$$

$\sigma_i = \mu_i \sigma_{i,\text{SM}}$ is then the hypothesized production cross section for the process i, and

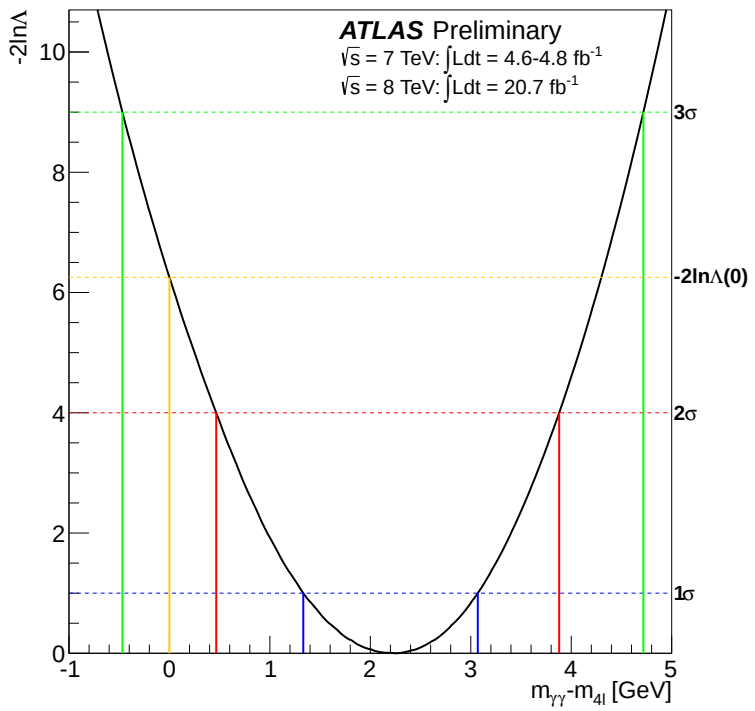


Figure 7.8: $-2\ln\Lambda$ versus Δm for the combined $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ likelihoods. The significance of the discrepency in measured mass between the two channels is indicated by the value of $-2\ln\Lambda(\Delta m = 0)$.

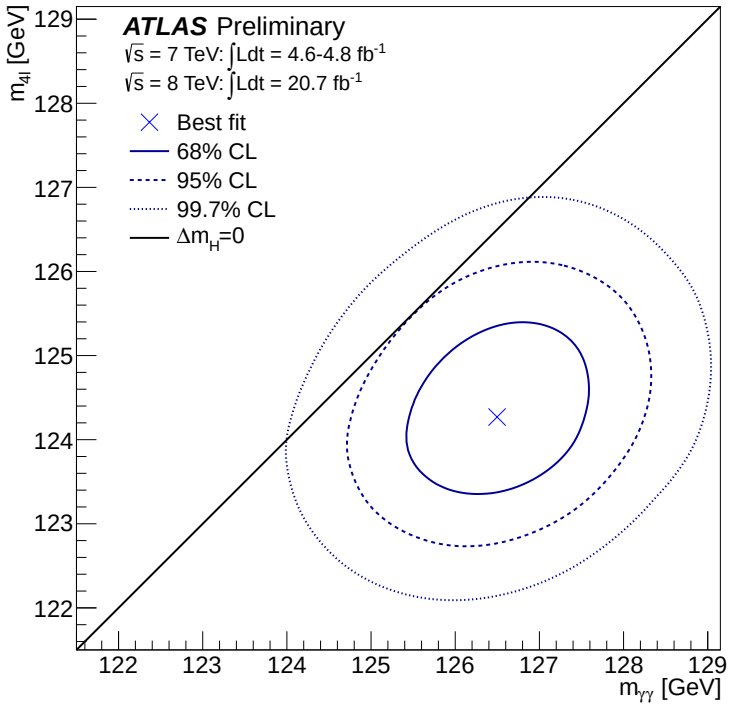


Figure 7.9: Contour of $-2\ln\Lambda$ for m_{4l} versus $m_{\gamma\gamma}$. The black line indicates the slice corresponding to $m_{4l} = m_{\gamma\gamma}$.

$\mathcal{B}_f = \mu_f \mathcal{B}_{f,\mathrm{SM}}$ is the hypothesized branching fraction for the decay mode f. The global SM strength parameter μ can be recovered by setting each product $\mu_i \mu_f$ to μ. This model with a global μ is not a well motivated BSM scenario, however, and fixing one or more μ_i or μ_f to its SM value may conceal tensions between SM and data, which contains in it the true values from nature.

A simple extension of the global μ is the division into ggF and VBF production processes. Because no channel offers substantial constraint on the $t\bar{t}H$ process and the contribution is small to all channels, this is assumed to scale like ggF. Similarly, the VH mode is assumed to scale like VBF (though this assumption is relaxed later). When considering the individual channels, the decay cannot be factorized from the production. The contour $\mu_{\mathrm{VBF+VH}} \times \frac{\mathcal{B}_f}{\mathcal{B}_{f,\mathrm{SM}}}$ versus $\mu_{\mathrm{ggF+t\bar{t}H}} \times \frac{\mathcal{B}_f}{\mathcal{B}_{f,\mathrm{SM}}}$ is therefore the simplest extension of the global model. This contour is shown in Figure 7.10 for each individual channel.

If one considers a slice through the contour defined by $\mu_{\mathrm{VBF+VH}} = \alpha \mu_{\mathrm{ggF+t\bar{t}H}}$, the parameter of interest becomes $\alpha = \frac{\mu_{\mathrm{VBF+VH}}}{\mu_{\mathrm{ggF+t\bar{t}H}}}$. The branching fraction for each channel cancels in this parameter and so a combined likelihood can be considered. This can be visualised in Figure 7.11.

The value of $-2\ln\Lambda$ as a function of $\frac{\mu_{\mathrm{VBF+VH}}}{\mu_{\mathrm{ggF+t\bar{t}H}}}$ is shown in Figure 7.12. The value of $\sqrt{-2\ln\Lambda}$ when $\frac{\mu_{\mathrm{VBF+VH}}}{\mu_{\mathrm{ggF+t\bar{t}H}}} = 0$ is the significance of production through vector bosons. Also shown is $-2\ln\Lambda$ as a function of $\frac{\mu_{\mathrm{VBF}}}{\mu_{\mathrm{ggF+t\bar{t}H}}}$, where the VH production process is profiled. In this case $\sqrt{-2\ln\Lambda}$ at $\frac{\mu_{\mathrm{VBF}}}{\mu_{\mathrm{ggF+t\bar{t}H}}} = 0$ is a statement purely about production through VBF. For the test of production through vector bosons, a significance of 3.3σ is observed, with an expected significance of 2.7σ. Similarly, the VBF only production of observed with 3.1σ significance and 2.5σ expected.

Similar to the statements purely about production, a parameter of interest can be formed that makes a statement purely about Higgs decay, as visualized in Figure 7.13 for $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ and $H \to \gamma\gamma$. With the parameter of interest as $\rho^{XX}_{YY} = \frac{\mathcal{B}(XX)\mathcal{B}_{\mathrm{SM}}(YY)}{\mathcal{B}_{\mathrm{SM}}(XX)\mathcal{B}(YY)}$, $\mu_{\mathrm{ggF+t\bar{t}H}} \times \mu_{\mathcal{B}(YY)}$ and $\frac{\mu_{\mathrm{VBF+VH}}}{\mu_{\mathrm{ggF+t\bar{t}H}}}$ are profiled in each channel. The curve $-2\ln\Lambda(\rho)$ for each

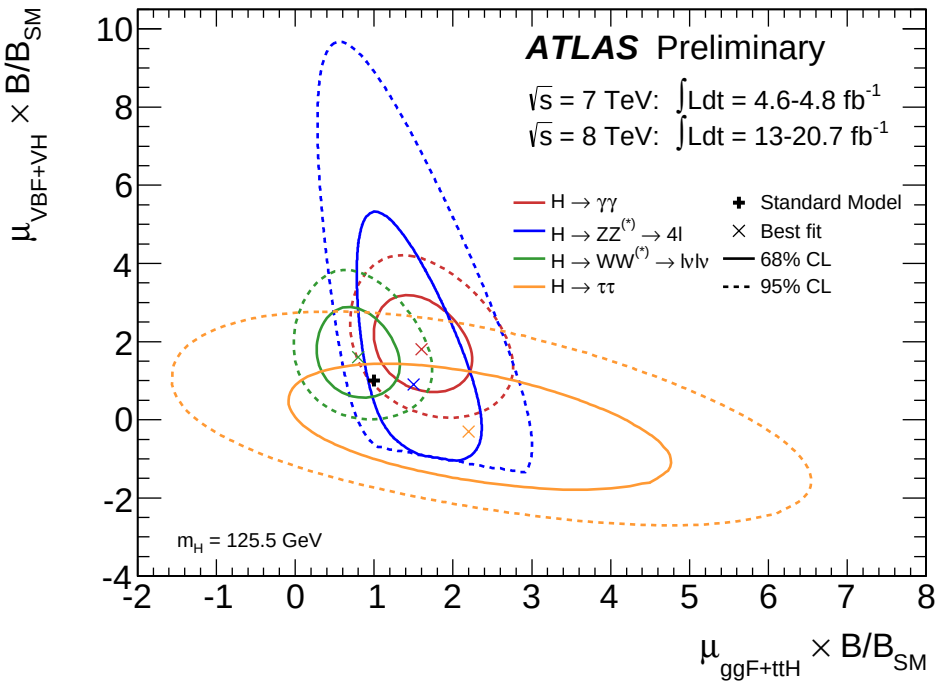


Figure 7.10: Likelihood contours for the $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to \ell\ell\ell\ell$, $H \to WW^{(*)} \to \ell\nu\ell\nu$, and $H \to \tau^+\tau^-$ channels in the $(\mu_{\mathrm{ggF}+t\bar{t}H}, \mu_{\mathrm{VBF}+VH})$ plane for a Higgs boson mass hypothesis of $m_H = 125.5$ GeV. Both $\mu_{\mathrm{ggF}+t\bar{t}H}$ and $\mu_{\mathrm{VBF}+VH}$ are modified by the branching ratio factors $\mathrm{B/B_{SM}}$, which are different for the different final states. The quantity $\mu_{\mathrm{ggF}+t\bar{t}H}$ ($\mu_{\mathrm{VBF}+VH}$) is a common scale factor for the gluon fusion and $t\bar{t}H$ (VBF and VH) production cross sections. The best fit to the data (×) and 68% (full) and 95% (dashed) CL contours are also indicated, as well as the SM expectation (+).

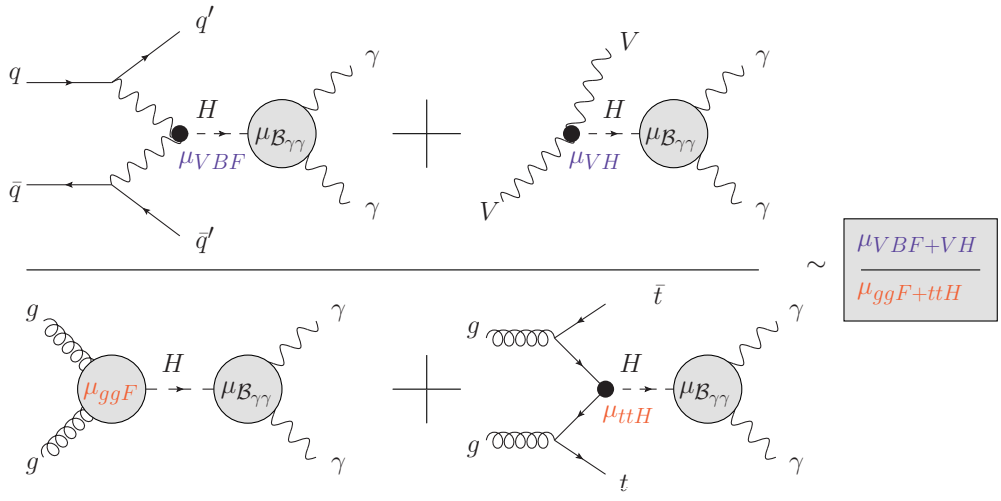
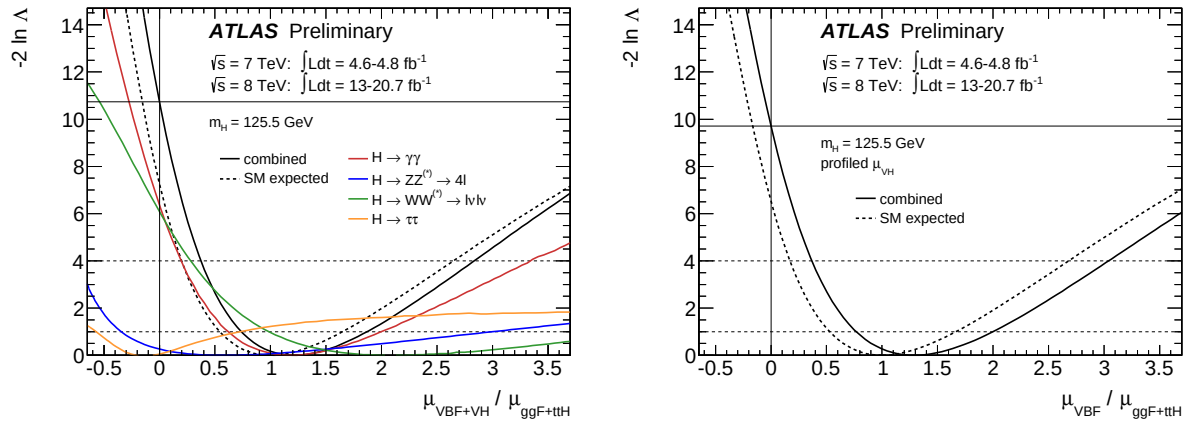


Figure 7.11: Visualization of the organization of the production scale factors into the ratio $\alpha = \frac{\mu_{\mathrm{VBF+VH}}}{\mu_{\mathrm{ggF+t\bar{t}H}}}$.



(a) (b)

Figure 7.12: Likelihood curves for the ratios (a) $\mu_{\mathrm{VBF}+VH}/\mu_{\mathrm{ggF}+t\bar{t}H}$ and (b) $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}+t\bar{t}H}$ for the $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to \ell\ell\ell\ell$, $H \to WW^{(*)} \to \ell\nu\ell\nu$, and $H \to \tau^+\tau^-$ channels and their combination for a Higgs boson mass hypothesis of $m_H = 125.5$ GeV. The branching ratios and possible non-SM effects coming from the branching ratios cancel in $\mu_{\mathrm{VBF}+VH}/\mu_{\mathrm{ggF}+t\bar{t}H}$ and $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}+t\bar{t}H}$, hence the different measurements from all four channels can be compared and combined. For the measurement of $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}+t\bar{t}H}$, the signal strength μ_{VH} is profiled. The dashed curves show the SM expectation for the combination. The horizontal dashed lines indicate the 68% and 95% confidence levels.

$\rho^{\gamma\gamma}_{ZZ}$, $\rho^{\gamma\gamma}_{WW}$, and ρ^{ZZ}_{WW} is shown in Figure 7.14.

7.2.4 Couplings

The leading order couplings with respect to the SM can be measured with a reparametrization of the likelihood in terms of detectable production and decay modes. Details and recommendations for this are outlined in Ref. [48]. The parametrization is shown explicitely in Table 7.7. Decay modes which are undetectable or which there is no sensitivity to must be assumed to scale as something similar to that which is detectable. For example, the partial width to $\mu^+\mu^-$ is assumed to scale like the width to $\tau^+\tau^-$, since they are both fermionic in nature. Two examples are represented by Figure 7.15. The first shows how the $gg \to H \to \gamma\gamma$ process scales in terms of detectable couplings. The second shows how the $gg \to t\bar{t}H \to t\bar{t}\mu^+\mu^-$ would have to be modified to scale with measurable parameters.

The effective loop vertices $gg \to H$ and $H \to \gamma\gamma$ can be written both as effective coupling

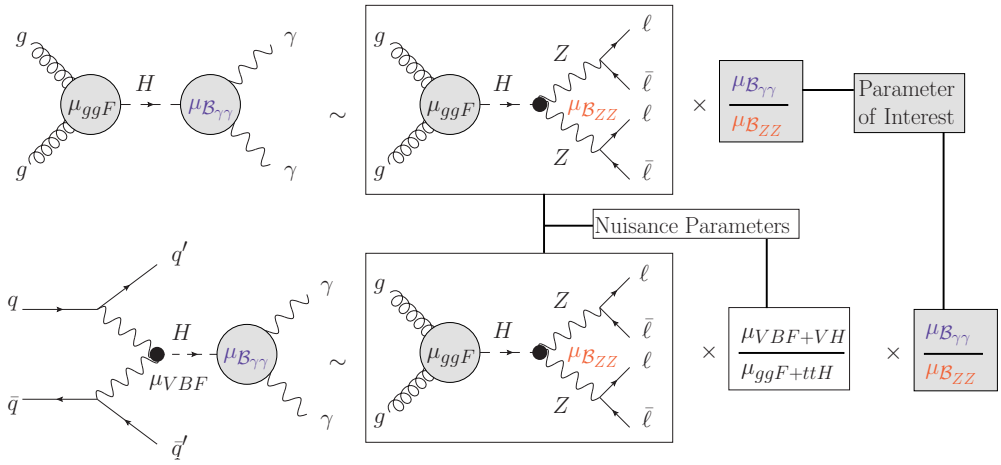


Figure 7.13: Visualization of the organization of decay modes into the parameter of interest $\rho^{\gamma\gamma}_{ZZ} = \frac{\mathcal{B}(\gamma\gamma)\mathcal{B}_{\mathrm{SM}}(ZZ)}{\mathcal{B}_{\mathrm{SM}}(\gamma\gamma)\mathcal{B}(ZZ)}$, as well as the profiled production processes.

scale factors κ_g and κ_γ, respectively, or it can be assumed that the particle content within the loops is known. When no assumption is made about the content, the effective couplings are profiled. The total width scale factor κ_H is treated similarly. Further, depending on the model considered, assumptions can be made about which parameters scale coherently with each other. Three primary models are considered. A description of these follows.

7.2.4.1 Fermion versus Vector Couplings

In the case that the fermion versus vector boson couplings are tested, it's assumed that all fermion couplings scale together as κ_F and all vector boson couplings scale together as κ_V. This can be divided into three submodels. In the first, it's assumed that only SM particles enter the production and decay. In this case, the functional form of κ_H and κ_γ are fixed. In the second submodel, the assumption about κ_H is relaxed and the parameter is profiled. The motivation of this is that the Higgs may couple substantially to non-SM particles that have not been detected, which would modify κ_H. In the third and final submodel, both κ_H and κ_γ are profiled. If the hypothetical non-SM particles in the second submodel exist and couple to photons, they may enter and modify the $H \to \gamma\gamma$ loop as well. Table 7.8 shows a summary of fermion versus vector coupling models considered.

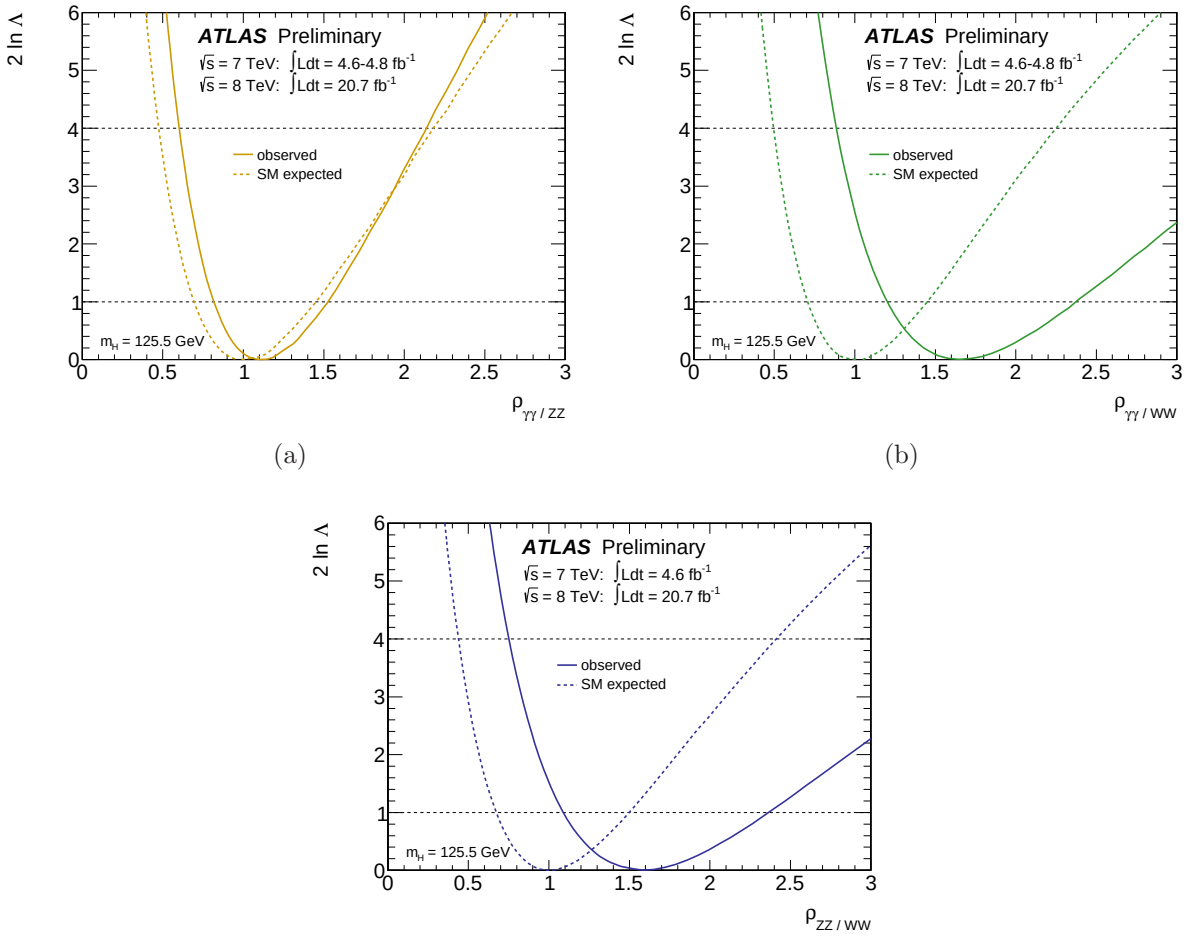
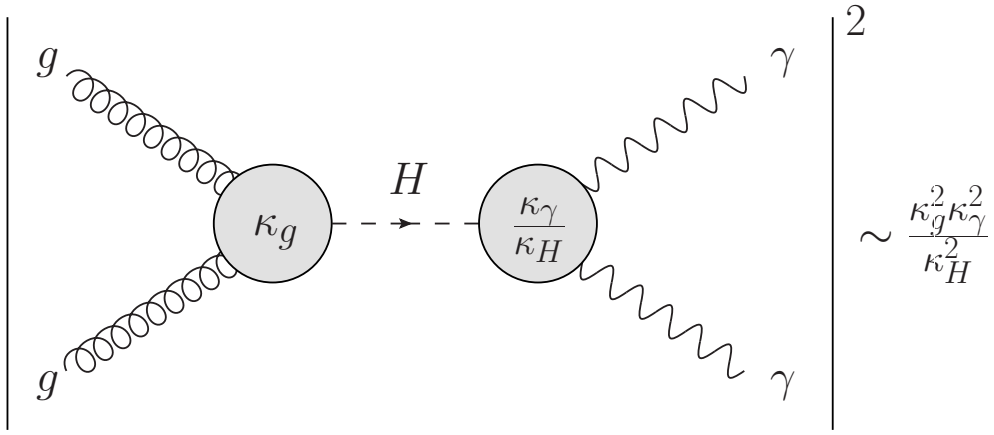
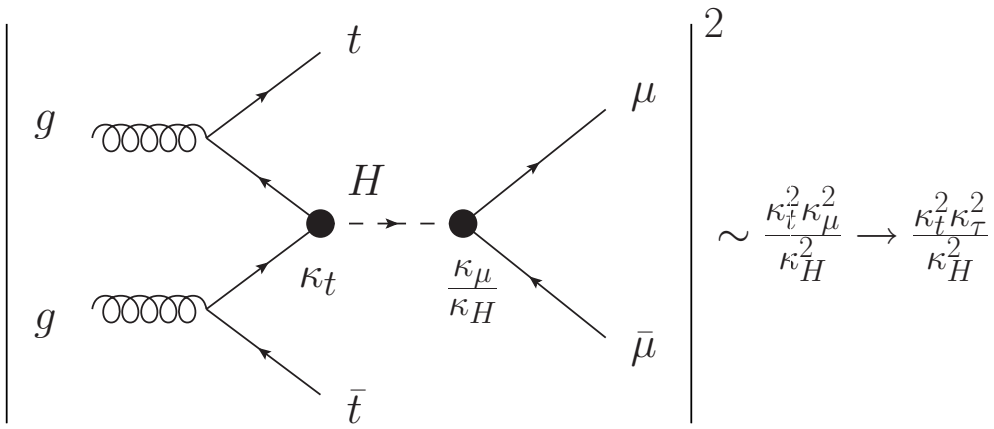


Figure 7.14: Likelihood curves for pairwise ratios of branching ratios normalized to their SM expectations (a) $\rho_{\gamma\gamma/ZZ}$, (b) $\rho_{\gamma\gamma/WW}$ and (c) $\rho_{ZZ/WW}$ of the $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to \ell\ell\ell\ell$, and $H \to WW^{(*)} \to \ell\nu\ell\nu$ channels, for a Higgs boson mass hypothesis of $m_H = 125.5$ GeV. The dashed curves show the SM expectation.



(a)



(b)

Figure 7.15: Visualization of (a) how the $gg \to H \to \gamma\gamma$ rate scales with coupling ratios, and (b) an example of how a currently undetectable mode is written in terms of a detectable one.

production processes

$$\frac{\sigma_{\mathrm{ggF}}}{\sigma^{\mathrm{SM}}_{\mathrm{ggF}}} = \begin{cases} \kappa_g^2(\kappa_b, \kappa_t, m_H) \\ \kappa_g^2 \end{cases}$$

$$\frac{\sigma_{\mathrm{VBF}}}{\sigma^{\mathrm{SM}}_{\mathrm{VBF}}} = \kappa^2_{\mathrm{VBF}}(\kappa_W, \kappa_Z, m_H)$$

$$\frac{\sigma_{\mathrm{WH}}}{\sigma^{\mathrm{SM}}_{\mathrm{WH}}} = \kappa_W^2$$

$$\frac{\sigma_{\mathrm{ZH}}}{\sigma^{\mathrm{SM}}_{\mathrm{ZH}}} = \kappa_Z^2$$

$$\frac{\sigma_{\mathrm{t\bar{t}H}}}{\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}}} = \kappa_t^2$$

Total width

$$\frac{\Gamma_H}{\Gamma_H^{\mathrm{SM}}} = \begin{cases} \kappa_H^2(\kappa_i, m_H) \\ \kappa_H^2 \end{cases}$$

Detectable decay modes

$$\frac{\Gamma_{\mathrm{WW}}}{\Gamma^{\mathrm{SM}}_{\mathrm{WW}}} = \kappa_W^2$$

$$\frac{\Gamma_{\mathrm{ZZ}}}{\Gamma^{\mathrm{SM}}_{\mathrm{ZZ}}} = \kappa_Z^2$$

$$\frac{\Gamma_{b\bar{b}}}{\Gamma^{\mathrm{SM}}_{b\bar{b}}} = \kappa_b^2$$

$$\frac{\Gamma_{\tau\tau}}{\Gamma^{\mathrm{SM}}_{\tau\tau}} = \kappa_\tau^2$$

$$\frac{\Gamma_{\gamma\gamma}}{\Gamma^{\mathrm{SM}}_{\gamma\gamma}} = \begin{cases} \kappa_\gamma^2(\kappa_b, \kappa_t, \kappa_\tau, \kappa_W, m_H) \\ \kappa_\gamma^2 \end{cases}$$

$$\frac{\Gamma_{Z\gamma}}{\Gamma^{\mathrm{SM}}_{Z\gamma}} = \begin{cases} \kappa^2_{(Z\gamma)}(\kappa_b, \kappa_t, \kappa_\tau, \kappa_W, m_H) \\ \kappa^2_{(Z\gamma)} \end{cases}$$

Currently undetectable decay modes

$$\frac{\Gamma_{t\bar{t}}}{\Gamma^{\mathrm{SM}}_{t\bar{t}}} = \kappa_t^2$$

$$\frac{\Gamma_{gg}}{\Gamma^{\mathrm{SM}}_{gg}} = \begin{cases} \kappa^2_{(H\to gg)}(\kappa_b, \kappa_t, m_H) \\ \kappa_g^2 \end{cases}$$

$$\frac{\Gamma_{c\bar{c}}}{\Gamma^{\mathrm{SM}}_{c\bar{c}}} = \kappa_t^2$$

$$\frac{\Gamma_{s\bar{s}}}{\Gamma^{\mathrm{SM}}_{s\bar{s}}} = \kappa_b^2$$

$$\frac{\Gamma_{\mu\mu}}{\Gamma^{\mathrm{SM}}_{\mu\mu}} = \kappa_\tau^2$$

Table 7.7: Parametrization of the production cross section, partial width, and total width with respect to their SM values into coupling ratio parameters κ_i. The undetectable modes are written in terms of similar detectable ones. The total width and effective coupling ratios can be written as both stand-alone parameters and as functions of other coupling ratios.

Fermion versus Vector Coupling Models

Model	Free Parameters
SM Particles Only	$\kappa_F (= \kappa_t = \kappa_b = \kappa_\tau = \kappa_g)$ $\kappa_V (= \kappa_W = \kappa_Z)$
Free Total Width	$\lambda_{FV} = \frac{\kappa_F}{\kappa_V} \leftarrow$ POI $\kappa_{VV} = \frac{\kappa_V^2}{\kappa_H}$
Free Total Width + Free $\gamma\gamma$ loop	$\lambda_{FV} = \frac{\kappa_F}{\kappa_V} \leftarrow$ POI $\kappa_{VV} = \frac{\kappa_V^2}{\kappa_H}$ $\lambda_{\gamma V} = \frac{\kappa_\gamma}{\kappa_V}$

Table 7.8: Summary of free parameters for the three submodels within the $\kappa_F - \kappa_V$ model.

In the submodel with no additional SM particles, the two parameters κ_F and κ_V are treated independently. The likelihood scan for each is shown in Figure 7.16 along with the 2D contour of κ_F versus κ_V. The second minimum in the κ_F-κ_V plane is due to the fermion-vector boson interference in the $H \to \gamma\gamma$ loop, shown in Figure 7.17. The fermions and bosons enter the loop with opposite signs, such that the matrix element squared gives a polynomial with a negative-signed term.

In the submodels where the assumptions are relaxed, the ratio $\lambda_{FV} = \frac{\kappa_F}{\kappa_V}$ is used to scan the likelihood. The results of this are shown in Figure 7.18. When κ_γ is profiled, the degeneracy in λ_{FV} is no longer broken as the interference term in the $H \to \gamma\gamma$ loop is not present explicitely in the likelihood. In this case the likelihood is arbitrary up to the sign of λ_{FV} since both κ_F and κ_V enter as squared everywhere.

7.2.4.2 Custodial Symmetry

Custodial symmetry [36] is highly constrained in the electroweak sector by both theory and experiment. It is therefore interesting to test this in the Higgs sector. For this, the vector couplings to W and Z bosons are separated, and the parameter $\lambda_{WZ} = \kappa_W/\kappa_Z$ is tested. This is tested in two submodels, outlined in Table 7.9. In the first, the total width and couplings to fermions are profiled. In the second, the effective coupling κ_γ is also profiled. The results for the two submodels are shown in Figure 7.19.

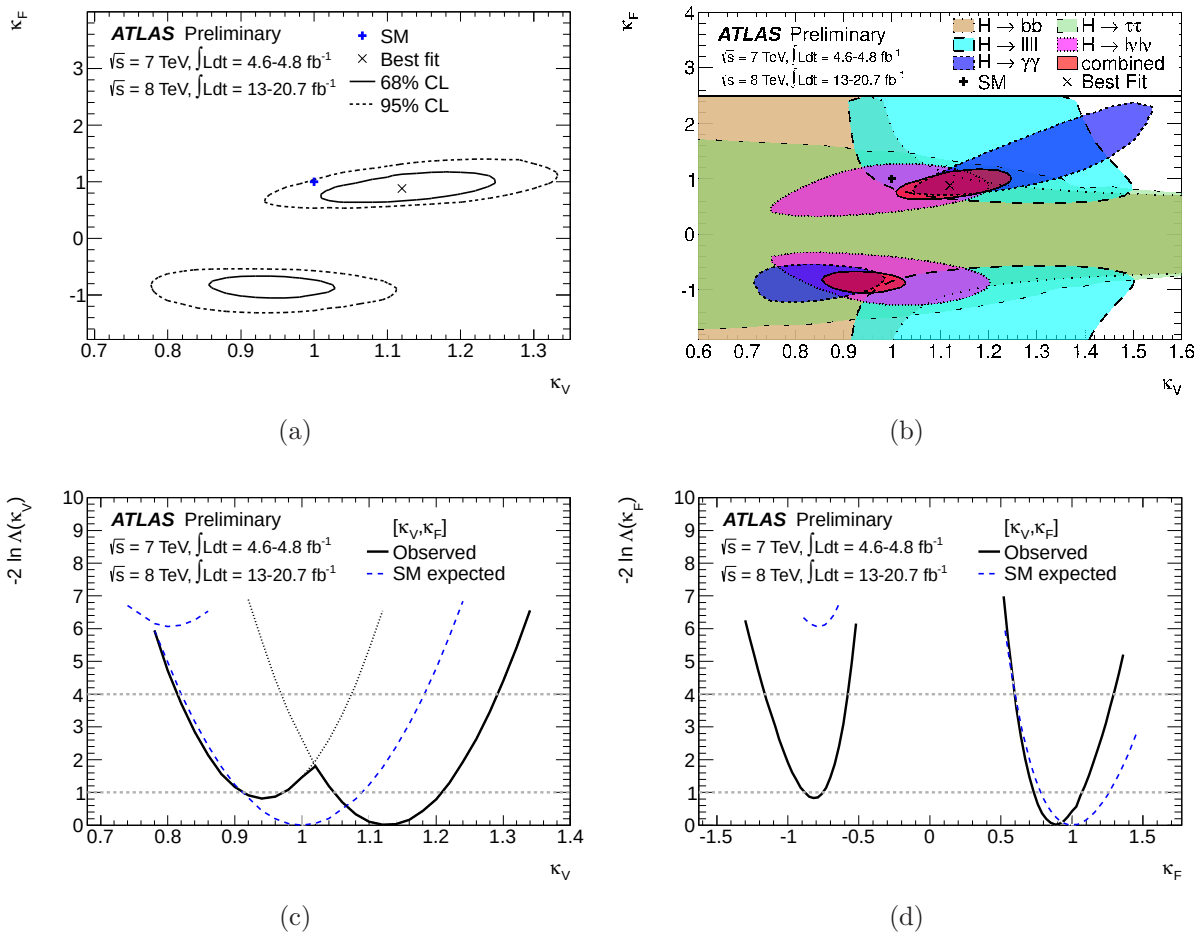


Figure 7.16: Fits for 2-parameter benchmark models probing different coupling strength scale factors for fermions and vector bosons, assuming only SM contributions to the total width: (a) Correlation of the coupling scale factors κ_F and κ_V; (b) the same correlation, overlaying the 68% CL contours derived from the individual channels and their combination; (c) coupling scale factor κ_V (κ_F is profiled); (d) coupling scale factor κ_F (κ_V is profiled). The dashed curves in (c) and (d) show the SM expectation. The thin dotted lines in (c) indicate the continuation of the likelihood curve when restricting the parameters to either the positive or negative sector of κ_F.

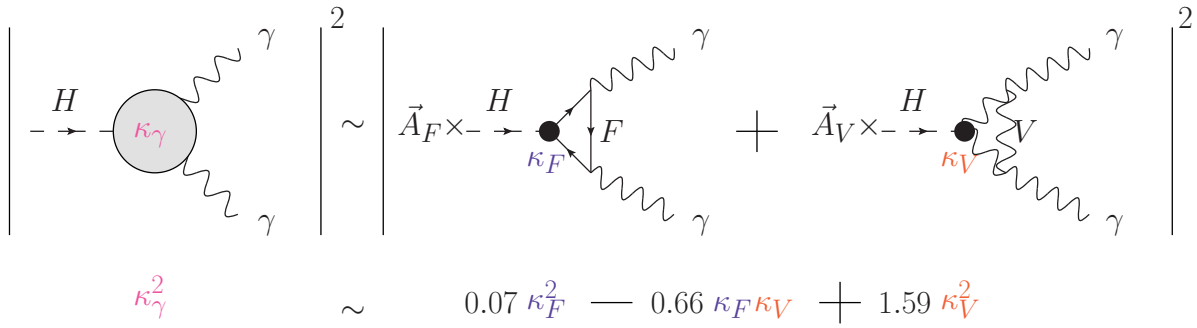


Figure 7.17: Visualization of the interference between fermions and bosons in the $H \to \gamma\gamma$ decay loop that gives rise to the two minima in the $\kappa_F - \kappa_V$ model.

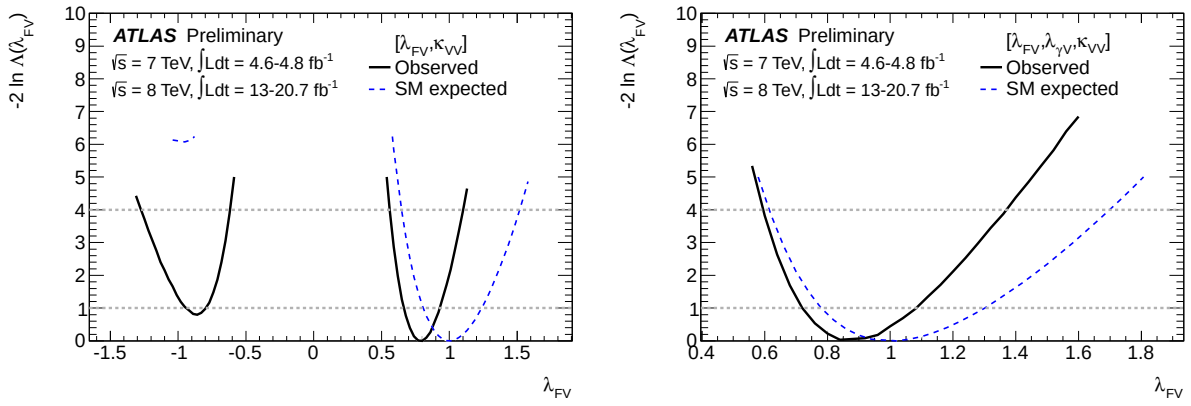
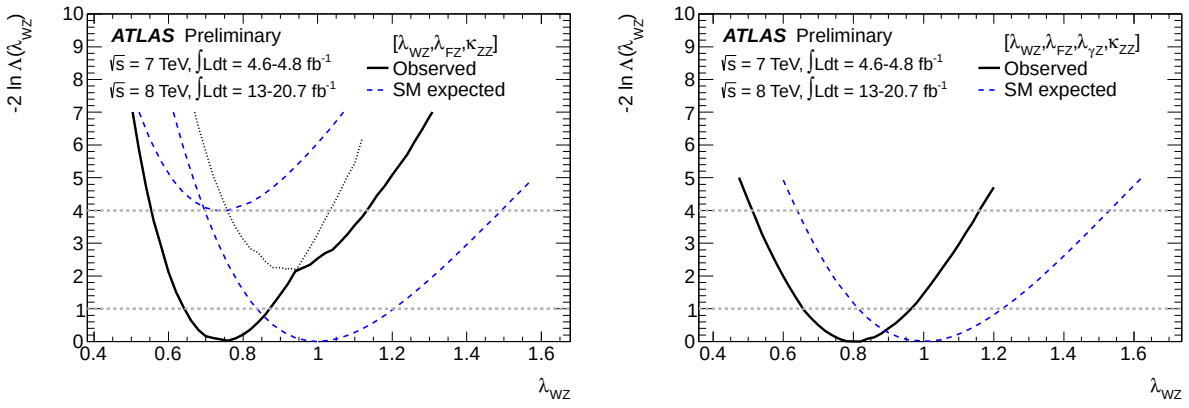


Figure 7.18: Likelihood curves for $\lambda_{FV} = \frac{\kappa_F}{\kappa_V}$ the fermion-vector boson model where (a) the assumption on the particle content in the total width is relaxed and (b) the assumptions on both the total width and the $H \to \gamma\gamma$ decay loop are relaxed. The second minimum disappears in (b) since the effective coupling κ_γ that gives rise to it is profiled.

Custodial Symmetry Models

Model	Free Parameters
Free Total Width	$\lambda_{WZ} = \frac{\kappa_W}{\kappa_Z} \leftarrow$ POI $\kappa_{ZZ} = \frac{\kappa_Z^2}{\kappa_H}$ $\lambda_{FZ} = \frac{\kappa_F}{\kappa_Z}$
Free Total Width + Free $\gamma\gamma$ loop	Above, + $\lambda_{\gamma Z} = \kappa_\gamma / \kappa_Z$

Table 7.9: Summary of free parameters for the two submodels within the $\kappa_W - \kappa_Z$ model that tests custodial symmetry in the Higgs sector.



(a) (b)

Figure 7.19: Fits for benchmark models probing the custodial symmetry through the ratio $\lambda_{WZ} = \kappa_W/\kappa_Z$ in two scenarios: (a) no assumption on the total width (b) additionally, no assumption on the $H \to \gamma\gamma$ decay loop. The dashed curves show the SM expectation. The thin dotted lines in (a) indicate the continuation of the likelihood curve when restricting the parameters to either the positive or negative sector of λ_{FZ}.

7.2.4.3 Probing Non-SM Contributions

The final model aims to probe potential couplings to non-SM particles. These can enter in the effective loops κ_g and κ_γ if the non-SM particles couple to gluons or photons. If they have a mass less than half the Higgs mass, they can substantially modify κ_H as well. There is therefore motivation to test the contributions in two scenarios. The first assumes that the non-SM particles can enter κ_g or κ_γ but do not satisfy the criteria to substantially modify κ_H. The second allows for the possibility of the latter scenario, and so κ_H is reparametrized

Vertex Loop Models

Model	Free Parameters	Example
SM only in Total Width	κ_g κ_γ	$\sigma(gg \to H) \cdot \mathcal{B}(H \to \gamma\gamma) \sim \frac{\kappa_g^2 \kappa_\gamma^2}{0.085\kappa_g^2 + 0.0023\kappa_\gamma^2 + 0.91}$
Free Total Width	κ_g κ_γ $\mathrm{BR}_{\mathrm{inv.,undet.}}$	$\sigma(gg \to H) \cdot \mathcal{B}(H \to \gamma\gamma) \sim \frac{\kappa_g^2 \kappa_\gamma^2 \cdot (1 - \mathrm{BR}_{\mathrm{inv.,undet.}})}{0.085\kappa_g^2 + 0.0023\kappa_\gamma^2 + 0.91}$

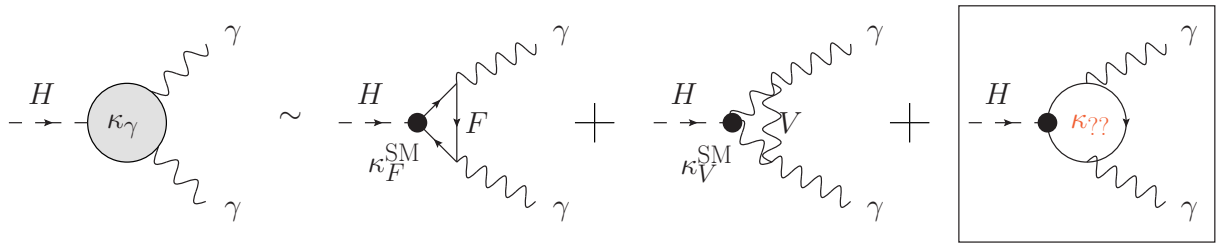
Table 7.10: Summary of free parameters for the two submodels that test for additional non-SM particles interacting with the Higgs in the scenarios that the total width both is and is not substantially affected by the new particles.

terms of the branching fraction to these invisible or undetected particles, $\mathrm{BR}_{\mathrm{inv.}}^{\mathrm{undet.}}$. Table 7.10 shows the details of parametrization.

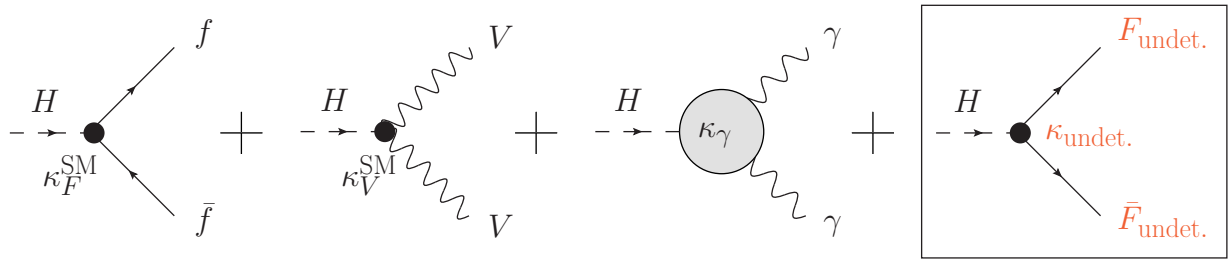
Since the Higgs itself is assumed to be SM, all non-loop coupling ratios κ_i are fixed to their SM value of unity. For the first submodel, the likelihoon scans of κ_g and κ_γ are shown in Figure 7.21 along with their 2D contour. The likelihood scan of $\mathrm{BR}_{\mathrm{inv.}}^{\mathrm{undet.}}$ in the second submodel is shown in Figure 7.22.

7.2.5 Summary

Table 7.11 summarizes the results of the global μ model, mass, and coupling measurements. Figure 7.23 further shows the summary of the coupling measurements centered around their corresponding SM expectation. All measurements agree within 2σ of the expected SM value. Because all results are based on the same data and are a simple reparametrization of the same likelihood, all deviations are highly correlated, so the plot should not be used in a χ^2 compatibility of the Higgs with the SM.



(a)

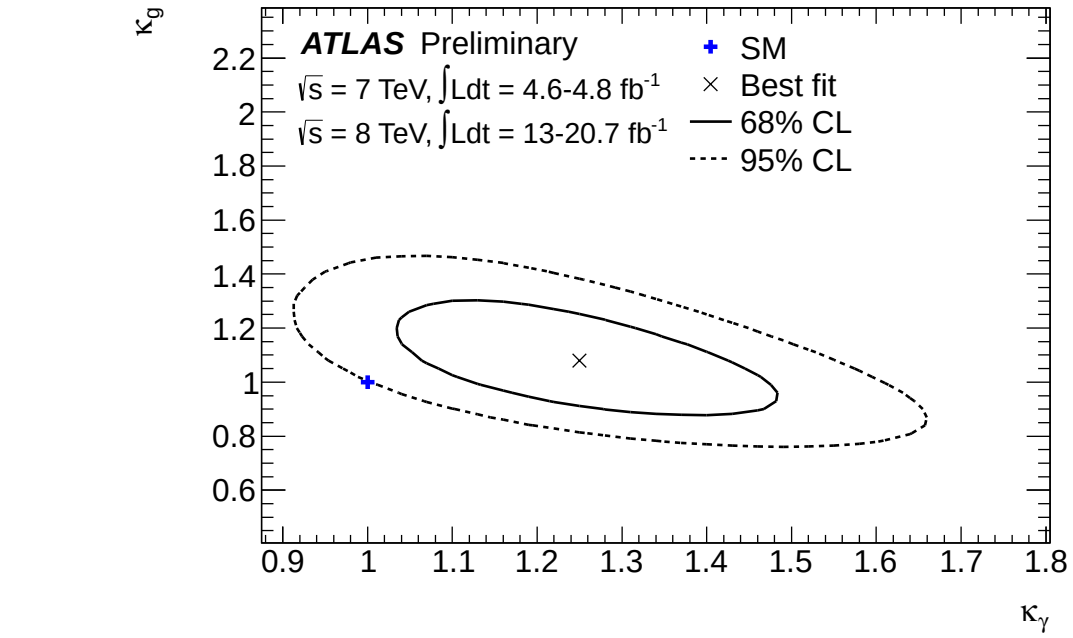


(b)

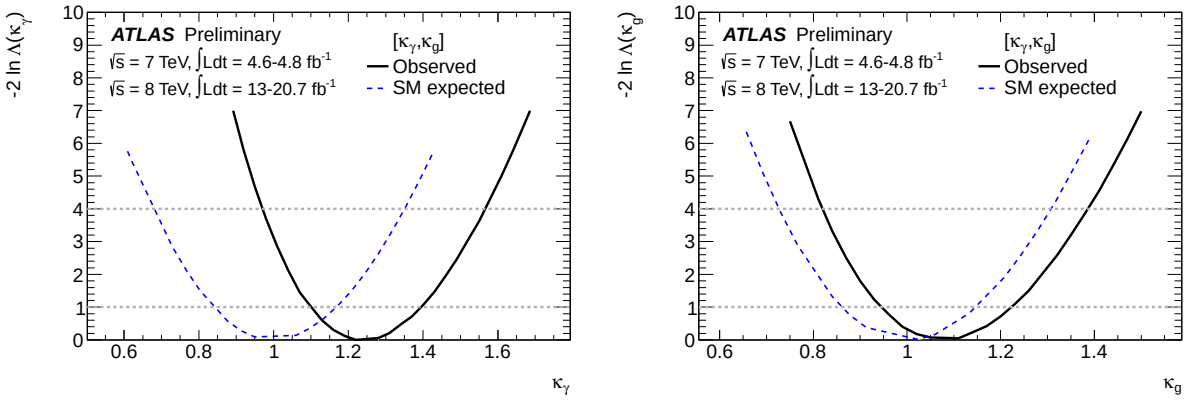
Figure 7.20: Visualization of non-SM scenarios that modify (a) the $H \to \gamma\gamma$ decay loop and (b) the total width of the Higgs

Global	Production	Decay
$\hat{m}_H = 125.5^{+0.5}_{-0.6}$ GeV $\hat{\mu} = 1.3 \pm 0.2$	$Z_{VBF+VH} = 3.3$ (2.7 exp) $Z_{VBF} = 3.1$ (2.5 exp)	$\rho_{ZZ}^{\gamma\gamma} = 1.1^{+0.4}_{-0.3}$ $\rho_{WW}^{\gamma\gamma} = 1.7^{+0.7}_{-0.5}$ $\rho_{ZZ}^{WW} = 1.6^{+0.8}_{-0.5}$
	Couplings	
Vector/Fermion	SM Particles	$\kappa_V \in [-0.88, -0.75] \cup [0.73, 1.07]$ $\kappa_F \in [0.91, 0.97] \cup [1.05, 1.21]$
	+Free Γ_H	$\lambda_{VF} \in [-0.94, -0.8] \cup [0.67, 0.93]$
	+Free $\Gamma_{\gamma\gamma}$	$\lambda_{VF} = 0.85^{+0.23}_{-0.13}$
Custodial Sym	Free Γ_H	$\lambda_{WZ} \in [0.64, 0.87]$
	+Free $\Gamma_{\gamma\gamma}$	$\lambda_{WZ} = 0.80 \pm 0.15$
Loop Vertices	Non-loop $\kappa_i = 1$	$\kappa_g = 1.08 \pm 0.14$ $\kappa_\gamma = 1.23^{+0.16}_{-0.13}$
	Inv./Undet. Decays	$\text{BR}_{\text{inv.}}^{\text{undet.}} < 0.6$ (95% CL)

Table 7.11: Summary of property measurements of the Higgs.



(a)



(b) (c)

Figure 7.21: Fits for benchmark models probing contributions from non-SM particles in the $H \to \gamma\gamma$ and ggF loops, assuming no sizeable extra contributions to the total width: (a) correlation of the coupling scale factors κ_γ and κ_g; (b) coupling scale factor κ_γ (κ_g is profiled); (c) coupling scale factor κ_g (κ_γ is profiled). The dashed curves in (b) and (c) show the SM expectation.

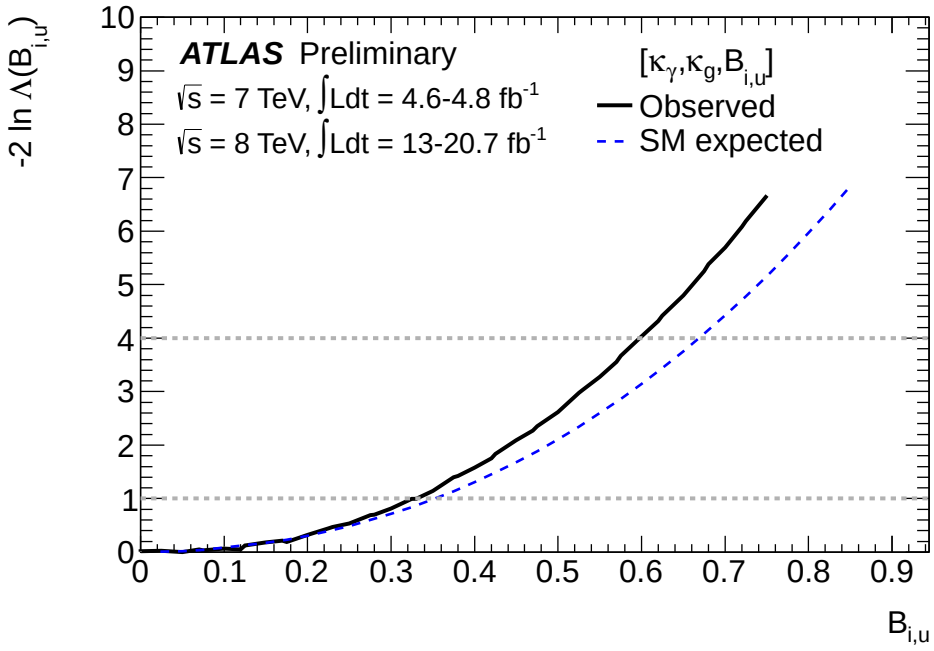


Figure 7.22: Fit for benchmark model probing contributions from non-SM particles in the total width. The dashed curves show the SM expectation.

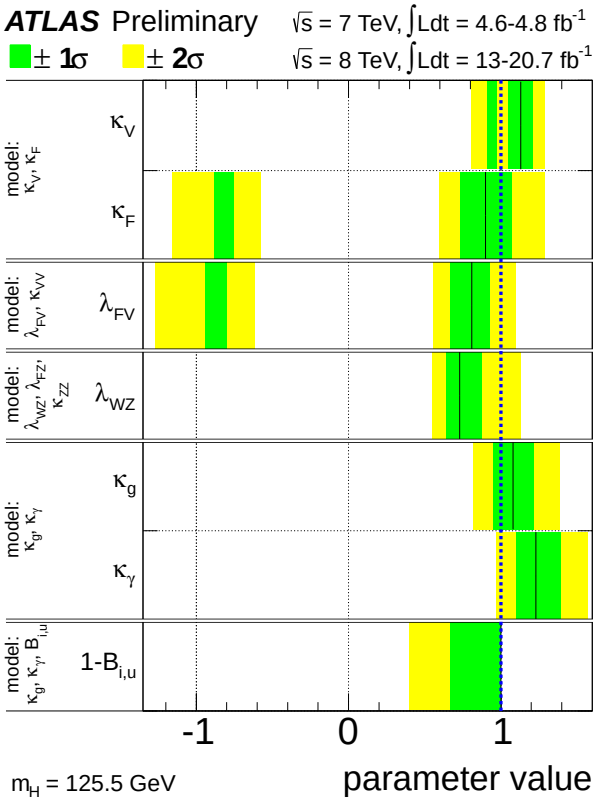


Figure 7.23: Summary plot showing measurements of the interesting parameters for all models probing the Lagrangian structure of the Higgs. Most deviations are largely correlated as they are reparametrizations of each other.

CHAPTER VIII

Conclusion

"Life is pretty simple: You do some stuff. Most fails. Some works. You do more of what works. If it works big, others quickly copy it. Then you do something else. The trick is the doing something else."

- Leonardo da Vinci

This dissertation (the length by date of which can be seen in Figure 8.1) has outlined the analysis of the ATLAS data that led to the discovery of a Higgs boson and the measurement of its properties. All properties measured as well as tests under BSM conditions show consistency with a SM Higgs boson. The mass of the boson is measured to be $m_H = 125.5 \pm 0.2$ (stat) $^{+0.5}_{-0.6}$ (sys) GeV, while the signal rate with respect to the SM is measured to be $\hat{\mu} = 1.30 \pm 0.20$ at this mass. The details of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis are elaborated upon as well, where a statistical significance of 3.8σ is observed, with 3.7σ expected at $m_H = 125$ GeV.

An unfortunate outcome of adding a Higgs to the SM and nothing else is that the particle suffers from a quadratic ultraviolet divergence in mass resulting from fermion and gauge boson loop diagrams. This would effectively bring the mass of the Higgs to the Planck scale unless there is an extreme fine-tuning of SM parameters. An elegant solution to this fine-tuning problem is to introduce an additional symmetry, dubbed Supersymmetry, between fermions and bosons that causes the loop diagrams to cancel, bringing the quadratic diver-

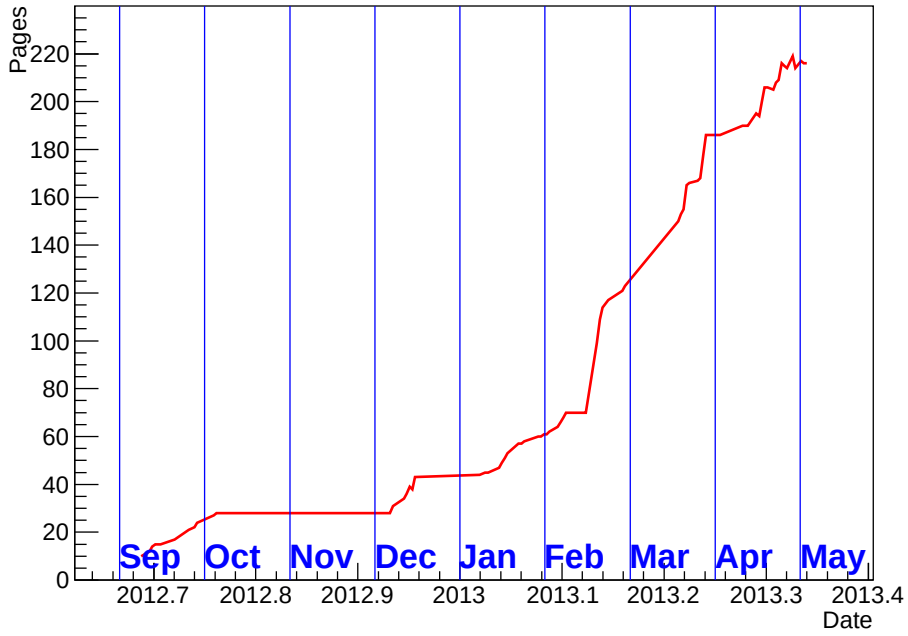


Figure 8.1: Length of this dissertation in pages by date. The periods of zero slope include efforts towards the Hadron Collider Physics conference and HSG3 workshop in Lisbon, Christmas break, and a mini-statistics worshop. Certain periods of heightened activity originate from mass figure additions and realizations of impending doom.

gence down to a logarithmic and renormalizable one. A convenient outcome of this addition is that it is also able to provide a dark matter candidate particle with properties consistent within current observations and limits [39]. This would also indicate that the new boson is only one of five supersymmetric Higgs bosons. One of the goals of the LHC program in the coming years, as well as that of many other physics and astrophysics experiments, will be to either exclude or substantiate this possibility; or, perhaps, something unexpected will be found that will revolutionize the way we think about particle physics.

APPENDICES

APPENDIX A

Appendix A: B-Spline Interpolation

All low resolution channels in the Higgs analysis use discretely binned likelihoods with signal distributions computed at fixed mass points. The spacing of these points is several times larger than the high mass resolution of the $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ channels, which use parametric signal models. Because of this, histogram interpolation needs to be employed to obtain the signal distributions required to test signal hypotheses at arbitrary points. This section outlines a method using B-Splines that can be used to provide analytical interpolation for the low resolution channels.

A.1 Introduction to B-Splines

B-Splines (short for basis splines) are parametric curves constructed from a set of order n basis polynomials $b_{i,n}(t)$ and $m+1$ knots t_i, with $t_0 \leq t_1 \leq ... \leq t_m$. A set of control points $\mathbf{P_i}$ defined over a subset of the knots $t \in [t_n, t_{m-n-1}]$, which are the points we wish to interpolate between, multiply the basis polynomials to obtain the B-Spline curve $\mathbf{S}(t) = \sum_{i=0}^{m-n-2} \mathbf{P}_i b_{i,n}(t)$.

The polynomials $b_{i,n}(t)$ are defined recusively:

$$\begin{aligned} b_{i,0}(t) &= \begin{cases} 1, & t_i \leq t < t_{i+1} \\ 0, & \text{otherwise} \end{cases}, \qquad i = 0, ..., m-2 \\ b_{i,n}(t) &= \frac{t-t_i}{t_{i+n}-t_i} b_{i,n-1}(t) + \frac{t_{i+n+1}-t}{t_{i+n+1}-t_{i+1}} b_{i+1,n-1}(t), \quad i = 0, ..., m-n-2 \end{aligned} \tag{A.1}$$

Note that the basis polynomials are dependent only on the spacing between the knots and are independent of any overall scaling between them. They are also independent of the control points. Figure A.1 shows an example interpolation of the gg→H production cross section with the basis polynomials overlayed. B-Splines have several important properties:

- Each control point P_i affects S(t) on the interval $[t_i, t_{i+n+1})$.
- S(t) is n-times differentiable at each knot.
- Given a set of splines $S_j(t) = \sum_{i=0}^{m-n-1} P_{ij} b_{i,n}(t)$, if $\sum_{j=0} P_{ij} = 1$ for each i, then $\sum_{j=0} S_j(t) = 1$ for all t (partition of unity property). This is important for preserving probability during PDF interpolation.
- A B-Spline of degree 1 linearly interpolates each P_i.
- S(t) is positive definite as long as each P_i is also positive.

A.2 B-Splines for Histogram Interpolation

B-Splines can be used to interpolate each individual bin of a histogram between mass hypotheses (mass-wise interpolation). This not equivalent to interpolating the distribution being binned in the histogram (observable interpolation), which would require an expensive integration to model within the likelihood. Thus, the likelihood is still binned with the j^{th} bin represented by a B-Spline curve $S_j(m_H)$ parametrized in m_H. In this case, the knots t_i become the coarse mass points $m_{H,i}$ at which the signal distributions are available.

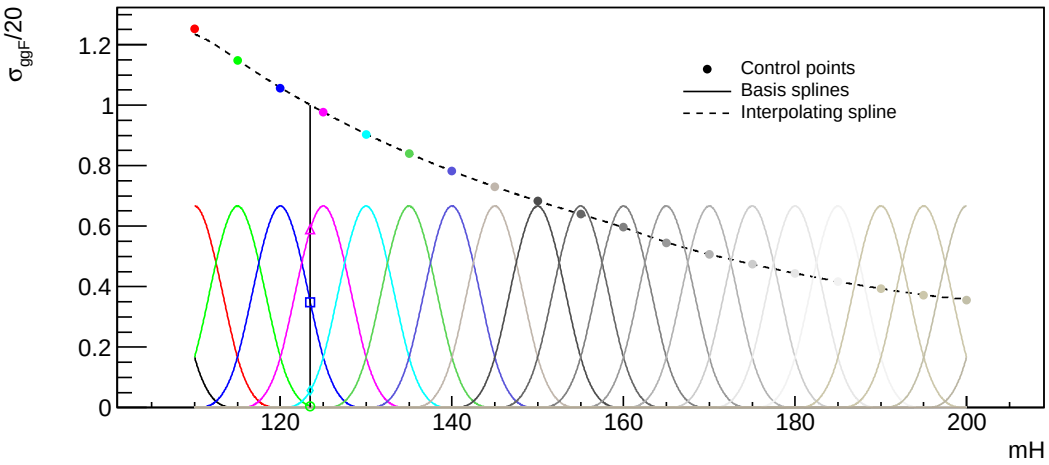


Figure A.1: B-Spline interpolation of the $gg \to H$ production cross section versus m_H. The colored points are the control points representing the cross section (scaled by 20) that enter the spline. The colored curves are the basis polynomials that multiply the control points of their corresponding color, which, when summed, give the interpolating spline.

One option for the control points P_{ij} is the expected events in bin E_{ij}, however it's more convenient to interpolate the total expected events $E_j = \sum_i^{bins} E_{ij}$ and normalized probability densities $p_{ij} = \frac{E_{ij}}{E_j}$ separately. This is because the latter will take advantage of the partition of unity property, while also being more physically motivated. The cross section, branching fractions, and acceptance are easier to model in a dedicated normalization spline, while the pure signal shape can be dedicated to the PDF splines.

To better understand the features of this interpolation scheme, a simple gaussian testbed is used. Figure A.2 shows the histogrammed distributions of a gaussian with various hypothesized mean values α_i and width 0.2. α_i is equivalent to the fixed $m_{H,i}$ points that we have in mind to interpolate between. Figure A.3 shows the spline interpolation of each bin in Figure A.2. There are two important notes. The first is the difference in the axes. The x-axis in Figure A.2 represents the observable of the distribution being binned. This is different from the x-axis of Figure A.3, which is the hypothesized mean of the gaussian. The second, which follows from the first, is that the expected events of each bin in Figure A.2 is represented as a curve in Figure A.3. The control points for each curve in Figure A.3 is

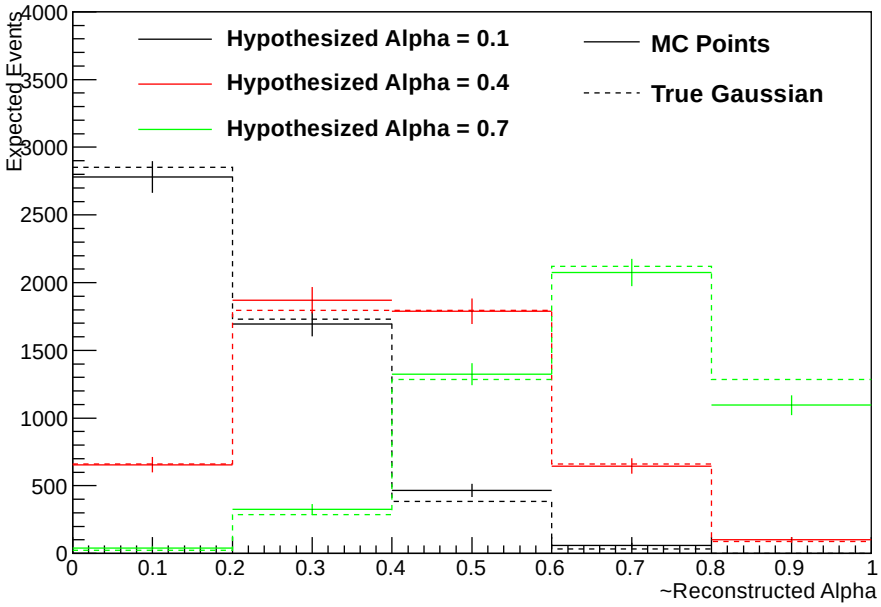


Figure A.2: Histogrammed gaussian distribution for various mean values α versus the reconstructed (or observable) α, all with width 0.2.

thus the bin content for each hypothesized α_i in Figure A.2. Figure A.3 shows two spline interpolations. One is for a high order spline ($n = N_\alpha - 2$), and one is for a low order spline ($n = 3$). The spline tends to underestimate the curvature of the distribution, with the effect increasing with the order of the spline. This bias will propagate to results obtained from the model if left untreated. To alleviate the bias, an iterative reweighting sequence $P_{ij}^{n+1} = P_{ij}^0 - (S_{ij}^n(\alpha_i) - P_{ij}^0)$ is applied to the control points in order to aid the spline in passing closer to the control points. This sequence preserves the partition of unity property and naturally converges towards $S_{ij}^N(\alpha_i) = P_{ij}^0$ for sufficiently large N. For the gaussian model, $N = 5$ provides sufficient convergence.

To understand the effects of limited MC statistics on the spline, the histograms are filled with entries drawn from true gaussian distributions. The number of entries is further drawn from a Poisson distribution with a mean of 5,000. For this test, the interpolation factorizes the normalization component of the spline from the shape component. Figures A.4- A.5 shows the interpolation splines for different initial seeds. One can see that the reweighting

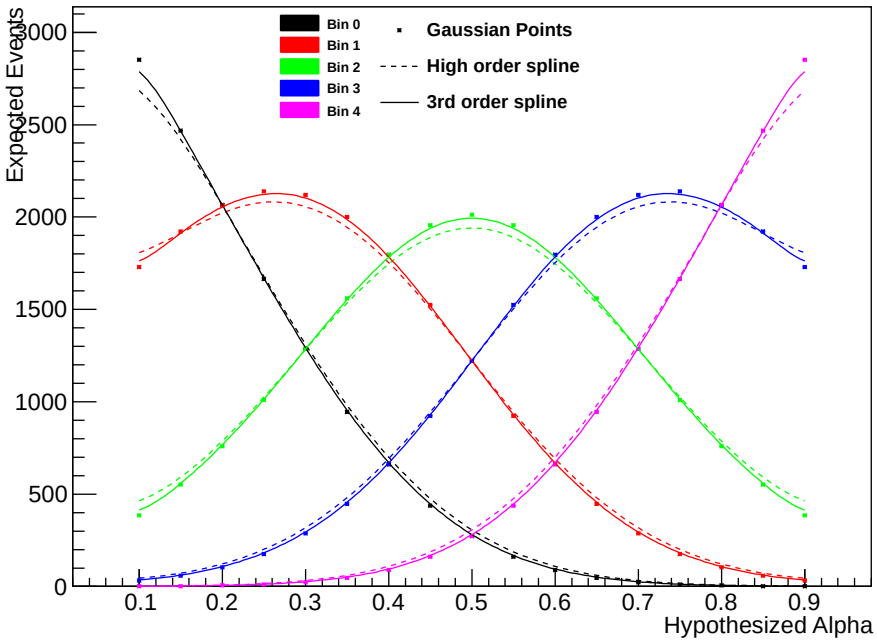


Figure A.3: Each colored curve represents a B-Spline interpolation of one bin in Figure A.2. The points on each curve represent the control points being interpolated. The solid curve is a high order spline, while the dashed curve represents a low order spline.

that removed the bias causes the low order spline to follow the statistical fluctuations to a much higher degree than the high order spline, which is much smoother. For the purpose of reconstructing the true distribution, the high order spline seems preferable. One drawback of this is that high order splines are computationally expensive, since the recursive sequence to obtain $b_{i,n}(t)$ must be redone for each t. To capture the advantages of each low and high order spline, a hybrid spline can be constructed with a second and final reweighting of a low order spline to the reweighted high order spline. In this case, the control points at each knot need only be computed once at the time of construction. Each subsequent computation of $b_{i,n}(t)$ can use these cached control points for low n. The final low order interpolation spline follows the reweighted high order spline almost exactly.

Potential biases and systematics from the method are evaluated by generating an ensemble with a procedure identical to the one used to obtain the four pseudo-experiments above. A binned likelihood $\mathcal{L}(\mu, \alpha)$ is constructed. μ is a parmameter that multiplies the overall normalization, while α is the interpolation parameter. For each pseudo-experiment the maximum likelihood estimators $\hat{\mu}(\alpha_i)$ and $\hat{\alpha}(\mu = 1)$ are computed for each α_i. The uninterpolated likelihood for each fixed α_i, $\mathcal{L}(\mu, \alpha_i)$, is also constructed as a control. The difference of the median of the distribution of $\hat{\mu}$ between the interpolated and control model, as well as the difference of the median of the distribution of $\hat{\alpha}$ from α_i, represent biases induced by the method. The standard deviation of each estimator represents systematics from the method that arise due to MC statistics. Figure A.6 shows the median and standard deviation of the distribution of $\hat{\mu} - 1$ and $\hat{\alpha} - \alpha_i$ for the fixed model and interpolated model before reweighting, while Figure A.7 shows these after reweighting. For $\hat{\mu}$ the bias from the non-reweighted spline is clear, but is completely corrected by the reweighting. Further, as can be seen by the smaller standard deviation in the interpolated model, systematics due to MC statistics are reduced by the interpolation. This is caused by the smoothing of the gaussian distribution from the high order spline. Additional information from neighboring α_i points allows a better estimate of the observable distribution at a fixed value of α. The

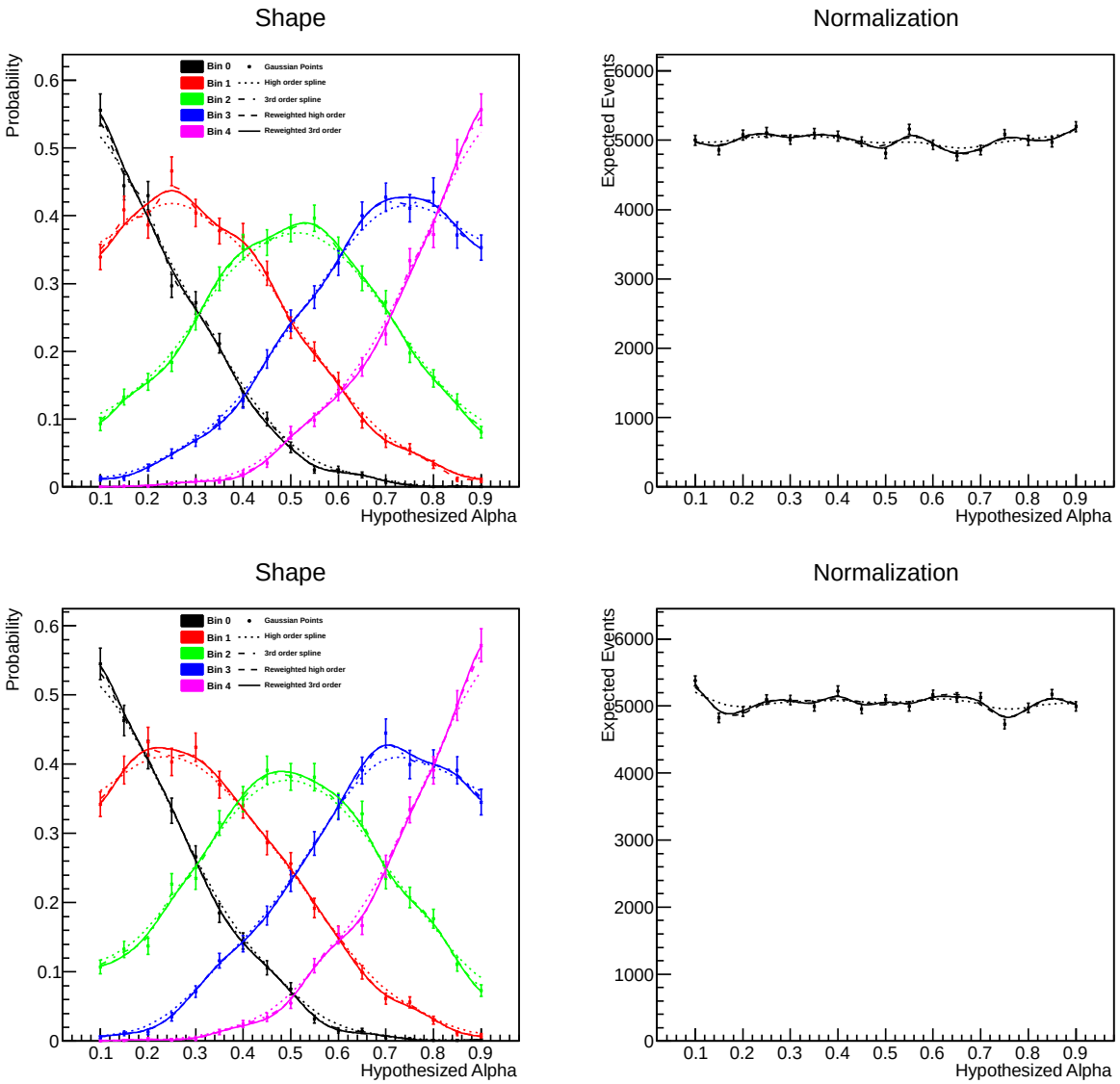


Figure A.4: Spline interpolations for two pseudo-experiments generated from gaussian distribution. The left plot shows the interpolation of the normalized histogram PDF, while the right shows the interpolation of the normalization of the histogram.

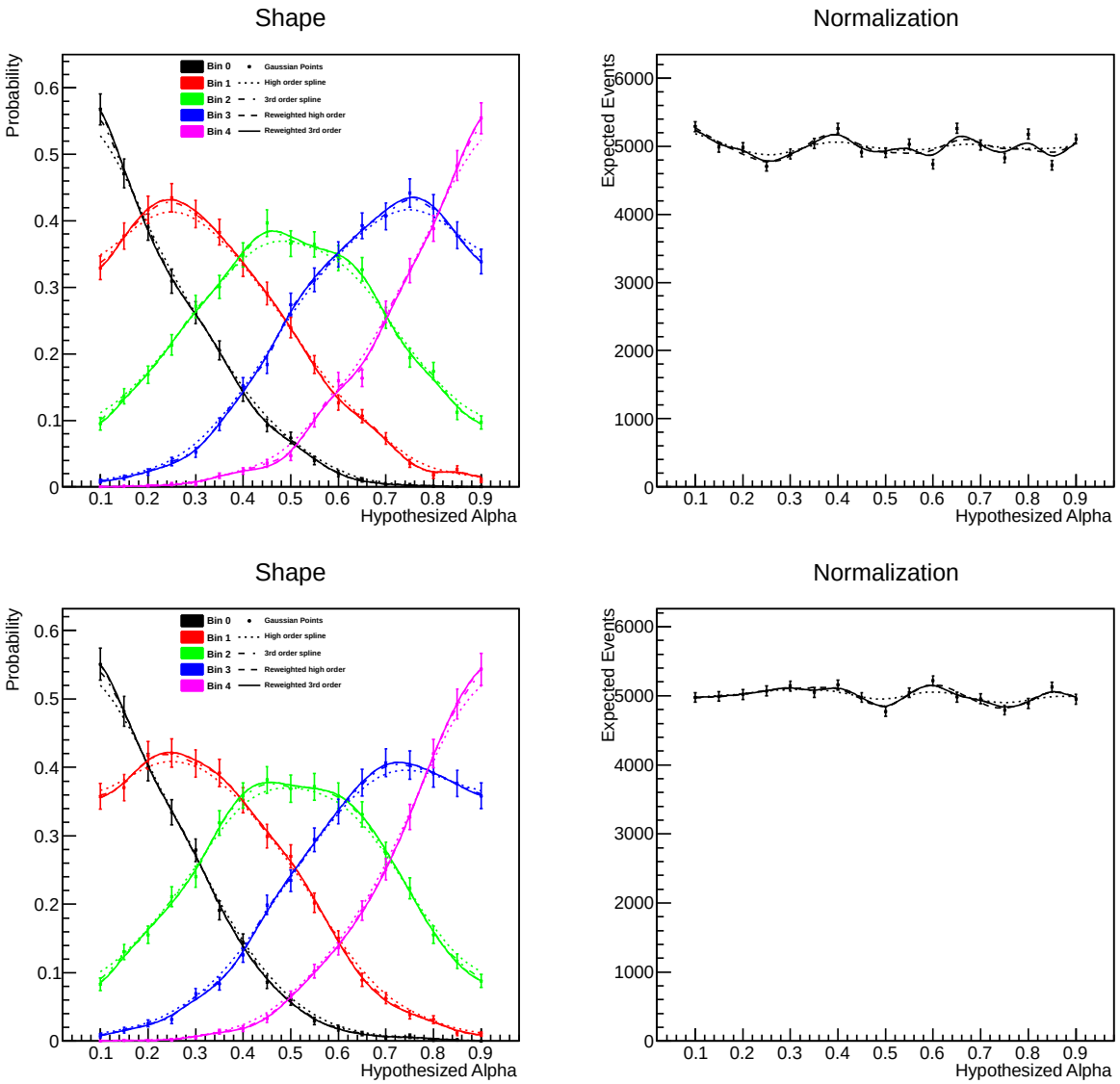


Figure A.5: Spline interpolations for two pseudo-experiments generated from gaussian distribution. The left plot shows the interpolation of the normalized histogram PDF, while the right shows the interpolation of the normalization of the histogram.

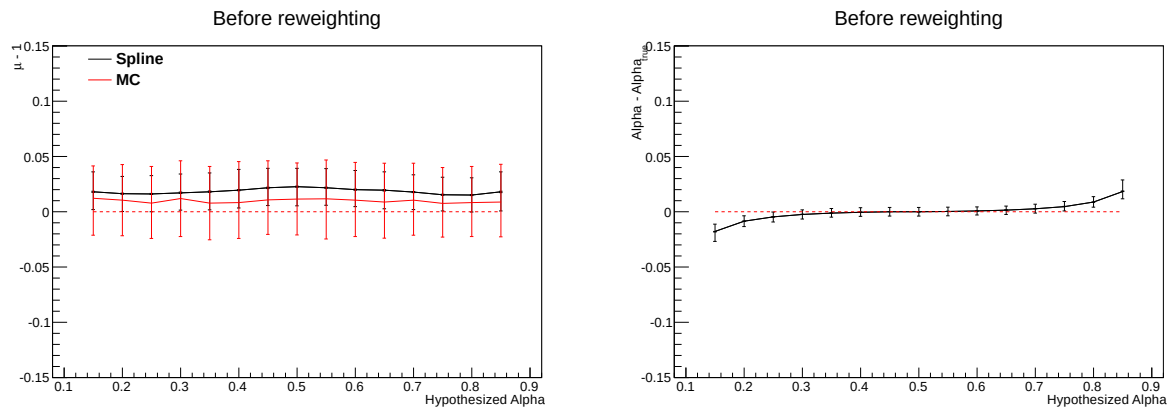


Figure A.6: Measurement of the bias induced by the spline interpolation in both the measured normalization and in the reconstructed mean value of the gaussian before the reweighting procedure.

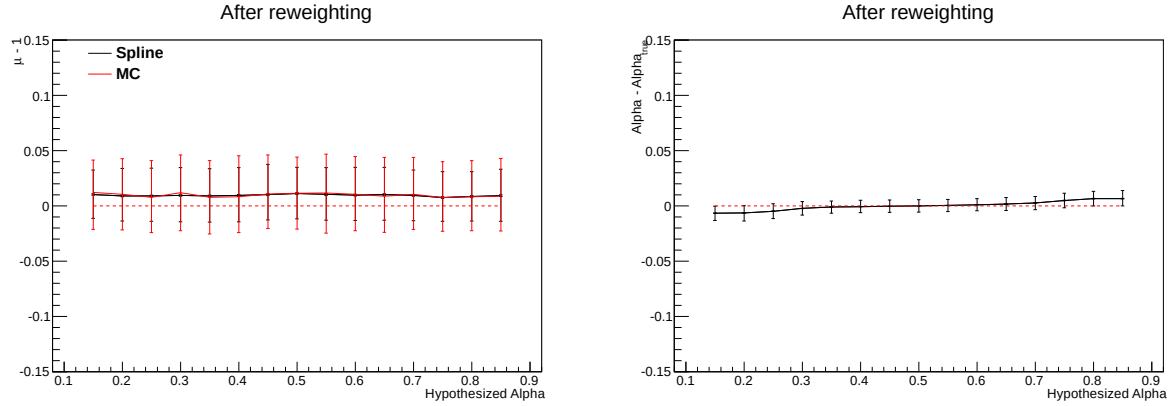


Figure A.7: Measurement of the bias induced by the spline interpolation in both the measured normalization and in the reconstructed mean value of the gaussian after the reweighting procedure. The bias in the normalization disappears. The bias in the reconstructed mean is alleviated but still present at the endpoints.

bias in $\hat{\alpha}$ before reweighting is alleviated but not completely solved by the reweighting. It is, however, within the 1σ interval of the distribution. This bias is only present at the endpoints. For point sufficiently far away (greater than the order of the spline), the bias is not present.

A.3 Application to Higgs Searches

The procedure outlined in the last sub-section is applied to the three low resolution channels in the low mass Higgs search: $H \to WW^{(*)} \to \ell\nu\ell\nu$, $H \to b\bar{b}$, and $H \to \tau\bar{\tau}$. Further consideration is required for the treatment of systematics in these channels. Both shape and normalization systematics, as outlined in Section 5.3, are present in the likelihood, and in general are m_H dependent. Interpolation of these is therefore required.

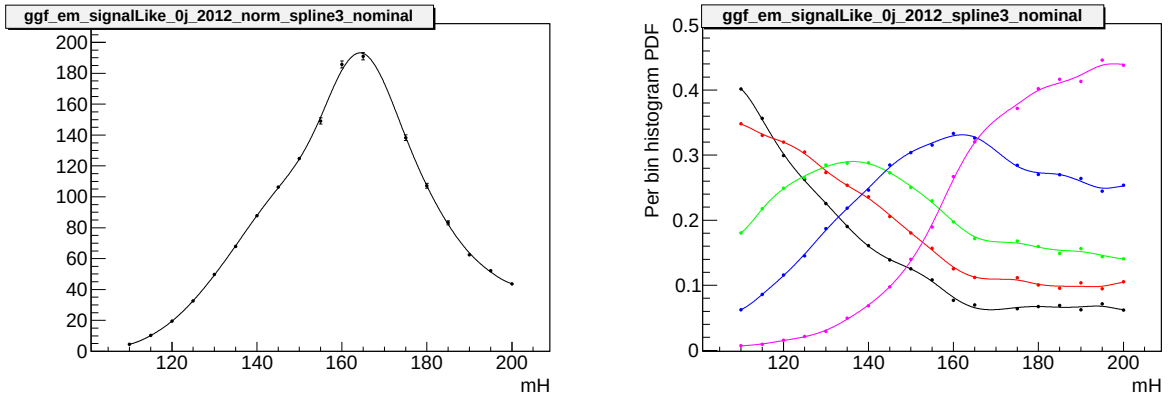
For a binned distribution with a discrete observable x_i, the expected events at each Higgs mass m can be written as follows:

$$E_m(x_i;\vec{\theta}) = (E_{0m}(x_i) + \sum_{j}^{shape}(E_{jm}(x_i) - E_{0m}(x_i))\theta_j) \times \prod_{k}^{flat}\kappa_{km}^{\theta_k} \tag{A.2}$$

$E_{0m}(x_i)$ represents the nominal distribution. For each shape systematic, there is a representative unit gaussian distributed nuisance parameter θ_j and varied distribution $E_{jm}(x_i)$. For each normalization systematic, there is a NP θ^k and response function $\kappa_{km}^{\theta_k}$. This is a simplistic form, however, given the polynomial nuisance parameter interpolation described in Section 5.3.4. The general form of the expected events can be written as:

$$E_m(x_i;\vec{\theta}) = R_m S(p_{0m}(x_i), \vec{p}_m(x_i), \vec{\theta}_{shape}) K_m(\vec{\theta}_{norm}) F(\vec{\theta}_{norm}) \tag{A.3}$$

R_m represents the nominal expected events at each m, such that $E_m = R_m$ when $\vec{\theta}$ are at their nominal values. S is the response function for the shape systematics, which in turn is a function of the nominal and varied histogram PDFs $p_{0m}(x_i)$ and $\vec{p}_m(x_i)$, respectively. K_m represents the mass dependent response function to the normalization systematics, and F the mass independent normalization systematics response function that needs no interpolation. The spline is employed such that R_m, $p_{0m}(x_i)$, $\vec{p}_m(x_i)$, and K_m are the control points of spline functions $R(m)$, $p_0(x_i;m)$, $\vec{p}(x_i;m)$, and $K(\vec{\theta}_{norm};m)$, respectively, yielding in total M+3 spline functions for M shape systematics. The total expected events for an observable



(a) (b)

Figure A.8: Interpolation of the normalization (left) and shape (right) of the transverse mass of the ggF signal in the $e\mu - 0j$ signal region in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis.

distribution is finally obtained:

$$E(x_i, \vec{\theta}; m) = R(m)S(p_0(x_i; m), \vec{p}(x_i; m), \vec{\theta}_{shape})K(\vec{\theta}_{norm}; m)F(\vec{\theta}_{norm}) \tag{A.4}$$

$K(\vec{\theta}_{norm}; m)$ requires a special note. Due to its unfactorizable dependence on the nuisance parameters $\vec{\theta}_{norm}$, it has no unique reweighting. As a consequence, it is treated with a standard 3^{rd} order spline with no reweighting. So long as the systematics are not heavily mass dependent, this bears negligible side effects due to end-point fringing.

The splines $R(m)$ and $p_0(x_i; m)$ are shown in Figure A.8- A.10 for select distributions in each channel. The splines $K(\vec{\theta}_{norm}; m)$ for the WW model is also shown in Figure A.11 for the MLEs $\hat{\theta}$ evaluated at $\mu = 0$, $\mu = 1$, and $\hat{\mu}$. Figure A.12- A.14 shows $\hat{\mu}$ versus m_H for the individual models, while Figure A.15 shows the $\hat{\mu}$ and p_0 versus m_H for the combined WW, $\tau\tau$, and bb parametrized likelihood.

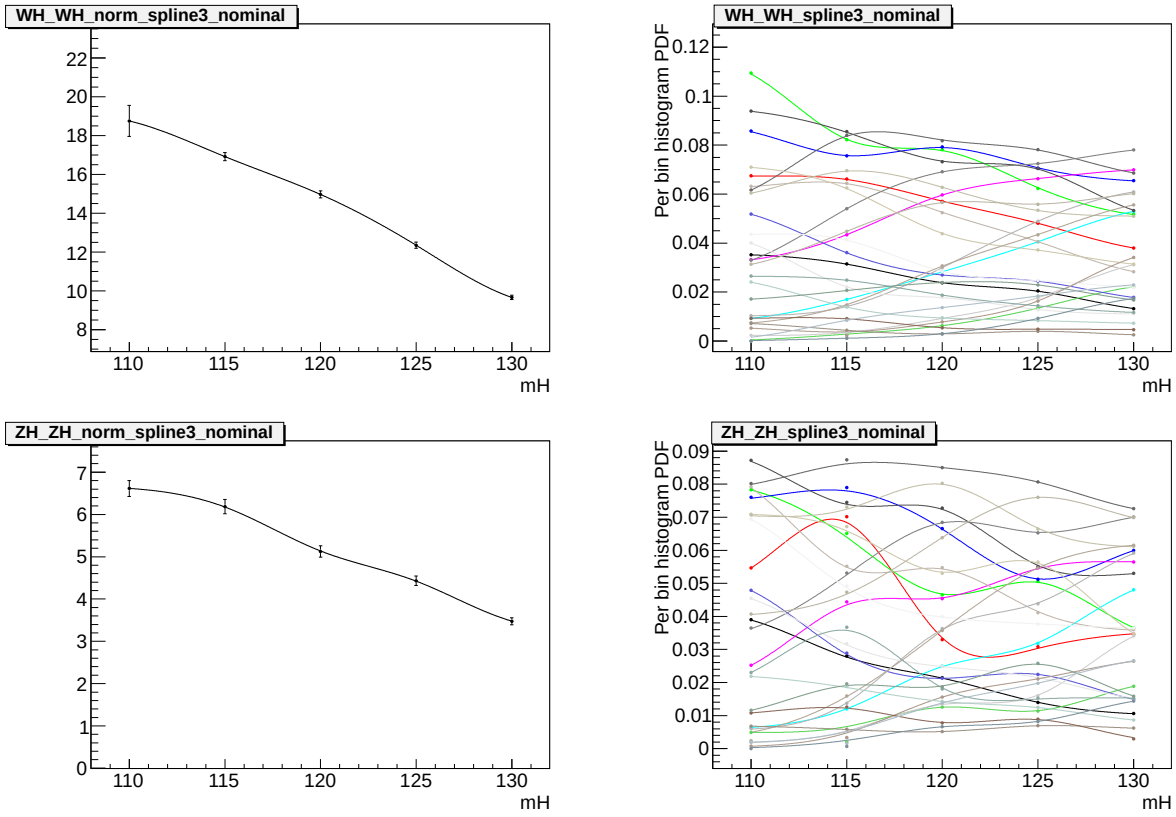


Figure A.9: Interpolation of the normalization (left) and shape (right) of the invariant $b\bar{b}$ mass of the WH (top) and ZH (bottom) signal in the signal region in the $H \to b\bar{b}$ analysis.

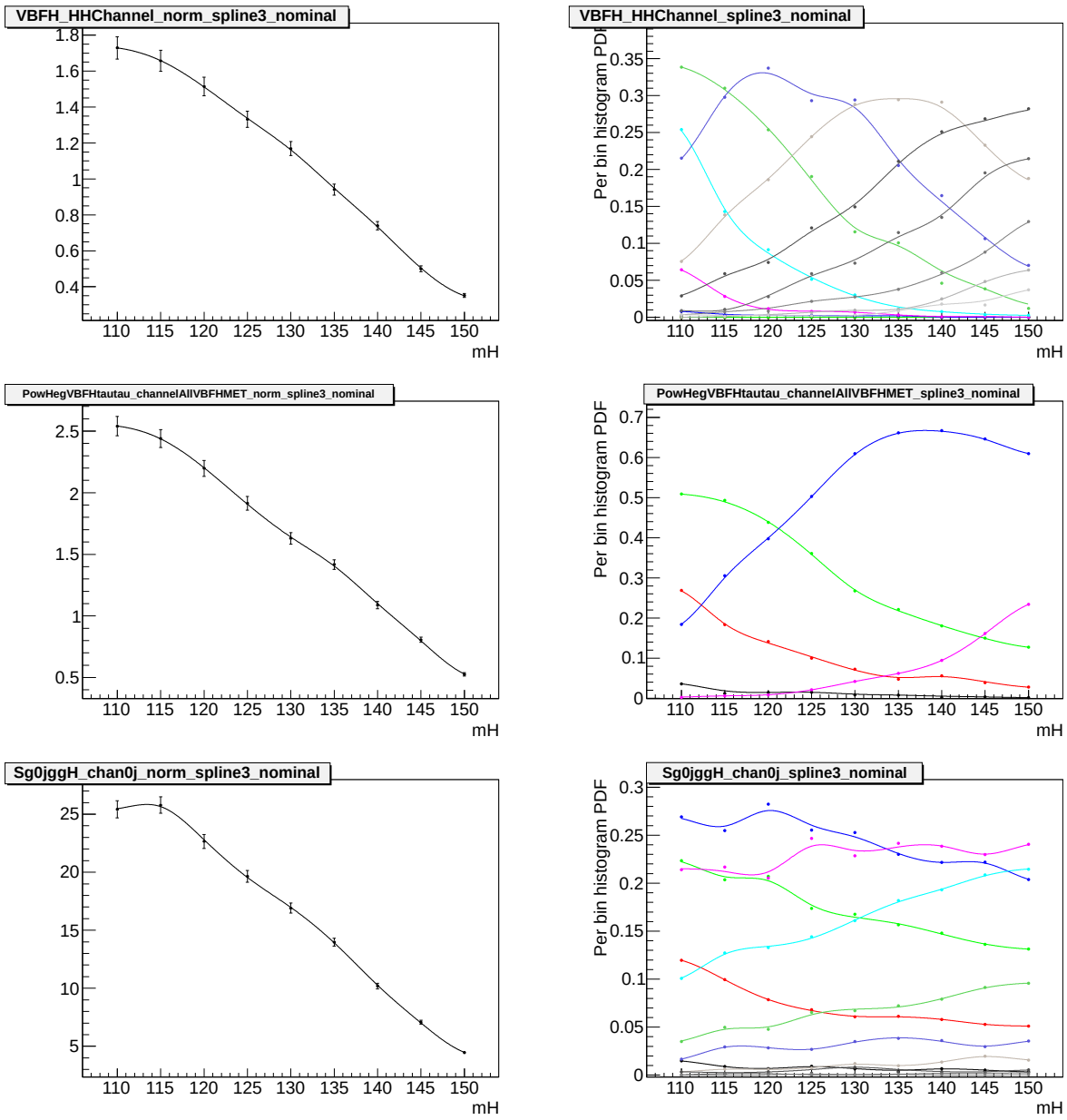


Figure A.10: Interpolation of the normalization (left) and shape (right) of the reconstructed $\tau^+\tau^-$ mass of the $\tau_{\text{had}}\tau_{\text{had}}$ (top), $\tau_{\text{lep}}\tau_{\text{had}}$ (center), and $\tau_{\text{lep}}\tau_{\text{lep}}$ (bottom) signal in the signal region in the $H \to \tau^+\tau^-$ analysis.

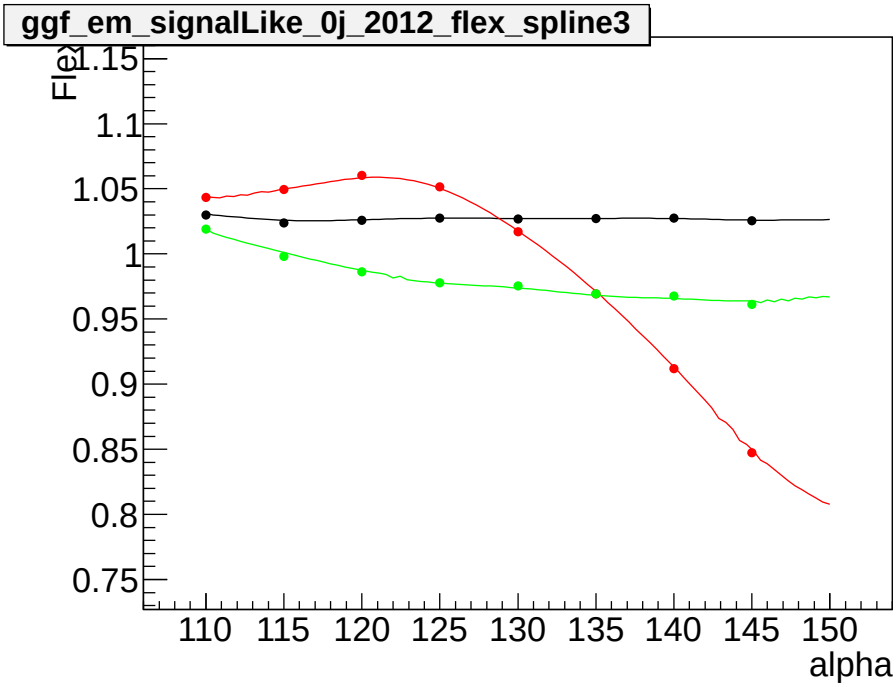


Figure A.11: Interpolation of the mass dependent normalization systematic ggF signal response term for a fit to $\hat{\mu}$ (black), $\mu = 1$ (red), and $\mu = 0$ (green) in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis.

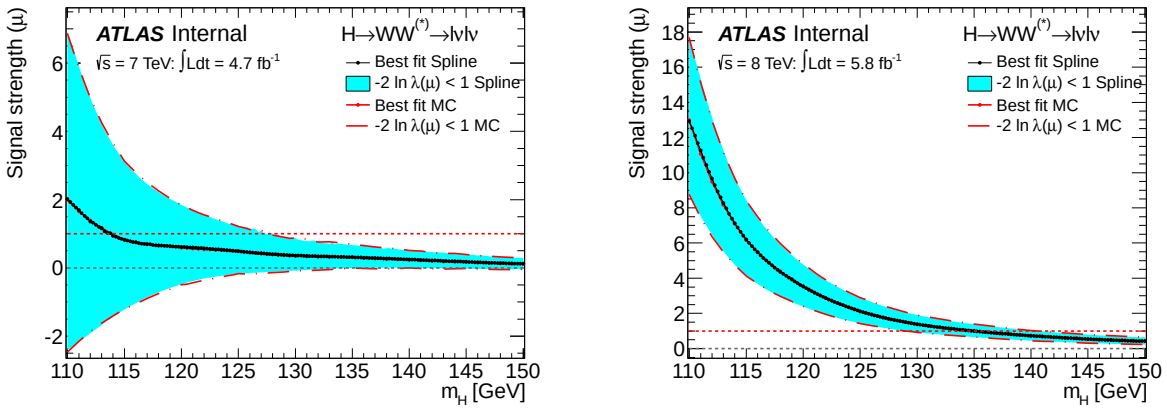


Figure A.12: The best fit signal strength versus m_H for 7 TeV (left) and 8 TeV (right) data in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis for the spline interpolated signal, such as that shown in Figure A.8, and the standard fixed-point MC analysis.

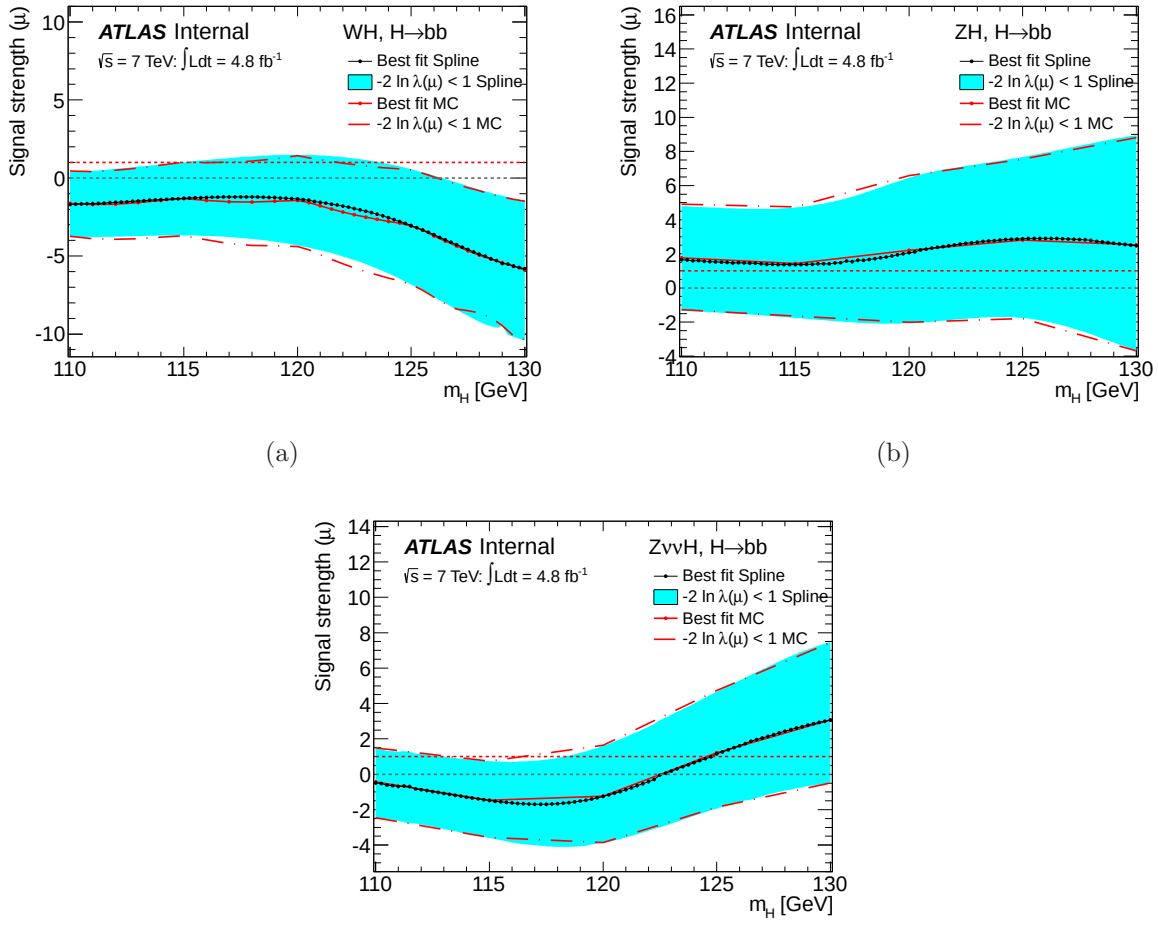
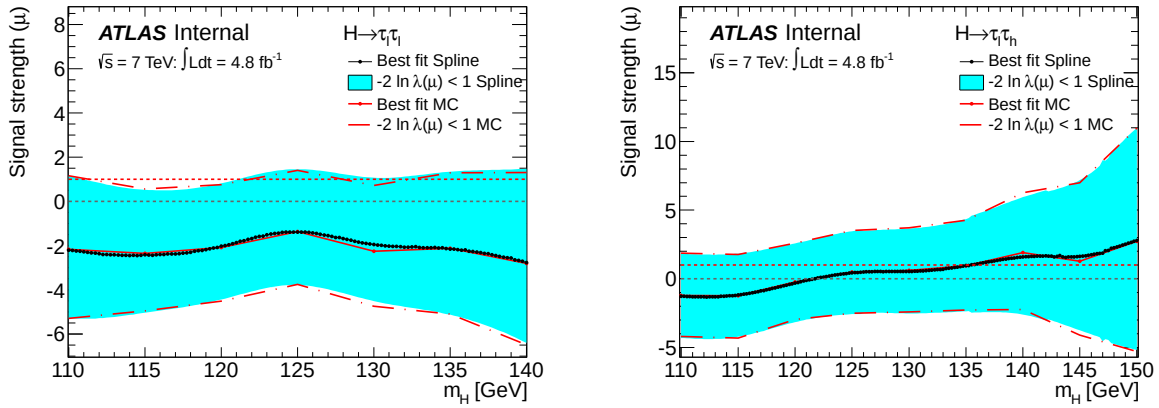
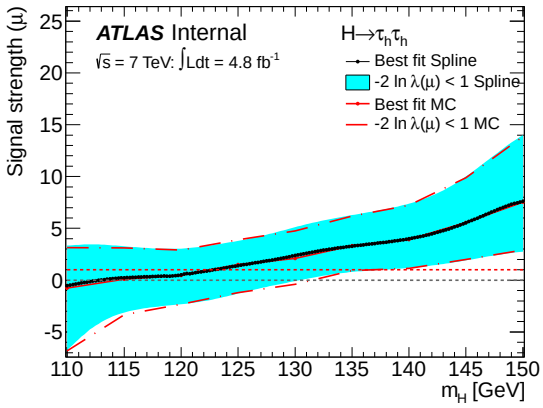


Figure A.13: The best fit signal strength versus m_H for 7 TeV data in the (a) WH, (b), ZH, $Z \to \ell\ell$, and (c) ZH, $Z \to \nu\nu$ analyses for the spline interpolated signal, such as that shown in Figure A.9, and the standard fixed-point MC analysis with ALR interpolation [66]. Some deviation occurs between the two at the interpolated points and is due to known features in the ALR interpolation.

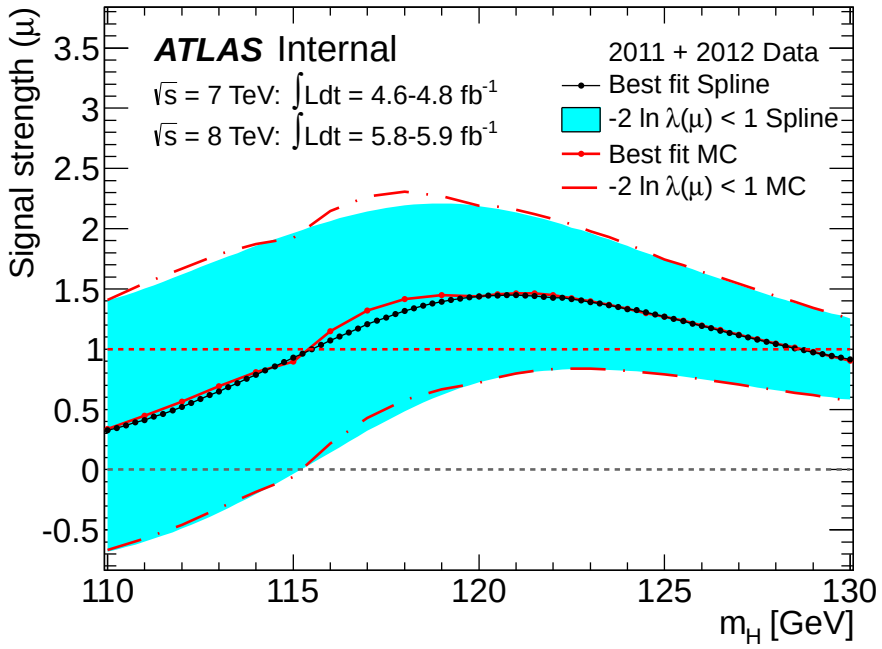


(a) (b)

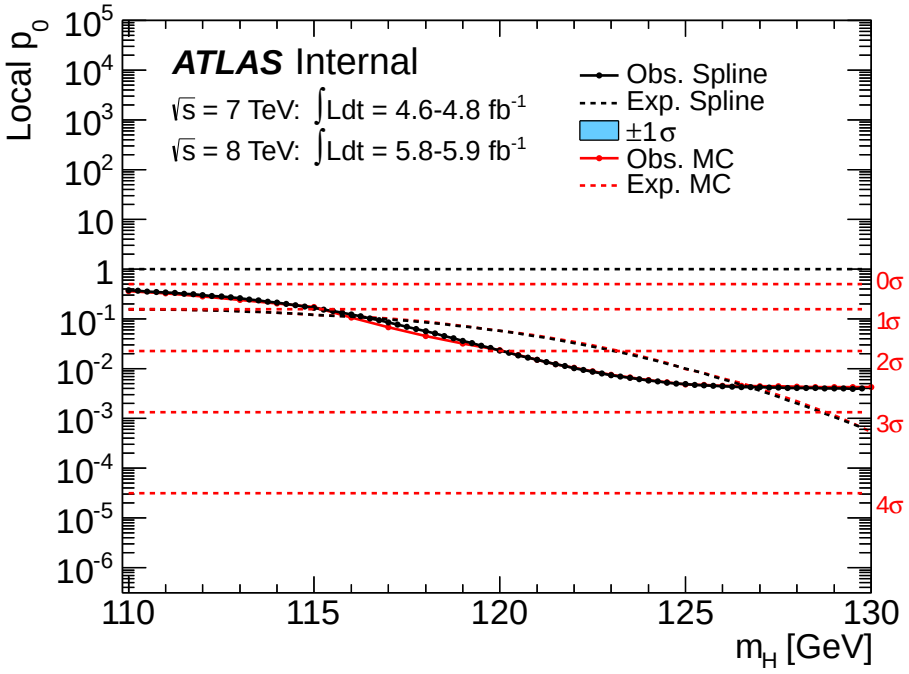


(c)

Figure A.14: The best fit signal strength versus m_H for 7 TeV data in the (a) $\tau_{\text{lep}}\tau_{\text{lep}}$, (b), $\tau_{\text{lep}}\tau_{\text{had}}$, and (c) $\tau_{\text{had}}\tau_{\text{had}}$ analyses for the spline interpolated signal, such as that shown in Figure A.10, and the standard fixed-point MC analysis.



(a)



(b)

Figure A.15: The (a) best fit signal strength and (b) p_0 versus m_H for 7+8 TeV data for the combined WW+$\tau\tau$+bb model for the spline interpolated signal in black, and the standard fixed-point MC analysis with ALR interpolation in red. The arching between 115 and 120 GeV in the red curve is due to known features in the ALR interpolation method.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Data-quality requirements and event cleaning for jets and missing transverse energy reconstruction with the atlas detector in proton-proton collisions at a center-of-mass energy of $sqrts = 7$ tev. Technical Report ATLAS-CONF-2010-038, CERN, Geneva, Jul 2010.

[2] Commissioning of the atlas high-performance b-tagging algorithms in the 7 tev collision data. Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011.

[3] Combination of standard model higgs boson searches and measurements of the properties of the new boson with a mass near 125 gev. Technical Report CMS-PAS-HIG-12-045, CERN, Geneva, 2012.

[4] Improved electron reconstruction in atlas using the gaussian sum filter-based model for bremsstrahlung. Technical Report ATLAS-CONF-2012-047, CERN, Geneva, May 2012.

[5] Observation and study of the higgs boson candidate in the two photon decay channel with the atlas detector at the lhc. Technical Report ATLAS-CONF-2012-168, CERN, Geneva, Dec 2012.

[6] Search for the standard model higgs boson in h→ tau tau decays in proton-proton collisions with the atlas detector. Technical Report ATLAS-CONF-2012-160, CERN, Geneva, Nov 2012.

[7] Search for the standard model higgs boson in produced in association with a vector boson and decaying to bottom quarks with the atlas detector. Technical Report ATLAS-CONF-2012-161, CERN, Geneva, Nov 2012.

[8] Search for the standard model higgs boson produced in association with top quarks in proton-proton collisions at $\sqrt{s} = 7$ tev using the atlas detector. Technical Report ATLAS-CONF-2012-135, CERN, Geneva, Sep 2012.

[9] Combined coupling measurements of the higgs-like boson with the atlas detector using up to 25 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-034, CERN, Geneva, Mar 2013.

[10] Combined measurements of the mass and signal strength of the higgs-like boson with the atlas detector using up to 25 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-014, CERN, Geneva, Mar 2013.

[11] Evidence for a particle decaying to w+w- in the fully leptonic final state in a standard model higgs boson search in pp collisions at the lhc. Technical Report CMS-PAS-HIG-13-003, CERN, Geneva, 2013.

[12] Measurements of the properties of the higgs-like boson in the four lepton decay channel with the atlas detector using 25 fb-1 of proton-proton collision data. Technical Report ATLAS-CONF-2013-013, CERN, Geneva, Mar 2013.

[13] Measurements of the properties of the higgs-like boson in the two photon decay channel with the atlas detector using 25 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-012, CERN, Geneva, Mar 2013.

[14] Measurements of the properties of the higgs-like boson in the $ww^{(*)} \to \ell\nu\ell\nu$ decay channel with the atlas detector using 25 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-030, CERN, Geneva, Mar 2013.

[15] Study of the spin of the higgs-like boson in the two photon decay channel using 20.7 fb-1 of pp collisions collected at $\sqrt{s} = 8$ tev with the atlas detector. Technical Report ATLAS-CONF-2013-029, CERN, Geneva, Mar 2013.

[16] Study of the spin of the new boson with up to 25 fb^{-1} of atlas data. Technical Report ATLAS-CONF-2013-040, CERN, Geneva, Apr 2013.

[17] Study of the spin properties of the higgs-like particle in the $h \to ww^{(*)} \to e\nu\mu\nu$ channel with 21 fb^{-1} of $\sqrt{s} = 8$ tev data collected with the atlas detector. Technical Report ATLAS-CONF-2013-031, CERN, Geneva, Mar 2013.

[18] Updated measurements of the higgs boson at 125 gev in the two photon decay channel. Technical Report CMS-PAS-HIG-13-001, CERN, Geneva, 2013.

[19] 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, 1998.

[20] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. 2009.

[21] Georges Aad et al. Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data. *Eur.Phys.J.*, C72:1909, 2012.

[22] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012.

[23] Stefano Actis, Giampiero Passarino, Christian Sturm, and Sandro Uccirati. NLO electroweak corrections to Higgs boson production at hadron colliders. *PLB*, 670:12, 2008.

[24] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini. Two loop light fermion contribution to Higgs production and decays. *Phys.Lett.*, B595:432–441, 2004.

[25] Charalampos Anastasiou and Kirill Melnikov. Higgs boson production at hadron colliders in NNLO QCD. *Nucl.Phys.*, B646:220–256, 2002.

[26] L. Bergstrom and G. Hulth. Induced higgs couplings to neutral bosons in e+e- collisions. *Nuclear Physics B*, 259(1):137 – 155, 1985.

[27] E. Braaten and J. P. Leveille. Higgs-boson decay and the running mass. *Phys. Rev. D*, 22:715–721, Aug 1980.

[28] A. Bredenstein, A. Denner, S. Dittmaier, and M. Weber. Radiative corrections to the semileptonic and hadronic Higgs-boson decays H $\rightarrow$ WW/ZZ $\rightarrow$ 4 fermions. *JHEP*, 0702:080, 2007.

[29] A. Bredenstein, Ansgar Denner, S. Dittmaier, and M.M. Weber. Precise predictions for the Higgs-boson decay $H \rightarrow WW/ZZ \rightarrow 4$ leptons. *Phys.Rev.*, D74:013004, 2006.

[30] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 0804:063, 2008.

[31] R.N. Cahn, M.S. Chanowitz, and N. Fleishon. Higgs particle production by z-¿hgamma. *Physics Letters B*, 82(1):113 – 116, 1979.

[32] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012.

[33] Serguei Chatrchyan et al. On the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs. *Phys. Rev. Lett.*, 110:081803, 2013.

[34] Kyle S. Cranmer. Kernel estimation in high-energy physics. *Comput.Phys.Commun.*, 136:198–207, 2001.

[35] S. Dawson. Radiative corrections to Higgs boson production. *Nucl.Phys.*, B359:283–300, 1991.

[36] Rodolfo A. Diaz and R. Martinez. The Custodial symmetry. *Rev.Mex.Fis.*, 47:489–492, 2001.

[37] A. Djouadi, M. Spira, and P.M. Zerwas. Production of Higgs bosons in proton colliders: QCD corrections. *Phys.Lett.*, B264:440–446, 1991.

[38] Abdelhak Djouadi. The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model. *Phys.Rept.*, 457:1–216, 2008.

[39] Abdelhak Djouadi. The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model. *Phys.Rept.*, 459:1–241, 2008.

[40] A. Elagin, P. Murat, A. Pranko, and A. Safonov. A New Mass Reconstruction Technique for Resonances Decaying to di-tau. *Nucl.Instrum.Meth.*, A654:481–489, 2011.

[41] John R. Ellis, Mary K. Gaillard, and Dimitri V. Nanopoulos. A Phenomenological Profile of the Higgs Boson. *Nucl.Phys.*, B106:292, 1976.

[42] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *PRL*, 13:321–323, 1964.

[43] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[44] J. Gaiser. CHARMONIUM SPECTROSCOPY FROM RADIATIVE DECAYS OF THE J / psi AND psi-prime. 1982.

[45] H. M. Georgi, S. L. Glashow, M. E. Machacek, and D. V. Nanopoulos. Higgs bosons from two-gluon annihilation in proton-proton collisions. *Phys. Rev. Lett.*, 40:692–694, Mar 1978.

[46] A. Grau, G. Pancheri, and R.J.N. Philips. Contributions of off-shell top quarks to decay processes. *Physics Letters B*, 251(2):293 – 298, 1990.

[47] CDF Group, D0 Collaborations, the Tevatron New Physics, and Higgs Working. Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data. 2012.

[48] LHC Higgs Cross Section Working Group, A. David, A. Denner, M. Duehrssen, M. Grazzini, et al. LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle. 2012.

[49] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *PRL*, 13:585–587, 1964.

[50] W. Heitler. The quantum theory of radiation; 3rd ed. 1954.

[51] P. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *PRL*, 13:508–509, 1964.

[52] Andrei L. Kataev and Victor T. Kim. QCD corrections to decays of intermediate mass Higgs boson. 1996.

[53] Wai-Yee Keung and William J. Marciano. HIGGS SCALAR DECAYS: H -¿ W+- X. *Phys.Rev.*, D30:248, 1984.

[54] Benjamin W. Lee, C. Quigg, and H. B. Thacker. Weak interactions at very high energies: The role of the higgs-boson mass. *Phys. Rev. D*, 16:1519–1531, Sep 1977.

[55] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011.

[56] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. *CERN-2012-002*, CERN, Geneva, 2012.

[57] C. H. Llewellyn Smith. High energy behaviour and gauge symmetry. *Phys.Lett.*, 46B:233–236, 1973.

[58] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189, 2009.

[59] D W Miller, A Schwartzman, and D Su. Jet-vertex association algorithm. Technical Report ATL-COM-PHYS-2008-008, CERN, Geneva, Jan 2008.

[60] J. Neyman and E. S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:289–337, 1933.

[61] M. Oreglia. A STUDY OF THE REACTIONS psi-prime $\rightarrow$ gamma gamma psi. 1980.

[62] Particle Data Group, C. Amsler, M. Doser, and M. et al. Antonelli. Review of Particle Physics. *Physics Letters B*, 667:1–6, September 2008.

[63] Antonio Pich. The Standard model of electroweak interactions. 2007.

[64] R. Harlander and W. B. Kilgore. Next-to-next-to-leading order Higgs production at hadron colliders. *PRL*, 88:201801, 2002.

[65] V. Ravindran, J. Smith, and W. L. van Neerven. NNLO corrections to the total cross-section for Higgs boson production in hadron hadron collisions. *Nucl.Phys.*, B665:325–366, 2003.

[66] A.L Read. Linear interpolation of histograms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 425(1-2):357 – 360, 1999.

[67] L. Resnick, M. K. Sundaresan, and P. J. S. Watson. Is there a light scalar boson? *Phys. Rev. D*, 8:172–178, Jul 1973.

[68] R.K. Ellis et al. Higgs decay to $\tau^+\tau^-$: A possible signature of intermediate mass Higgs bosons at the SSC. *Nucl. Phys. B*, 297:221, 1988.

[69] Tomasz Skwarnicki. A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances. 1986.

[70] M. Spira, A. Djouadi, D. Graudenz, and P.M. Zerwas. Higgs boson production at the LHC. *Nucl.Phys.*, B453:17–82, 1995.

[71] Iain W. Stewart and Frank J. Tackmann. Theory Uncertainties for Higgs and Other Searches Using Jet Bins. *Phys.Rev.*, D85:034011, 2012.

[72] Abraham Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3):pp. 426–482, 1943.

[73] Steven Weinberg. A Model of Leptons. *Phys.Rev.Lett.*, 19:1264–1266, 1967.

[74] Frank Wilczek. Decays of heavy vector mesons into higgs particles. *Phys. Rev. Lett.*, 39:1304–1306, Nov 1977.

[75] HWW working group. Object selection for the $H \to WW$ search with the ATLAS detector at $\sqrt{s} = 8$ TeV. Technical Report ATL-COM-PHYS-2013-147, CERN, Geneva, Feb 2013.