A search for Higgs-portal dark matter and new phenomena with monojet signatures in pp collisions at $\sqrt{s} = 8$ TeV

by

James D. Pearce
B.Sc., McGill University, 2009
M.Sc., University of Victoria, 2012

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Physics and Astronomy

A search for Higgs-portal dark matter and new phenomena with monojet signatures in pp collisions at $\sqrt{s} = 8$ TeV

by

James D. Pearce
B.Sc., McGill University, 2009
M.Sc., University of Victoria, 2012

Supervisory Committee

Dr. R. Kowalewski, Supervisor
(Department of Physics and Astronomy)

Dr. R. Keeler, Departmental Member
(Department of Physics and Astronomy)

Dr. M. Roney, Departmental Member
(Department of Physics and Astronomy)

Dr. C. Bradley, Outside Member
(Department of Engineering)

Supervisory Committee

Dr. R. Kowalewski, Supervisor
(Department of Physics and Astronomy)

Dr. R. Keeler, Departmental Member
(Department of Physics and Astronomy)

Dr. M. Roney, Departmental Member
(Department of Physics and Astronomy)

Dr. C. Bradley, Outside Member
(Department of Engineering)

ABSTRACT

A search for new phenomena in final states with one or more energetic jets and large missing transverse momentum, E_T^{miss}, is presented. An integrated luminosity of 20 fb^{-1} is collected from $\sqrt{s} = 8$ TeV pp collisions at the LHC with the ATLAS detector during 2012 operations. A selection criteria is imposed requiring events to have monojet signatures, with $E_T^{miss} > 350$ GeV and no identified leptons. Standard Model backgrounds and systematics uncertainties are estimated using a maximum likelihood procedure. The number of events passing this selection criteria is in good agreement with Standard Model expectations. These events are then divided up into three orthogonal signal regions based on the outputs of two Random Forest classifiers trained to classify invisible decays of the Higgs boson produced through the vector boson fusion and Higgs-strahlung production modes. These results are then interpreted in terms of three different Higgs-portal models and translated into upper limits on WIMP-nucleon cross-sections for comparison with direct detection experiments.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

First I would like to thank the University of Victoria for all of the support and opportunity this amazing institution has afforded me. Thank you, to all of the professors, lecturers, staff and, of course, my fellow graduate students.

I would also like to thank my thesis committee, Richard Keeler, Mike Roney and Colin Bradley for all their valuable inputs and corrections. I'd also like to thank Sampa Bhadra for reading my thesis and acting as the external examiner.

Of course none of my work would have been possible without the ATLAS and LHC collaborations at CERN. It is hard to even fathom the immense amount science and engineering that goes into such an endeavor. I would like to thank the many different ATLAS working groups that provided the necessary data and tools, especially the Monojet and Missing ET Trigger groups, the work of which was integral to my own. In particular, I would like to thank Philippe Calfayan, Florian Bernlochner and Allen Mincer for their important work, as well as their generosity in extending a helping hand whenever I needed it.

Finally, I would like to thank my supervisor, professor Robert Kowalewski, who guided me all these years, through both my Masters and PhD. Bob's close and continuous involvement in my thesis and work at CERN was in every way essential to my success and the completion of this dissertation. I could not have done it without his wisdom, patience and unwavering support. Not only did he give me the guidance that I needed, but also the freedom to choose my own path, allowing me to pursue the ideas and topics that I am truly passionate about. Even in my transition from academia to industry he continued to support and encourage me. I couldn't have asked for a better supervisor and mentor. Thank you Bob.

DEDICATION

To my father, James E. Pearce, for all of his love and encouragement. Thank you for always being there for me.

Chapter 1

Introduction

The Standard Model of particles physics embodies our current understanding of all known elementary particles and their interactions. The Standard Model is a theoretical framework that combines quantum chromodynamics (QCD) and the electroweak model into an internally consistent theory that incorporates the electromagnetic, weak and strong interactions. Over the past few decades the Standard Model has been enormously successful in predicting experimental results. It has successfully predicted the existence of the weak neutral current, charm and top quarks as well as the W and Z bosons and, most recently, the Higgs boson [1, 2, 3, 4, 5]. Indeed, when used with general relativity the Standard Model accounts for almost all natural phenomena observed. However, despite all of the Standard Model's triumphs, it has a number of known limitations, such as having 20 arbitrary parameters, lacking a viable dark matter candidate and failing to incorporate a quantum theory of gravitation. The Standard Model's unprecedented accuracy juxtaposed to its known limitations implies that it may be part of—or incorporated into—a more comprehensive theory. Many theories *beyond* the Standard Model (BSM) have been developed to address these flaws and omissions. The main program of experimental high energy particle physics (HEP) is to develop and conduct experiments that test both the Standard Model and BSM theories.

One of the greatest mysteries in astronomy and cosmology is the observation that 84% of the matter in our universe is composed of dark matter [6], an unknown substance that neither absorbs nor emits light or any observable amount of electromagnetic radiation. Evidence for dark matter is found in galaxy rotation curves, gravitational lensing and the cosmic microwave background [7]. The simplest, and most favoured, model of dark matter that accommodates all astrophysical observations is that of Weakly Interacting Massive Particles (WIMPs). WIMPs cannot be accounted for by any of the known particles in the Standard Model of particle physics. Thus direct detection of WIMPs would be a clear indication of new and exciting physics beyond our currently accepted theories.

If one assumes dark matter mass is generated through the Standard Model Higgs mechanism then the Higgs boson will couple to WIMPs directly [8]. Such a coupling would allow for WIMPs to be produced at the LHC via any of the Higgs production modes. Such a discovery would shed light on the nature of dark matter and point in a clear direction for future BSM theories and experiments.

1.1 The Standard Model

Before the Standard Model of particle physics can be described in any meaningful way one must be familiar with the theoretical framework in which it is embedded. The Standard Model can be formally establish in several ways. Some are motivated by the historical discoveries that support the theory, others by means of aesthetics or simplicity. The theoretical formalism presented here only aims to take the most direct route in penetrating the—often abstract and lofty—issues that motivate the current theories and experiments central to the field today.

The arguments and reasoning presented in what follows draws heavily upon the works of Refs. [9, 10, 11].

1.1.1 Field theories and the principle of least action

The Euler-Lagrange equation for a Lagrangian density[1] $\mathscr{L}$ of a continuously varying real scalar field over space-time $\phi(x)$ is given as

$$\partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathscr{L}}{\partial \phi} = 0, \tag{1.1}$$

where ∂_μ is shorthand for $\frac{\partial}{\partial x^\mu}$. Substituting various choices for the Lagrangian $\mathscr{L}$ into equation 1.1 will yield different (well known) physical equations. For example,

$$\mathscr{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2} m^2 \phi^2$$

---

[1] Often referred to as simply “the Lagrangian”.

yields the Klein-Gordon equations[2,3] for a field with mass m,

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0.$$

The Euler-Lagrange equation is a result of the *principle of least action.* Where the action, denoted $\mathcal{S}$, of the system is given as the integral of the Lagrangian between spacetimes x_1 and x_2

$$\mathcal{S} = \int_{x_1}^{x_2} \mathcal{L} \mathrm{d}^4 x \tag{1.2}$$

The principle of least action states that the path taken by the system between spacetimes x_1 and x_2 is the one for which the action is stationary (no change) to first order [12]. Here "path" refers to the curves traced out in configuration space (sometimes referred to as "phase space") in terms of the chosen coordinates. Mathematically this is equivalent to extremizing the action, succinctly expressed as

$$\delta \mathcal{S} = 0, \tag{1.3}$$

where δ means small change. The action and Lagrangian are central to the development of theoretical physics as they both contain the dynamics of the system for all times and are independent of the chosen coordinate system. The principle of least action holds true across classical field theory, special relativity and quantum theory, allowing for all three to be combined into a unifying theory of quantum fields (QFT).

In QFT the "paths" being considered correspond to all possible evolutions of the field over space-time. The action is a functional[4] of the field, $\mathcal{S} = \mathcal{S}[\phi]$, where the field $\phi(x)$ is itself a function of spacetime. In theory all dynamics and interactions of a quantum system can be derived from integrating over all possible paths, a procedure known as *path integration.* Of particular interest in collider physics is the probability amplitude $\mathcal{M}$ which is used to calculate cross-sections and decay widths. The probability amplitude for the field ϕ to transition from some initial configuration I to a final configuration F is

$$\mathcal{M}_{I \to F} = \frac{1}{Z} \int \mathcal{D}\phi e^{i\mathcal{S}} \phi_I \phi_F, \tag{1.4}$$

where Z is a normalization factor and $\mathcal{D}\phi$ is the measure over all possible config-

---

[2] Notational note: Einstein's summation convention will be used for all tensor equations, meaning repeated indices are always summed over, e.g. $\sum_i x_i y^i = x_i y^i$.

[3] For simplicity all physical expressions are expressed in natural units, where $\hbar = c = 1$.

[4] A functional is a function that takes in a function and returns as number.

urations of ϕ. The path integral can be thought of as a weighted average over the different configurations of the field where the phase $e^{i\mathcal{S}}$ suppresses configurations that disobey the principle of least action.

However, in practice such integrals are intractable, forcing one to apply approximate methods such as perturbation theory or Monte Carlo integration. In perturbation theory each order of a perturbative series can be represented as one or more Feynman diagrams (e.g. figure 7.1). Typically only up to second order diagrams are considered as they dominate the expansion. Higher order corrections may be calculated when the calculation requires extra precision.

While the probability amplitudes $\mathcal{M}$ are necessary to calculate relevant observables such as cross-sections and decay widths much can be gleaned from the Lagrangian itself, such as the dynamics, interactions and symmetries of the system.

1.1.2 Matter particles and force mediators

In QFT particles are treated as excitations, called *quanta*, of their associated field which is modelled as a quantum oscillator. The fundamental particles that furnish the Standard Model are distinguished by the symmetries they observe. The most familiar presentation of the Standard Model is given in Table 1.1 where particles are listed by their mass eigenstates—the eigenvalue of which is a readily measured observable. In addition to mass, particles are often identified by their quantum numbers that correspond to internal symmetries, such as electric charge or spin. For example, quarks and gluons carry a colour charge, which is analogous to electric charge but with three distinct charges that are associated with the strong interaction. Quarks and leptons also carry weak isospin, with two distinct charges which are conserved in weak interactions. In addition particles may be distinguished by their space-time symmetries, e.g. each of the listed particles is also associated with an antiparticle partner, which carries the same quantum numbers, but has opposite charge(s) unless neutral.

In the mass basis there are 12 fermions (spin $= \frac{1}{2}$) and 5 bosons (spin $= 0, 1$) for a total of 17 fundamental particles that—with the exception of gravity—compose all known fields and matter in the universe. The gauge bosons are often described as force carriers as they mediate interactions between particles. Photons mediate the electromagnetic interaction creating both attractive and repulsive forces between all particles that carry electric charge. Similarly the W and Z bosons mediate the weak

Particle Type	Name	Label	Spin	Charge	Mass
Quarks	down	d	$\frac{1}{2}$	$-\frac{1}{3}$	3.5-6.0 MeV
	up	u		$\frac{2}{3}$	1.5-3.3 MeV
	strange	s		$-\frac{1}{3}$	104^{+26}_{-34} MeV
	charm	c		$\frac{2}{3}$	$1.27^{+0.07}_{-0.11}$ GeV
	bottom	b		$-\frac{1}{3}$	$4.20^{+0.17}_{-0.07}$GeV
	top	t		$\frac{2}{3}$	171.2 ± 2.1 GeV
Leptons	electron	e	$\frac{1}{2}$	-1	511 keV
	e-neutrino	ν_e		0	< 2 eV
	muon	μ		-1	106 MeV
	μ-neutrino	ν_μ		0	< 2 eV
	tau	τ		-1	1.77 GeV
	τ-neutrino	ν_τ		0	< 2 eV
Gauge Bosons	gluons	g	1	0	0
	photon	γ		0	0
	W-boson	$\mathrm{W}^{\pm}$		± 1	80.4 GeV
	Z-boson	Z^0		0	91.2 GeV
Scalar Boson	Higgs Boson	H^0	0	0	125.09 ± 0.23 GeV

Table 1.1: Elementary particles of the Standard Model. Masses taken from [13]

interaction between particles that carry weak isospin while gluons mediate the strong interaction between particles that carry colour charge. While photons are electrically neutral, the W and Z bosons as well as gluons, carry their own respective charges allowing for self-interactions.

Fermions, on the other hand, compose all known matter. Due to a phenomenon called colour confinement quarks are perpetually bound to one another forming composite colour-neutral particles called hadrons. The most common hadrons, protons and neutrons, are formed from the quark subset $\{u, d\}$. Exchange of gluons between protons and neutrons keeps nuclei bound together. Electrons then tend to form bound states with nuclei via the electromagnetic interaction through exchange of photons between the electron and nucleus. Thus this small subset of fermions, $\{e, u, d\}$— a mere three out of the known 12—form the atoms and all the elements of the periodic table.

The Higgs boson plays a special role in the Standard Model, it couples to particles with varying strength endowing them with a unique mass.

1.1.3 Gauge theories and Quantum Electrodynamics

A gauge theory is a type of field theory in which the Lagrangian is invariant under certain types of transformations, called gauges transformations. These are continuous transformations that, when taken together, form the basis of their Lie symmetry group.

The simplest example of a gauge transformation is that of a global phase shift. The $U(1)$ abelian group transformation applied to the Dirac Lagrangian demonstrates the physical significance of such symmetry groups. The free Dirac Lagrangian[5] for a spin-$\frac{1}{2}$ field $\psi(x)$

$$\mathcal{L} = \bar{\psi}(i\partial\!\!\!/ - m)\psi \tag{1.5}$$

can be seen to be invariant under the $U(1)$ transformation[6] $\psi \rightarrow e^{i\alpha}\psi$. This invariance in the Lagrangian, and hence the equations of motion, reveals its underlying symmetry $U(1)$. By Noether's theorem [14], this symmetry implies a conserved current, which in this case can be shown [11] to be the electromagnetic charge current density, enforcing that the electrical charge is a globally conserved quantity.

---

[5] Notational note: Dirac's slash notation is being used here, where $\partial\!\!\!/ = \gamma^\mu \partial_\mu$ and γ^μ are Dirac's gamma matrices and $\bar{\psi} = \psi^\dagger \gamma^0$ is the Dirac conjugate of ψ.

[6] Since $\bar{\psi} \rightarrow \bar{\psi} e^{-i\alpha}$.

The Standard Model—and many of its extensions—are gauge theories based on $SU(N)$ symmetries. These are the groups of **S**pecial (determinate equals unity) **U**nitary (each element has an inverse) $N \times N$ matrices. These $N \times N$ matrices operate on N component vectors, spinors or scalars, which are said to belong to the group's *fundamental* representation. In this representation the full set of matrices that furnish $SU(N)$ can be constructed from a set of $N^2 - 1$, $N \times N$ traceless hermitian matrices called *generators*. Thus the generators form a basis of $SU(N)$ spanning a $N^2 - 1$ dimensional space in which the group elements live.

The Standard Model and its extensions are examples of what are called *local* gauge groups. In a $SU(N)$ local gauge group the elements can be represented as unitary operators that are functions of space-time, $U(x)$. In gauge theories a spin $= 0$ or $\frac{1}{2}$ field, $\psi(x)$, transforms under the operation of $U(x)$ as [1]

$$\psi \to \psi' = U\psi = e^{i\vec{\beta}\cdot\vec{T}}\psi,$$

where $\vec{\beta} = \vec{\beta}(x)$ is a vector of arbitrary functions of space-time and $\vec{T}$ are the Lie generators of the symmetry group. In this equation tensor indices are suppressed, however it should be understood that spin-$\frac{1}{2}$ fields are represented as multiplets of Dirac spinors, spin-0 fields as multiplets of scalars and generators as hermitian matrices. Lagrangians that correspond to physical theories require kinetic terms involving the derivative operator ∂^μ. For a typical Lagrangian this term would break the gauge invariance since it operates on $\vec{\beta}(x)$ in the exponential. To enforce gauge invariance one is required to add additional terms to the Lagrangian to cancel these symmetry breaking terms. It is conventional to absorb these extra terms into a redefinition of ∂^μ called the gauge covariant derivative:

$$\partial^\mu \to D^\mu = \partial^\mu + ig\vec{Q}^\mu \cdot \vec{T}, \tag{1.6}$$

where g is an arbitrary gauge coupling and $\vec{Q}^\mu$ are real vector gauge fields, one per Lie generator, that transform as $\vec{Q}_\mu \to \vec{Q}_\mu - \partial_\mu\vec{\beta}(x)$. These are spin $= 0$ massless fields that when quantized can be identified with the gauge bosons of the theory. Hence by postulating gauge invariance of the Lagrangian one finds the existence of gauge bosons is required. Indeed, even gauge boson couplings to fermions are specified by the gauge symmetry.

One could argue that the fact that $SU(N)$ gauge theories are local in space-time is the most important aspect of the symmetry, since all gauge theories are guaranteed to be renormalizable as a consequence of this locality [9]. And Noether's theorem also tells us that there is a conserved charge that is attributed to every continuous symmetry group. This charge is mediated by the gauge fields (or in the quantized theory gauge bosons) allowing them to couple with fermions and spin-0 fields. All modern quantum field theories are based on some symmetry group that exhibits local gauge invariance as it seems to capture, or indeed require, the essence of particle interactions.

For example, in Quantum Electrodynamics (QED) the gauge symmetry $U(1)$ requires by equation 1.6 the introduction of the vector field A^μ

$$\mathcal{L} = \bar{\psi}(i\partial\!\!\!/ - m)\psi - e\bar{\psi}\gamma^\mu\psi A_\mu \tag{1.7}$$

where e is a coupling constant and the quanta of the vector field A^μ are identified as photons. The conserved charge is the familiar electric charge, which is mediated by the photon. The photon couples to other charged particles through the $e\bar{\psi}\gamma^\mu\psi A_\mu$ term creating a (conserved) current. In order to regard A_μ as the physical photon field, the corresponding kinetic energy term needs to be added as well. The requirement of gauge and Lorentz invariance only leave one possibility, the field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.$$

Combining this kinetic term with equation 1.7 yields the full QED Lagrangian

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(iD\!\!\!\!/ - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.8}$$

Since the mass terms of the form $\frac{1}{2}m^2 A_\mu A^\mu$ are forbidden by gauge invariance, the theory correctly predicts that photons are massless.

QED is the simplest example of a gauge theory since it is derived from the trivial Lie group $U(1)$. The Standard Model has a complex group structure, as it is the group product of the $SU(3)$ colour group and the $SU(2) \otimes U(1)$ weak isospin and hyperchage group. Hence QCD and the electroweak model are combined to form the Standard Model based on the $SU(3) \otimes SU(2) \otimes U(1)$ gauge symmetry.

1.1.4 Quantum Chromodynamics

Quantum Chromodynamics is the modern theory of the strong interaction based on the $SU(3)$ symmetry group. In QCD each of the different quark flavours, $q = \{u, d, s, c, b, t\}$, carry an additional quantum number called *colour* , which is the conserved charge that results from the gauging of the $SU(3)$ symmetry group. The quark fields transform as a colour triplet under the fundamental representation of $SU(3)$, where each quark flavour is described by a three component field

$$q = \begin{pmatrix} \psi_r(x) \\ \psi_b(x) \\ \psi_g(x) \end{pmatrix}$$

where the components $\psi_k(x)$ carry the colour index $k = \{r, b, g\}$. To distinguish the $SU(3)$ colour gauge group from other possible $SU(3)$ groups it is labeled with a subscript "C", $SU(3)_C$. Since every particle of the Standard Model has an antiparticle twin with opposite charge there exists an anticolour current as well. A state with one colour index and one anticolour index of the same type will be colour neutral. This is analogous to electromagnetism where a proton with anti-electric charge (positive) and a electron with electric charge (negative) form a neutral bound state. However in addition to this neutral state one could have the three different colour charges combined in a single state which would also be colour neutral. It is because of this that the name colour was chosen, as when one combines light beams of the three primary colours a colourless white light is produced. This is why the three possible values of k are often labeled as *red, green* and *blue.*

$SU(3)_C$ gauge invariance of the QCD Lagrangian requires the introduction of the covariant derivative acting on the quark fields

$$D^\mu = \partial^\mu - \frac{ig_s}{2}\vec{\lambda} \cdot \vec{G}^\mu,$$

where g_s is the strong coupling constant and $\vec{\lambda}$ are the Gell-Mann matrices which form a representation of the $SU(3)$ generators.

There are eight generators of $SU(3)_C$ and hence eight gauge bosons, $\vec{G}^\mu$, associated with QCD. These gauge bosons are the gluons, which carry two colour indices, one colour and one anti-colour . This allows gluons to exchange colour between quarks, mixing colour indices.

The full QCD Lagrangian takes on a very similar form as the QED Lagrangian:

$$\mathcal{L}_{\mathrm{QCD}} = \bar{\psi}(i\not{D} - m)\psi - \frac{1}{4}\vec{G}_{\mu\nu}\cdot\vec{G}^{\mu\nu}, \tag{1.9}$$

where

$$G^j_{\mu\nu} = \partial_\mu G^j_\nu - \partial_\nu G^j_\mu - g_s f^j_{kl} G^k_\mu G^l_\nu$$

are the gluon field strength tensors, analogous to $F_{\mu\nu}$ in QED. However, unlike in QED, the $SU(3)$ group adds gluon-gluon interactions governed by the $SU(3)$ non-abelian commutation relation $[T^j, T^k] = if^{jkl}T^l$, where this relationship manifests itself through the last term via the $SU(3)$ structure constants f^{jkl}. This term allows gluons to couple to one another lending to a rich and intricate phenomenology.

1.1.5 Glashow-Weinberg-Salam Electroweak model

The Glashow-Weinberg-Salam electroweak model is a gauge theory that unifies the electromagnetic and weak interaction based on the $U(1)\otimes SU(2)$ symmetry group. The conserved quantum numbers are *weak* isospin from gauging $SU(2)$ and hypercharge from gauging $U(1)$. Electric charge is given as a combination of weak isospin and hypercharge, thus unifying the two interactions. Weak isospin current can transmute charged leptons into their associated neutrinos or different flavours of quarks into one another. For such an interaction the mass eigenstate basis is inconvenient since particle type is not conserved. This type of weak interaction invites one to interpret leptons and their associated neutrinos as components of a single field, where say the electron and e-neutrino would transform together as a doublet, analogous to how quarks transform as a colour triplet in QCD. A similar treatment needs to be applied to quark fields, however, they transmute between generations requiring quark states to be in a linear combination of one another.

Experimentally it's been observed that $SU(2)$ transformations of the electroweak interaction are particular about the handedness of the fields they operate on, where left and right handed fields are defined as $\psi^L = \frac{1}{2}(1-\gamma_5)\psi$ and $\psi^R = \frac{1}{2}(1+\gamma_5)\psi$ [15]. So fermions are introduced into the electroweak model as "left-handed" (L) doublets

$$L_i = \begin{pmatrix} l_i \\ \nu_i \end{pmatrix}_L, \quad Q_i = \begin{pmatrix} q^u_i \\ q^d_i \end{pmatrix}_L,$$

and “right-handed” (R) singlets,

$$l_i = l_{R,i}, \quad u_i = q^u_{R,i}, \quad d_i = q^d_{R,i}$$

where l labels charged leptons, ν neutrinos, q^u up-type quarks $\{u, c, t\}$ and q^d is a linear combination of down-type $\{s, d, b\}$ quark states. The index i runs over the three lepton generations and quark flavours. The $SU(2)$ symmetry only acts on left-handed fields, giving it the subscript “L”, $SU(2)_L$, while $U(1)$ acts on both left-handed and right-handed fields with hypercharge denoted $U(1)_Y$. The resulting covariant derivative for the $SU(2)_L \otimes U(1)_Y$ symmetry group is

$$D^\mu = \partial^\mu - ig_Y Y B^\mu - \frac{ig_W}{2}\vec{\sigma} \cdot \vec{W}^\mu,$$

where g_Y and g_W are coupling constants and Y and $\vec{\sigma}$ are representations of the generators of $U(1)_Y$ and $SU(2)_L$ respectively. $\vec{W}_\mu = (W^1_\mu, W^2_\mu, W^3_\mu)$ and B_μ are the necessary gauge fields that need to be introduced to make the Lagrangian gauge invariant. Like in QED and QCD these gauge fields are accompanied by their respective field strength tensors $\vec{W}_{\mu\nu}$ and $B_{\mu\nu}$.

1.1.6 Spontaneous symmetry breaking and the Higgs mechanism

As stated earlier gauge bosons need to be massless for gauge invariance. However the W and Z bosons are observed to be massive particles. In fact, due to $SU(2)$ transformations only acting on left-handed states, none of the fermion fields are allowed mass terms in the Lagrangian since $m\psi\bar{\psi} = m(\bar{\psi}^L\psi^R + \bar{\psi}^R\psi^L)$ is not invariant under $SU(2)_L$. This implies that the $SU(2)_L \otimes U(1)_Y$ symmetry is in fact not obeyed—or at least not at the low energy levels from which we observe nature. This observation is made consistent with the electroweak model by postulating that the $SU(2)_L \otimes U(1)_Y$ symmetry is *spontaneously broken* at a higher energy scale. The simplest instrument that achieves this is the Higgs mechanism, which is introduced into the Standard Model through a scalar spin-0 field that transforms as a doublet under $SU(2)_L \times U(1)_Y$. This field, Φ, known as the Higgs field, introduces a potential, $V = V(|\Phi|^2)$ into the Lagrangian.

Consider, for example the Higgs Lagrangian

$$\mathcal{L}_{\text{Higgs}} = D_\mu \Phi D^\mu \Phi + V(|\Phi|^2) \tag{1.10}$$

where

$$V(|\Phi|^2) = \mu^2 |\Phi|^2 + \lambda |\Phi|^4$$

and

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}.$$

With the Higgs field the full electroweak Lagrangian is given by

$$\begin{aligned} \mathcal{L}_{EW} = {} & \underbrace{\bar{Q}_i i \not{D} Q^i + \bar{u}_i i \not{D} u^i + \bar{d}_i \not{D} d^i + \bar{L}_i i \not{D} L^i + \bar{l}_i i \not{D} l^i}_{\text{fermions}} \\ & \underbrace{-\frac{1}{4}\vec{W}^{\mu\nu} \cdot \vec{W}_{\mu\nu} - \frac{1}{4} B^{\mu\nu} \cdot B_{\mu\nu}}_{\text{gauge bosons}} \underbrace{+ D_\mu \Phi D^\mu \Phi - V(|\Phi|^2)}_{\text{Higgs boson}} \\ & \underbrace{+ \bar{Q}_i Y_u u^i \Phi + \bar{Q}_i Y_d d^i + \bar{L}_i Y_l l^i \Phi^\dagger + h.c.}_{\text{Yukawa}} \end{aligned} \tag{1.11}$$

where Y_u, Y_d and Y_l are complex matrices representing Yukawa couplings induced by the Higgs field and *h.c.* stands for hermitian conjugates.

If the parameters μ and λ happen to take on the configuration $\mu^2 < 0$ and $\lambda > 0$ then one can see from inspection of $V(|\Phi|^2)$ that $|\Phi|^2 = 0$ does not correspond to a local minimum, but instead the minimum is found at $|\Phi|^2 = -\frac{\mu^2}{2\lambda} \equiv \nu^2$. When quantized ν is known as a non-vanishing vacuum expectation value. In such a case the Higgs field can be expressed in terms of the vacuum expectation value as

$$\Phi(x) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ \nu + H(x) \end{pmatrix}$$

where $H(x)$ is the physical Higgs field. This result shows that spontaneous symmetry breaking has picked out a preferred direction in weak isospin plus hypercharge space breaking the $SU(2)_L \times U(1)_Y$ symmetry. However the global $U(1)$ symmetry of QED still exists.

The breaking of $SU(2)_L \times U(1)_Y$ induces a mixing of the $\vec{W}^\mu$ and B^μ gauge fields in their mass eigenstates. The results are

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W,$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W,$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2),$$

where θ_W is the weak mixing angle, an empirical parameter that is defined by $e = g_Y \cos\theta_W = g_W \sin\theta_W$. After spontaneous symmetry breaking W and Z bosons kinetic terms such as $\frac{1}{2}m_Z^2 Z_\mu Z^\mu$ where $m_Z = \frac{1}{2}\nu\sqrt{g_W^2 + g_Y^2}$ appear in the Lagrangian endowing the vector boson with mass. Indeed, the Higgs field itself picks up a mass term $\frac{1}{2}m_h^2 H^\dagger H$ where $m_h = \sqrt{2\lambda}\nu$. The different couplings between the Higgs field and the fermions in the Yukawa term determine the masses of fermions and leptons.

The quanta of these four gauge fields correspond to the observed gauge bosons of the Standard Model. Strictly speaking the Standard Model Higgs boson does not couple to neutrinos (or any right handed particles) and therefore cannot explain observed neutrino oscillations, as this requires neutrinos to have non-zero mass eigenstates. However minimal extensions to the Standard Model, such as adding right handed neutrinos (Dirac mass) or combining left-handed neutrino with their complex conjugate (Majorana mass), allow the Higgs boson to couple to neutrinos giving them a non-zero mass.

1.2 Evidence for dark matter

The Swiss astrophysicist Fritz Zwicky is usually credited with the first observation of dark matter. In 1933 Zwicky used the virial theorem to argue the existence of unseen matter in the Coma galaxy cluster [16]. The virial theorem states that the total kinetic energy of the Coma cluster should be half the total gravitational binding energy. However, when Zwicky studied the motion of galaxies on the Coma cluster's outer edge and compared their kinetic energy to the gravitational binding energy of the cluster, Zwicky found that the Coma cluster contained 400 times more mass than was accounted for from luminous matter. Zwicky inferred that there must exist some non-luminous matter to account for this discrepancy.

Since Zwicky's initial observation there has seen a plethora of evidence put forward supporting the idea of a pervasive distribution of non-luminous matter in the

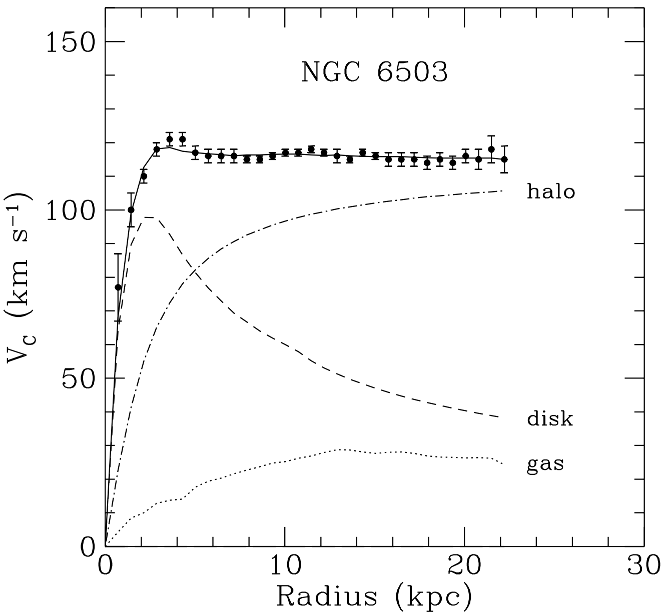


Figure 1.1: Rotation curve for the spiral galaxy NGC6503. The points are the measured circular rotation velocities as a function of distance from the galactic centre. The dashed and dotted curves are the contribution to the rotational velocity due to the observed disk and gas, respectively, and the dot-dash curve is the contribution from the dark halo. Image taken from [17]

universe. Evidence has been collected from many different astrophysical and cosmological structures on vastly different scales; evidence that ranges from individual galaxies to the universe on the whole. Below briefly summarizes some of the strongest evidence for dark matter to date.

1.2.1 Galaxy rotation curves

Zwicky's initial discovery was largely ignored until discrepancies in spiral galaxy rotation curves started to emerge. Observations from numerous galaxies showed that the stellar rotational velocities remained constant, or "flat", with increasing distance from the galactic centre. This result was surprising and highly counterintuitive since Newton's law of gravitation suggests that the rotational velocities should decrease with distance from the galactic centre, resulting in a distribution with a "falling tail." Fig. 1.1 shows a typical rotation curve of a spiral galaxy. The expected curve would be the sum from the disk and gas contributions, resulting in a curve that drops off

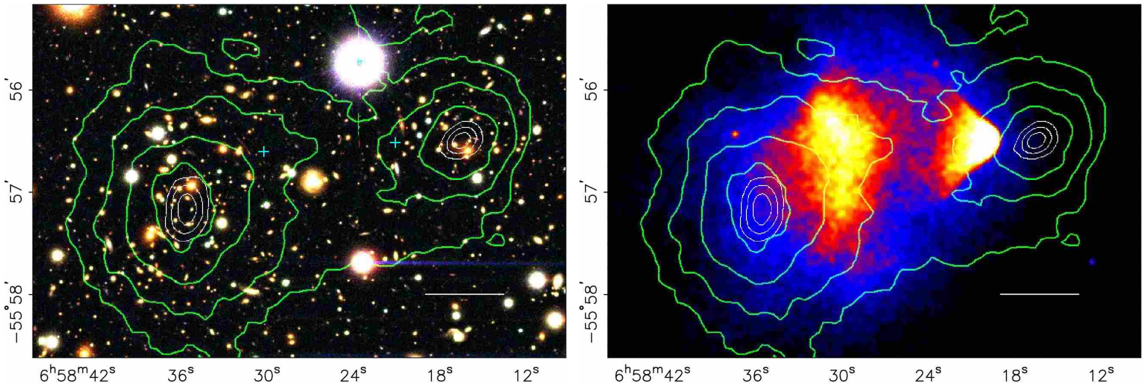


Figure 1.2: (Left) Hubble Space Telescope image of merging cluster 1E0657–558 (also known as the Bullet Cluster). (Right) Chandra X-ray Observatory image of the same system, showing x-ray emission intensities. In both images green contours show the distribution of mass as calculated using weak gravitational lensing techniques. The white bar indicating 200 kpc at the distance of the cluster. Images taken from [19].

as the radius increases. However, the observed data does not show this trend; the data shows that stellar velocities remain constant as the radius from the galactic centre increases. The observed rotation curves can be accounted for if one assumes a spherical dark matter distribution, often called a "halo", around the galaxy. This allows the mass density to remain constant, rather than concentrated at the galactic bulge. It has been found that the average mass-luminosity ratio for spiral galaxies is about six times larger than estimated [18], showing that most—if not all—galaxies are dominated by dark matter.

1.2.2 Gravitational lensing effects and the Bullet Cluster

In addition to rotation curves, gravitational lensing techniques can be used to estimate the mass of large astrophysical structures such as galaxy clusters. Gravitational lensing can be observed when light from a distant object, such as a quasar, is bent around a massive object, such as a galaxy cluster, and into the observers line of sight. This can lead to double images in the case of strong gravitational lensing or a shearing distortion in the case of weak lensing. By examining deformations of background galaxies the distribution of dark matter has been found to be consistent with other dark matter measurements of large scale structures [19].

The most direct evidence for dark matter comes from a system of two colliding galaxy clusters, together referred to the Bullet Cluster. Fig. 1.2 shows the Bullet

Cluster as imaged by the Hubble Space Telescope (left) and the Chandra X-ray Observatory (right). The green contours indicate the mass distribution as calculated through weak lensing techniques [7]. The X-ray observations from Chandra show that baryonic matter, in the form of gas or plasma, is concentrated in the centre of the system. As the two clusters passed through one another these gases were slowed by electromagnetic interactions causing them to get stuck in the centre. However, the weak gravitational lensing measurement shows that much of the systems mass lies outside of this central region. The interpretation being that the dark matter (as well as the compact stellar matter), which does not interact via the electromagnetic interaction, passed straight through unaffected by the baryonic gas or the dark matter from the other cluster. This is an important observation as it implies that dark matter has a small interaction cross-section. It should also be noted that this demonstrates clear evidence for dark matter that is independent of Newtonian gravity, in contrast to rotation curves.

1.2.3 The Cosmic Microwave Background

The Cosmic Microwave Background (CMB) contains evidence on cosmological scales of dark matter through temperature fluctuations imprinted in the early universe. Fig. 1.3 shows these angular fluctuations in the CMB spectrum as measured from the Planck space observatory. The anisotropies in the CMB were caused by acoustic oscillations in the photon-baryon plasma of the early universe prior to the emission of the CMB, when photons decouple from the baryonic plasma [6]. The restoring force for these oscillations was due to gravity produced from baryonic, as well as, dark matter. Inflation then caused these very small anisotropies to be stretched out onto cosmological scales. Since baryonic matter interacts with photons while dark matter does not the two types of matter are expected to have different effects on the power spectrum. Baryonic matter plays a role in both propagation and dampening of the acoustic oscillations while dark matter only effects the dampening through its gravitational attraction. In Fig. 1.3 the red curve shows the best fit for the ΛCDM model; the leading cosmological model, containing a flat space-time, dark energy and cold (non-relativistic) dark matter. The best fit (maximum likelihood) value for the current dark matter density, $\Omega_{DM}h^2$, is 0.1199 ± 0.0027. Counting from left to right the first peak is primarily caused by a flat spacetime curvature, the second from baryonic matter and the third is largely contributed to the presence of dark matter

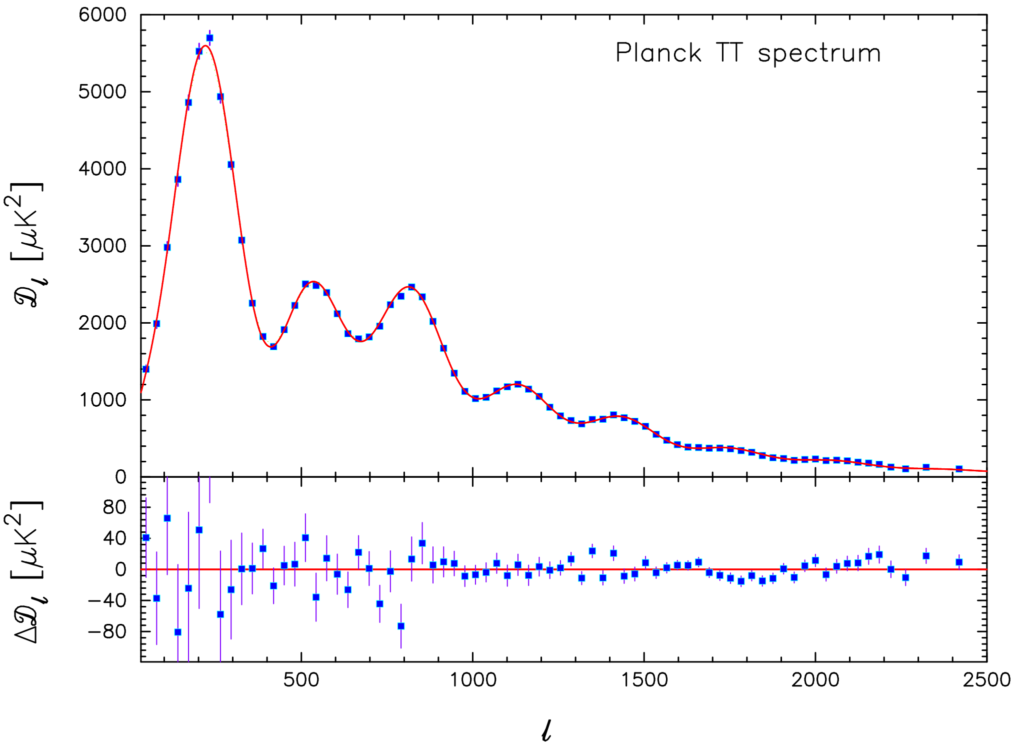


Figure 1.3: Planck temperature power spectrum. Points in upper panel show foreground-subtracted power spectrum as measured by the Planck space observatory. Red curve shows best fit for ΛCDM model. (Lower panel) residuals of measure Planck power spectrum to best fit of theoretical model. Plot taken from [20].

in the early universe. The ΛCDM model can been seen to have an excellent fit to Planck data and, since it has cold dark matter built into it, the CMB spectrum can be viewed as evidence for cold dark matter.

1.3 Dark matter candidates

The first dark matter candidates proposed were made up of ordinary baryonic matter, such as black holes, neutrons stars and faint white dwarfs. However, searches for these objects in our galaxy have shown that they can only contribute a small portion of the missing matter [21]. Today the most commonly held view is that dark matter is primarily non-baryonic, composed of one or more new elementary particles that are not part of the Standard Model. Some of the best motivated candidates for dark matter go under the label of Weakly Interacting Massive Particles, or WIMPs. WIMPs, as the name implies, are usually assumed to have weak scale interactions and mass; however, the relic abundance, Ω_{DM}, only constrains the ratio of the two [22]. It is also possible that WIMPs interact with one another through some unknown mediator particle, which would constitute a new fundamental force. Such a scenario allows for a rich and elaborate phenomenology in the “dark sector” that could be completely independent of the Standard Model.

The observed value of the relic abundance motivates what is usually regarded as the strongest argument for the WIMP interpretation of dark matter, the so called “WIMP miracle”.

1.3.1 The WIMP miracle

The WIMP miracle refers to the fact that the observed abundance of dark matter in the universe today is consistent with thermally produced dark matter having a mass and interaction scale roughly equal to that of the weak interaction, defined by the vacuum expectation value $\nu = 246$ GeV. Figure 1.4 illustrates this argument. In the early universe Standard Model and dark matter particles were in thermal equilibrium as they were created and annihilated at equal rates. Then the universe started to expand and cool resulting in an exponential drop in matter production. At a certain point the energy in the universe dropped below the production threshold and the amount of remaining dark matter “froze out”, or remained constant. If the relic abundance is calculated assuming weak scale mass and interactions one finds a

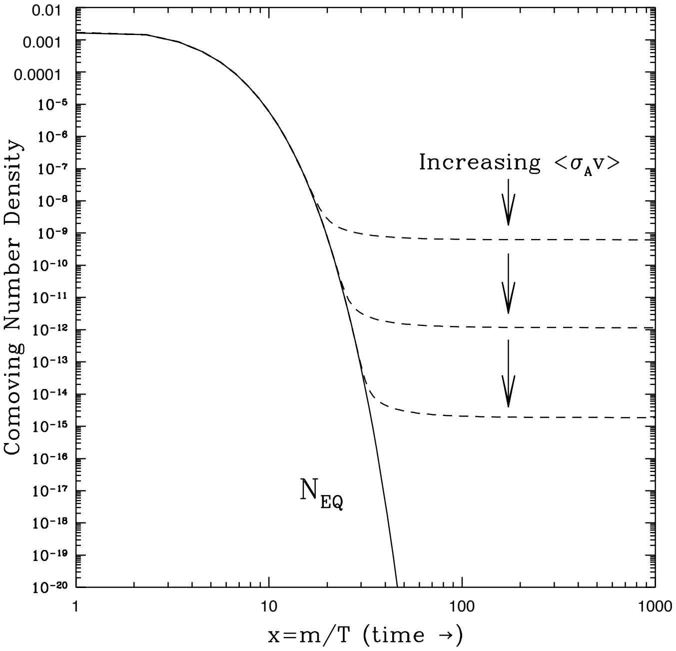


Figure 1.4: Comoving number density (the number density after adjusting for volume expansion) of WIMPs in the early Universe. The dashed curves are the actual abundance, for a given thermally averaged annihilation cross section $\langle \sigma v \rangle$) and the solid curve is the equilibrium abundance. The x-axis, WIMP mass over average temperature m_{DM}/T, increases with time as the universe expands and cools. Image taken from [17]

number that is on the same order of magnitude as what is observed, $\Omega_{DM} \approx 0.24$. Considering the scale on figure 1.4, which spans roughly 20 orders of magnitude, this would seem pretty miraculous if it were in fact just a coincidence.

However, the mass and interactions of the particles does not necessarily have to be on the order of the weak scale. The constraint given by the relic density is on the ratio of the two [22]:

$$\Omega_{DM} \propto \frac{1}{\langle \sigma v \rangle} \sim \frac{m_{DM}^2}{\lambda_{DM}^4} \tag{1.12}$$

where m_{DM} and λ_{DM} are the mass and coupling of the dark matter particles and $\langle \sigma v \rangle$ is the thermally averaged annihilation cross section.

1.4 Dark Matter searches

In developing a model for WIMP interactions and dynamics it is instructive to summarize what properties of about dark matter are already known:

1. It's neutral under electric charge, since it does not produce photons,
2. It's stable, or at least has a lifetime on cosmological scales,
3. It's non-baryonic, to preserve the success of ΛCDM,
4. It has a mass and Standard Model coupling consistent with the observed relic abundance.

This points one to a weakly interacting massive particle that is stable, neutral, and not a quark bound state. In order to observe dark matter one must assume that it interacts with Standard Model particles in some way beyond just gravity. However, the above observations tell one nothing about the spin of dark matter nor that of its interactions in the dark sector (also known as "dark forces").

One class of theories that satisfy the above requirements are so called "Higgs-portal" theories. These theories make very few assumptions about the nature of dark matter or dark forces. Indeed, the only considerable assumption made is that dark matter obtains mass through the same mechanism as Standard Model particles, that is the Higgs mechanism.

1.4.1 The Higgs-portal

This section describes the Higgs-portal model of dark matter originally put forward by Ref.[23]. In the Higgs-portal model WIMPs are introduced by extending the Standard Model gauge symmetries to include a global Z_2 parity. This is the simplest symmetry that can be introduced that guarantees the stability of dark matter. In this scenario WIMP dark matter experiences an odd charge while all Standard Model particles are even. Dark matter transforms as a singlet under the Standard Model symmetries and thus does not couple with any of the gauge bosons (W/Z bosons, photons, gluons). However, it is assumed that dark matter mass is generated through the same Higgs mechanism as in the Standard Model. Three cases for the spin of dark matter are considered: real scalar dark matter S, Majorana fermion dark matter χ, and real vector dark matter V_μ.

The Lagrangians, which are invariant under the Standard Model symmetries, can be written as

$$\mathcal{L}_S = \mathcal{L}_{SM} + \frac{1}{2}(\partial_\mu S)(\partial^\mu S) - \frac{1}{2}M_S^2 S^2 - \frac{1}{2}\lambda_S|\Phi|^2 S^2 - \frac{1}{4!}\kappa_S S^4, \tag{1.13}$$

$$\mathcal{L}_F = \mathcal{L}_{SM} + \frac{1}{2}\bar{\chi}(i\gamma^\mu\partial_\mu - M_F)\chi - \frac{\lambda_F}{2\Lambda}|\Phi|^2\bar{\chi}\chi - \frac{\kappa_F}{2\Lambda}\bar{\chi}\sigma^{\mu\nu}\chi B_{\mu\nu}, \tag{1.14}$$

$$\mathcal{L}_V = \mathcal{L}_{SM} - \frac{1}{4}V^{\mu\nu}V_{\mu\nu} + \frac{1}{2}M_V^2 V_\mu V^\mu + \frac{1}{2}\lambda_V|\Phi|^2 V_\mu V^\mu - \frac{1}{4!}\kappa_V(V_\mu V^\mu)^2, \tag{1.15}$$

where $V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$, $B_{\mu\nu}$ is the field strength tensor of the hypercharge gauge boson, and $\mathcal{L}_{SM} = \mathcal{L}_{QED} + \mathcal{L}_{QCD} + \mathcal{L}_{EW}$ is the full Standard Model Lagrangian. The quartic terms in equations 1.13 and 1.15 correspond to dark matter self-interactions. For the fermion case the theory is not renormalizable, and thus relies on a energy cut-off scale Λ. It can also been seen that the last term proportional to κ_F shows that there is an interaction between dark matter and the hypercharge gauge boson, however this interaction most likely proceeds through higher order loop diagrams of new physics dynamics since the dark matter considered here does not carry hypercharge.

The terms proportional to λ show that dark matter fields couple to the Higgs field directly, allowing for Higgs-WIMP interactions—this is what is referred to as the “Higgs-portal”. These Higgs-WIMP interactions are observable through a Higgs exchange in WIMP-nucleon scattering or through Higgs decay to WIMPS. The former may be observed at direct detection experiments such as XENON100, CoGeNT and CRESST-II while the latter at collider experiments such as the Tevatron and the LHC.

1.4.2 Direct detection searches and effective field theories

Direct detection experiments such as XENON100 [24], CRESST [25] and CoGeNT [26] rely on WIMP-nucleon scattering in ultra-low background laboratories deep underground. Direct detection experiments tend to use one of two detector technologies. The first are cryogenic detectors, such as CRESST and CoGeNT, operating at temperatures below 100 mK to detect heat produced when a particle hits germanium crystals in the detector’s fiducial volume. The second type are noble liquid detectors, such as XENON100, that use xenon or argon to detect scintillation light produced by a particle collision.

Compared to collider experiments such as the LHC the collision energy, or *mo-*

mentum transfer Q_{tr}, between the WIMP and nucleon is very small. In this low energy regime heavy particles such as the Higgs will not be produced on mass shell. Instead interactions will be mediated through a Higgs resonance which can be modelled with an effective field theory. Effective field theories can be constructed by "integrating out"[7] the mediator particle . After the integration one is left with only a contact interaction term between the Standard Model and the dark matter particles. The easiest way to see this is by taking the propagator and expanding it around the momentum transfer Q_{tr}. If one then drops everything except the leading term all that is left is a constant which describes the contact interaction:

$$\frac{1}{Q_{tr}^2 - m_h^2} = -\frac{1}{m_h^2}\left(1 + \frac{Q_{tr}^2}{{m_h}^2} + \mathcal{O}\left(\frac{Q_{tr}^4}{m_h^4}\right)\right) \approx -\frac{1}{m_h^2} \tag{1.16}$$

This approximation is only valid if $Q_{tr}^2 \ll m_h^2$ otherwise all other terms in the expansion, i.e. the UV complete theory, must be considered since Q_{tr} and m_h scale together.

After integrating out the Higgs boson equations 1.13–1.15 lead to the effective Lagrangians [23]

$$\mathcal{L}_S^{Eff} = \frac{\lambda_S}{2m_h^2} S^2 \left(\sum_q m_q \bar{q} q - \frac{g_s}{2\pi} G_{\mu\nu} G^{\mu\nu}\right), \tag{1.17}$$

$$\mathcal{L}_F^{Eff} = \frac{\lambda_F}{2\Lambda m_h^2} \bar{\chi}\chi \left(\sum_q m_q \bar{q} q - \frac{g_s}{2\pi} G_{\mu\nu} G^{\mu\nu}\right), \tag{1.18}$$

$$\mathcal{L}_V^{Eff} = \frac{\lambda_V}{2m_h^2} V_\mu V^\mu \left(\sum_q m_q \bar{q} q - \frac{g_s}{2\pi} G_{\mu\nu} G^{\mu\nu}\right), \tag{1.19}$$

where q represents the light quark field $q = \{u, d, s\}$ with masses m_q. It can be seen that in this effective field theory the dark matter fields couple directly with the quark and gluon fields. This allows for WIMP-nucleon scattering (e.g. $\chi N \to \chi N$) of the type that direct detection experiments are sensitive to.

---

[7] The integration here refers to the path integral formulation of QFT

Chapter 2

ATLAS and the LHC experiments

2.1 Hadron collider physics and the LHC

Particle colliders accelerate charged particles with electromagnetic fields in opposite directions along paths that are diverted into one another at certain interaction points (IP). The Large Hadron Collider (LHC) is currently the most powerful particle collider in the world. It is designed to collide beams of hadrons—protons or lead ions—at unprecedented energies. The LHC is installed 100 m below the Franco-Swiss border in a 27 km long tunnel that was formerly occupied by the LEP accelerator at the European Organization for Nuclear Research (CERN). Protons are accelerated through a succession of smaller accelerators before being injected into the LHC where they are boosted to a terminal velocity of 99.9999991% the speed of light [27]. Acceleration of hadrons in the LHC is achieved through the use of radio frequency accelerator cavities that are tuned to a frequency and field orientation that gives the protons a push forward through each cavity. The proton beams are directed and focused around the ring though a series of dipole and quadrupole magnets.

There are four main interaction points around the LHC ring where the proton beams are squeezed and directed into one another. Four independent detectors are installed at these interaction points to record the resulting proton-proton collisions. ALICE and LHCb are specialized experiments devoted to the study of heavy ion collisions and B-physics respectively, while CMS and ATLAS are often referred to as “discovery machines” as they are general purpose experiments designed to be sensitive to a wide variety of particles signatures.

2.1.1 LHC parameters and observables

When two protons collide at the IPs the total energy squared in the centre-of-momentum frame (CoM) can be expressed in terms of the Lorentz-invariant Mandelstam variable

$$s = (E_1 + E_2)^2 \tag{2.1}$$

where E_1 and E_2 are the energies of the two protons. The LHC has an impressive CoM energy of up to $\sqrt{s} = 14$ TeV [28]. However, during 2012 operations the LHC ran at only $\sqrt{s} = 8$ TeV due to a faulty electrical connection that led to a magnet quench in 2008 [29]. The CoM energy $\sqrt{s}$ is possibly the most important beam parameter for a collider, as it determines what physical processes the collider is capable of accessing, since the cross-section σ of a scattering event is a function of the CoM energy, $\sigma = \sigma(s)$. This relationship is true for both the *underlying event* of the proton-proton scattering as wells as the *hard scattering*—defined by a large momentum transfer—of the quarks and gluons (σ_{hard}) in which one is typically interested. In the case of the hard scattering this relationship shows up in both the Lorentz invariant flux $\mathcal{F}$ and the probability amplitude $\mathcal{M}$ that encodes the Lagrangian (equation 1.2) through the differential equation

$$\mathrm{d}\sigma_{hard} = \frac{1}{\mathcal{F}} |\mathcal{M}|^2 \mathrm{d}\Pi \tag{2.2}$$

where $\mathrm{d}\Pi$ is the Lorentz invariant phase space volume element.

Another important parameter for colliders is the instantaneous luminosity, defined as the the number of particles passing each other per unit time through unit transverse area at the interaction point. At the LHC proton beams come in discontinuous *bunches* of particles. If there are n_1 particles in each bunch of the first beam and n_2 for the second, then the instantaneous luminosity L is given by

$$L = \frac{f n_1 n_2}{4\pi \sigma_x \sigma_y}, \tag{2.3}$$

where f is the frequency at which the bunches collide and σ_x and σ_y are the root mean square deviation of the transverse beam size in the horizontal and vertical directions. In 2012 the LHC ran with roughly 1380 bunches per beam with a collision frequency of $f = 20$ MHz corresponding to an instantaneous luminosity of about 2.5×10^{33} cm^{-2}s^{-1} [28]. It should be noted, however, that the beam parameters are not constant and vary somewhat during operations, and thus should be thought of as a function of time.

The main observable in experimental particle physics is the reaction rate of the process of interest, which is given by the instantaneous luminosity times the scattering cross-section

$$R = \sigma L. \tag{2.4}$$

To get the total number of collisions one simply needs to integrate the instantaneous luminosity over time and multiply by the relevant cross-section

$$N(t) = \int_0^t R \mathrm{d}t = \sigma \int_0^t L \mathrm{d}t. \tag{2.5}$$

Since one typically gives the cross-section in units of *barns*, where 1 $\mathrm{cm}^2 = 10^{24}$ b, it is useful to measure the LHC's integrated luminosity in units of inverse barns, b^{-1}. During 2012 operation the LHC delivered a total of 22.8 fb^{-1} of integrated luminosity (see section 3.3 for more detail on 2012 data collection).

2.1.2 Hadrons, jets and colour confinement

Proton-proton scattering at the LHC can be quite complex since hadrons are composite objects, composed of quarks and gluons—often together referred to as *partons*. Gluons are appropriately named as they hold quarks in colour-charge-neutral, or *colourless*, bound states. For example, colour interactions between three quarks will form a colour singlet baryon bound state by contraction of the anti-symmetric tensor $\epsilon_{\alpha\beta\gamma}q^\alpha q^\beta q^\gamma$ [10]. This state is colourless since the anti-symmetric tensor ϵ ensures that all three indices are different, resulting in a colour-neutral state. The other possible colour singlet states are the antibaryon state, $\epsilon^{\alpha\beta\gamma}\bar{q}_\alpha\bar{q}_\beta\bar{q}_\gamma$, and the quark-antiquark meson state, $q^\alpha\bar{q}_\alpha$. Since gluons carry colour charge themselves they can couple to one another, theoretically creating colour-neutral bound states called "glueballs"—however such states have yet to be confirmed in nature. The quarks that make up colour neutral configurations of baryon and meson states are called *valence* quarks. In addition to valence quarks a fluctuating sea of virtual gluons and neutral $q\bar{q}$ pairs engulf the valence quarks within a hadron. These virtual partons are often ignored as they do not affect the quantum numbers of the hadron. However, in high energy collisions it is possible to scatter valence quarks off of sea partons.

This sea of virtual gluons also plays a role in the unexpected strength of the strong force. Since gluons are massless one may expect that the force required to separate two quarks would scale as the inverse squared force law as with photons. However gluons, unlike photons, carry colour charge themselves allowing exchanging gluons to induce a vacuum polarization in the virtual gluon sea surrounding valence quarks. This creates a string of gluons holding quarks together as if connected by a spring. Hence, as the distance between quarks increases so does the strong force holding them together. This phenomena is called *colour confinement.*

As a result of colour confinement the strength of the strong force scales linearly with distance. Thus scattered quarks within a hadron bound state will resist separating from the hadron. Instead it is energetically favourable for the system to create new quark-antiquark pairs in the sea, which may split to form a new bound state with the scattered quark creating a new hadron. This allows the string of gluons connecting the scattered quark and the incident quark to be broken into two, a process known as *fragmentation.* With high energy scattering the broken strings of gluons can continue to produce quark-antiquark pairs, which are produced with momenta collinear to the initial direction of the scattered quark. The quarks then coalesce into colourless bound states (hadronization) leading to a collimated spray of hadrons known as a *jet.* At hadron colliders it is these jets that are observed rather than quarks directly.

2.1.3 Hadron scattering and parton distribution functions

As mentioned, equation 2.2 is not directly applicable to processes at the LHC since hadrons are composite objects. Consider the scattering process of two incident hadrons A and B that produce an elementary particle c ($c =$ quark, lepton or W/Z boson) plus anything else X,

$$A + B \rightarrow c + X.$$

It is processes such as these that one may observe at LHC; however, typically it is the subprocess of the hadron constituents, that is the hard scatter, that one is interested in studying. Labelling the scattered partons from A and B as a and b respectively this process is

$$a + b \rightarrow c + X.$$

The momentum of the individual partons a and b will not be known. In hadron collider experiments all one knows is the momentum of the hadrons being collided and that this must be equal to the sum of the momenta of their parton constituents. However, if the parton momentum density distribution, often called the *parton density function* (PDF), in the hadron is known one can integrate over all possible momenta. This leads to the convention of calculating cross sections with a parton's fractional momentum $x = p(\text{parton})/p(\text{hadron})$, where only the component of momentum along the beam axis is considered. The PDF of parton a of hadron A, $f_{a/A}$, is given as a function of the momentum fraction of a, x_a, and the momentum transfer of the process, Q_{tr}. The cross-section $\sigma(AB \to cX)$ may be obtained by multiplying the subprocess cross-section $\sigma(ab \to cX)$ by $\mathrm{d}x_a f_{a/A}(x_a, Q_{tr}^2)$ and $\mathrm{d}x_b f_{b/B}(x_b, Q_{tr}^2)$, summing over parton and antipartion types a, b, integrating over x_a and x_b, and then averaging over the colours of a and b [9]. Thus the hadron process cross section is given by

$$\sigma(AB \to cX) = K \sum_{a,b} C_{ab} \int_0^1 \mathrm{d}x_a \mathrm{d}x_b f_{a/A}(x_a, Q_{tr}^2) f_{b/B}(x_b, Q_{tr}^2) \sigma(ab \to cX), \quad (2.6)$$

where C_{ab} are colour averaging factors and K is a constant that may be necessary for perturbative corrections, known as a *K-factor*.

It is not possible to calculate PDFs perturbatively due to non-perturbative QCD binding effects, instead they must be measured in the laboratory. PDFs for various values of Q_{tr}^2 are extracted from large datasets from various groups worldwide. Some such datasets and collaborations are:

- **CTEQ** [30], The CTEQ Collaboration;
- **MRST** [31], A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne;
- **GRV** [32], M. Gluck, E. Reya, and A. Vogt;
- **GJR** [33], M. Gluck, P. Jimenez-Delgado, and E. Reya;
- **NNPDF** [34], the NNPDF Collaboration.

The function used to fit the PDF and the number of free parameters will depend upon the value of Q_{tr}^2. In general, the total number of free parameters is quite large. For example the CTEQ6.6 PDF from the CTEQ Collaboration uses a total of 22 free parameters [35].

2.2 The ATLAS Detector

The unprecedented energy and luminosity of the LHC provides for a rich physics potential of discoveries and precision measurements. With the LHC, Standard Model parameters can be measured at world leading accuracies and the discovery reach for new physical phenomena is unrivalled. The ATLAS (**A** **T**oroidal **L**HC **A**pparatus**S**) detector is one the two general purpose detectors designed to exploit the full discovery potential of the LHC. To achieve this ATLAS was designed to be sensitive to a wide range of new and possibly unexpected physics signals, which requires that certain performance goals be achieved. The basic design criteria of the detector as stated in the technical design report [36] are:

- Very good electromagnetic calorimeter for electron and photon identification and measurements, complemented by full-coverage hadronic calorimetry for accurate jet and missing transverse energy (E_T^{miss}) measurements;
- High-precision muon momentum measurements, with the capability to guarantee accurate measurements at the highest luminosity using the external muon spectrometer alone;
- Efficient tracking at high luminosity for high-p_T lepton-momentum measurements, electron and photon identification, τ-lepton and heavy-flavour identification, and full event reconstruction capability at lower luminosity;
- Large acceptance in pseudorapidity (η) with almost full azimuthal angle (ϕ) coverage everywhere.
- Triggering and measurements of particles at low-p_T thresholds, providing high efficiencies for most physics processes of interest at LHC.

In order to achieve these requirements, ATLAS is composed of a number of sub-detector systems that operate largely independently of one another. Fig. 2.1 displays an overview of the ATLAS detector with its labeled sub-detectors and components. The main sub-detectors and components of ATLAS can be divided into four systems:

- **Inner Detector** for measuring the trajectories and vertices of charged particles,
- **Calorimeter** for energy measurements and identification of electromagnetic and hadronic particles,

- **Muon spectrometer** for measuring the trajectory and momenta of muons,
- **Magnet system** for bending the trajectories of charged particles providing momentum and charge measurements,
- **Trigger/DAQ** for quickly sorting through events, saving ones that are deemed to contain interesting physics based on a predefined set of selection criteria for offline analysis.

A brief overview of these systems, and their sub-systems, is provided in this section. For a more detailed description of these systems one is referred to [37].

2.2.1 Detector geometry, coordinate systems and nomenclature

The geometry, coordinate system and nomenclature used to describe the ATLAS detector and reconstructed detector objects is briefly described here. The geometry of the ATLAS detector is cylindrical with the origin defined to be the nominal IP, where the counter-rotating proton beams are directed into one another. In Cartesian coordinates the z-axis, also referred to as the longitudinal direction, is defined to lie along the beam axis, while the x-y plane, often referred to as the transverse plane, is normal to the beam axis. The positive x-direction is defined to point towards the centre of the LHC ring from the IP while the positive y-direction points upwards. In cylindrical coordinates the azimuthal angle ϕ is defined in the transverse plane, measured around the z-axis while a radius coordinate, R, defines the radial distance from the z-axis. In spherical coordinates the additional angle θ is defined as the polar angle measured from the z-axis.

The longitudinal momentum of scattered particles at hadron colliders has a relatively large associated uncertainty and can vary significantly from event to event. This uncertainty is due to the fact that the initial momentum of the incident partons is unknown and that the ATLAS detector has a limited polar acceptance. The longitudinal rapidity of a particle, defined as $y = 1/2 \ln[(E + p_z)/(E - p_z)]$, is often used at hadron colliders since rapidities are additive under Lorentz boosts; hence the difference between two rapidities is boost invariant. The pseudorapidity, η, approximates rapidity in the massless limit and is defined with only the polar angle θ: $\eta = -\ln\tan(\theta/2)$. For high momentum particles where $m \ll |p|$ this is a good

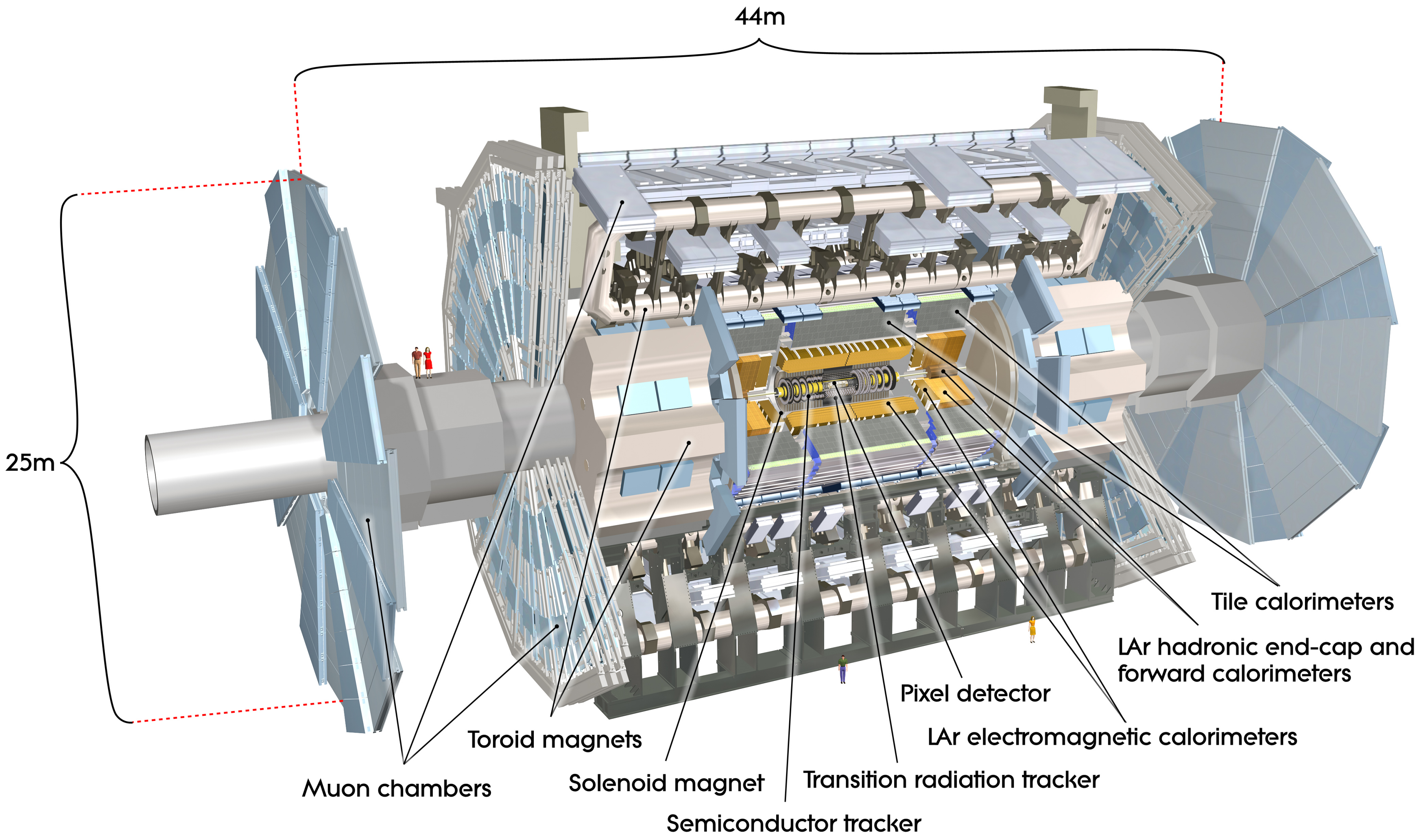


Figure 2.1: Cut-away view of the ATALS detector with its labeled sub-detectors. Image taken from [38]

approximation, and for this reason η has been adopted by ATLAS as the polar coordinate instead of θ. It is also often useful to use the distance ΔR in ϕ-η space between two points (ϕ_1, η_1) and (ϕ_2, η_2) which is defined as $\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}$.

The kinematic variables often used to describe particles such as momentum, p, energy, E, and mass, m, are more conveniently defined in the transverse plane due to the aforementioned uncertainties. For example, the *transverse momentum*, p_T, is simply the p projected into the transverse plane and the transverse energy, E_T, is defined by the projection $E_T = E \sin\theta$. E_T can again be projected along either the x-axis or y-axis in the transverse plane to define the components of the transverse energy vector, E_T. By conservation of energy-momentum the vector sum running over all scattered particle's E_T should be zero. However not all particles can be reconstructed with the ATLAS detector (e.g. neutrinos) and thus the measured sum E_T may not be zero. The magnitude of the negative vector sum of measured E_T is referred to as the *missing transverse energy*, E_T^{miss}, which is associated with scattered particles that escape detection.

2.2.2 Magnet systems

Moving charged particles in a magnetic field are subject to the magnetic force component of the Lorentz force, causing them to accelerate in a direction perpendicular to the magnetic field and their direction of motion. For a constant magnetic field charged particles will be pulled from their paths into a helix of a constant radius proportional to their momentum. By inducing a magnetic field of known field strength and by measuring a particle's radius of curvature, the Lorentz force can be exploited to make a momentum measurement. Additionally, particle charge can be measured by noting the direction of curvature, allowing discrimination between particles and antiparticles.

The ATLAS detector employs two magnet systems to measure the momentum and charge of traversing particles: a solenoid surrounding the ID and a system of three large toroids within the muon spectrometer. The central solenoid uses superconducting electromagnets to induce a 2 T magnetic field within the ID. The toroid system is composed of a barrel and two end-caps. There are eight barrel region toroidal coils that are arranged in a cylinder with an 8-fold symmetry around the calorimeter producing a magnetic field of 0.5 T for the barrel muon detectors. End-cap toroids are installed on both sides of the barrel toroid system producing a 1 T magnetic field within the detectors end-cap regions. Combined the magnet system provides a magnetic field of strength greater than 50 mT over a volume of approximately 12000 m^3. More details on the ATLAS magnet systems can be found in [39].

2.2.3 Inner Detector

The ATLAS Inner Detector (ID) is designed to measure particle tracks from both primary and secondary vertices with excellent p_T resolution within a pseudorapidity range of $|\eta| < 2.5$. Fig. 2.2 shows a cut-away diagram of the ID with labeled sub-detectors and components. The ID has a cylindrical structure of length 3.51 m and a radius of 1.15 m and is composed of three independent sub-detectors. From the beamline outwards, these detectors are: a high-resolution silicon pixel detector with 3 layers, a silicon microstrip semiconductor tracker (SCT) detector with 4 double layers and a transition radiation tracker (TRT) composed of many layers of straw tubes filled with a Xe-based gas mixture. These three sub-detectors are placed in a central solenoid, which extends over a length of 5.3 m with a radius of 1.25 m and generates a 2 T magnetic field. For a more detailed discussion of the ATLAS ID see [40] [41].

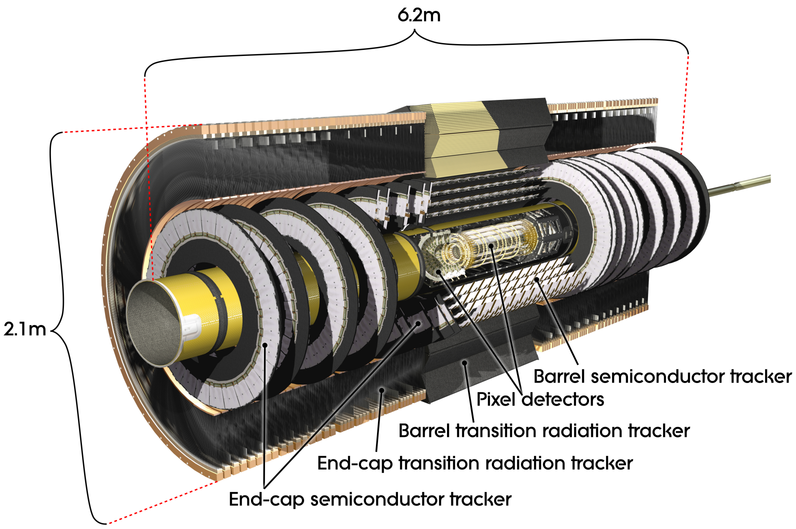


Figure 2.2: Cut-away diagram of the ATLAS Inner Detector with labeled sub-detectors and components. Image taken from [38]

Pixel detector

The Pixel Detector is the innermost sub-detector of the ID. It is composed of 3 "barrel" layers that wrap around the beampipe in concentric cylinders situated with radii, R, of 50.5 mm, 88.5 mm and 122.5 mm from the normal beam position and 3 layers of "end-caps" covering the ends of the barrels situated 494 mm, 580 mm and 650 mm from the collision point $z = 0$. The barrel and end-cap layers are covered by 31 million and 2.2 million identical silicon pixels respectively. Each pixel sensor has an individual readout channel—approximately 80.1 million in total. The pixel layers can be segmented into $R\phi$ and z units, where all pixels are identical and have a size in $R\phi \times z$ of 50 $\times$400 μm^2. The intrinsic accuracies for each pixel sensor are 10 μm ($R\phi$) and 115 μm (z) in the barrel and 10 μm ($R\phi$) and 115 μm (R) in the end-caps. A typical track will transverse all three of these layers, leaving a hit in each. The pixel detector provides for the highest granularity around the vertex region to give the most precise measurements of the tracks and vertex positions possible. For a more detailed description of the pixel detector see [42].

SCT

The Semiconductor Tracker is similar to the pixel detector in that they are both made from similar silicon sensors. The SCT is made from a pair of single-sided silicon micro-strip sensor connected end-to-end. Two such pairs glued back-to-back form modules 126 mm long. The total number of modules in the SCT is 4088 with approximately 6.3 million readout channels. The two layers of silicon strips are designed to be slightly off parallel so that the z-coordinate of a particle transversing both layers can be measured by the slight difference in its R measurement. The intrinsic accuracies of the strips per module are 17 μm ($R\phi$) and 580 μm (z) in the barrel and 17 μm ($R\phi$) and 580 μm (R) in the end-caps. Like the pixel detector the SCT is wrapped around the beampipe in 4 concentric cylinders with 9 end-cap disks at each end. The barrel layers are situated at R-coordinates 284 mm, 355 mm, 427 mm and 498 mm, while the 9 end-cap disks have a $|z|$ position of 854-2720 mm. For a more detailed description of the SCT see [43, 44].

TRT

The Transition Radiation Tracker is the largest of the ID sub-detectors, mounted around the pixel and SCT detectors. The main purpose of the TRT is to discriminate electrons from other charged particles through their unique transition radiation signature given off as they pass through the TRT. The TRT is built from straw tubes of length 144 cm and of diameter 4 mm that run parallel to the beampipe in the barrel region. End-caps are composed of 37 cm long straws that extend radially outward in a disk at the end of the TRT barrel. There are 73 straw planes in the barrel that cover a radial region of $554 < R < 1082$ mm and 160 in the end-caps covering a radial region of $617 < R < 1106$ mm. The TRT can only provide $R\phi$ measurements with an intrinsic accuracy of 130 μm per straw. Typically a particle track transversing the TRT will leave 36 hits. For a more detailed description of the TRT see [45].

2.2.4 Calorimetry

High-energy electrons and photons form a cascade of particles when incident upon dense materials. This cascade, referred to as a *shower*, is the result of pair production of electron-positron pairs from photons as well as bremsstrahlung from charged particles. Such electromagnetic showers are characterized by their radiation length, X_0, and transverse profiles. An analogous type of showering occurs when high-energy hadrons, such as nucleons, pions and kaons, are incident upon dense material. The mechanism involved in *hadronic* showers is partially electromagnetic, since the particles are often charged, but also involves interactions with nuclei via the strong force, where inelastic hadron-nuclear interactions produce particle multiplication. The ATLAS calorimeter exploits particle showers by stopping these particles in a dense material sampling the energy of the shower and hence the incident particle. Additionally, by studying the penetration depth and shower spread the type of incident particle can often be identified.

The ATLAS calorimeter is used to measure the energy of electrons, photons and jets as well as the total missing transverse energy, E_T^{miss}, in an event. It is also one of the central systems used for triggering. The calorimeter can be separated into three parts, the electromagnetic (EM) calorimeter, the hadronic calorimeter and the forward calorimeter, which together cover a range of $|\eta| < 4.9$. Over the pseudorapidity range of $|\eta| < 2.5$, which matches that of the ID, the EM calorimeter has a fine granularity for precision measurements of electrons and photons. The rest of the calorimeter has a courser granularity, but is sufficient for measuring jets and E_T^{miss}.

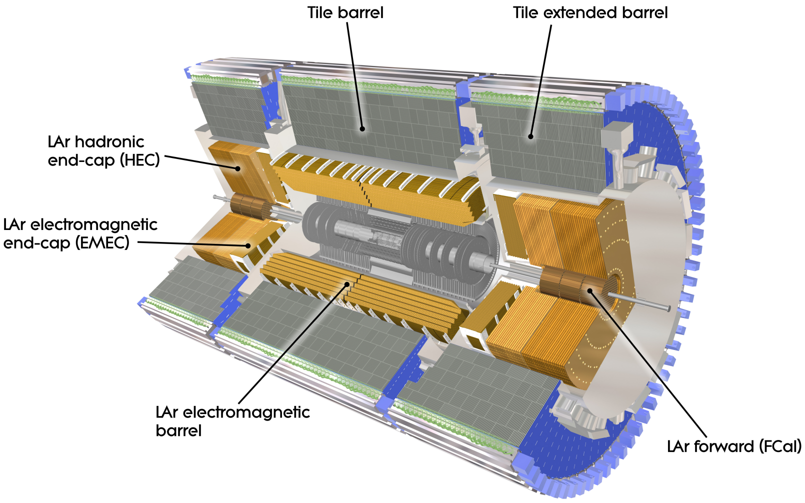


Figure 2.3: Cut-away view of the ATLAS calorimeter system with labeled subdetectors and components. Image taken from [38]

Electromagnetic Calorimetry

The EM liquid argon (LAr) calorimeter consists of 1.5 mm thick triangular-wave sheets of lead stacked upon one another immersed in a bath of liquid argon. When high-energy electrons and photons traverse these lead sheets they induce EM showers, converting their kinetic energies into lower energy shower electrons and photons. These shower electrons pass through the lead into 4 mm LAr filled gaps between the sheets. These electrons knock out argon atom electrons as they traverse the LAr, leaving a trail of electron-ion pairs in their wake. An electric field causes the displaced electrons to drift to readout electrode cells placed in the middle of the LAr gap. This motion of drifting electrons creates a current in an external circuit connected to the calorimeter. The number of produced shower electrons is proportional to the energy of the incident particle and thus the measured current from the calorimeter.

The EM calorimeter is divided into a barrel region, covering a pseudorapidity of $|\eta| < 1.475$, and two end-caps covering a pseudorapidity of $1.375 < |\eta| < 3.2$. The triangular-wave shape of the electrodes allows for continuous azimuthal coverage with minimal density variations. The total thickness of the EM calorimeter is greater than 22 X_0 in the barrel region and 24 X_0 in the end-caps, which provides good containment for high-energy jets. Electrons are identified by information from both the ID and shower shape. The high granularity of the EM calorimeter allows for reconstruction of the direction of the shower, allowing for discrimination of electrons and photons from secondary decays. A full description and technical specifications of the EM calorimeter can be found in [46].

Hadronic Calorimeter

The hadronic tile calorimeter surrounds the EM calorimeter, absorbing the energies of hadrons that escape it. The energy of particles such as protons, neutrons, pions and kaons is transformed into showers of hadrons when passing through steel absorbers that face normal to the beamline. These absorbers are separated by scintillating plastic tiles which emit light in an amount proportional to the incident particles. Wave-length shifting fibre readouts that line the edges of the tiles feed these light signals into photomultiplier tubes that convert them into a current, via the photoelectric effect, which is measured through an external circuit connected to the calorimeter.

At large pseudorapidities, radiation from the proton beam becomes increasingly intense. The scintillating tiles can be damaged by excessive radiation exposure. For this reason, the hadronic end-caps are made from a similar sampling material as the EM calorimeters. The main differences being that the lead plates are replaced by 2.5 cm thick copper plates and the argon gap region in between the plates is increased to 8 mm.

The barrel and extended barrel regions of the hadronic calorimeter cover a pseudorapidity of $|\eta| < 1.7$, while the hadronic end-caps cover the barrel ends, which extend over a pseudorapidity of $1.5 < |\eta| < 3.2$. A full description and technical specifications of the hadronic calorimeter can be found in [47].

Forward Calorimeter

The forward calorimeters are designed to absorb and measure the energies of particles with pseudorapidities in the range of $3.1 < |\eta| < 4.9$. The forward calorimeters have a LAr active material where the gap between absorption plates is reduced to < 2 mm. The forward calorimeter is split into three 45 cm thick modules: one electromagnetic module and two hadronic modules. For the electromagnetic modules copper is used as the absorbing element while tungsten is used in the two hadronic modules, which is better suited for shorter absorption lengths. A full description of the forward calorimeters can be found in [48].

2.2.5 Muon Spectrometer

Muons are unique in that with sufficient energy they will escape both the EM and hadronic calorimeters since they have relatively large masses (about ×200 the electron) and do not interact via the strong force. Since the power radiated by charged particles due to acceleration orthogonal (synchrotron radiation) and collinear (bremsstrahlung) to the particle's direction of motion scale as m^{-4} and m^{-6}, muons, unlike electrons, lose little energy through this mechanism. Instead high-momentum muons lose energy primarily through ionization. For this reason the muon spectrometer is made up of multiple gas filled chambers in which passing muons leave an ionization trail of electron-ion pairs. The free electrons drift to the closest anode under the influence of an applied electric field inducing a cascade of freed electrons. This flow of electrons creates a current in an external circuit connected to the muon spectrometer.

Like the other sub-detectors the muon spectrometer is divided up into a barrel region and end-cap regions. In the barrel region there are three layers of muon chambers consisting of monitor drift tube (MDT) chambers for precision tracking and resistive plate chambers (RPC) for triggering. The muon chamber layers are arranged in the form of a cylinder surrounding the calorimeter covering a pseudorapidity of $|\eta| < 1$. There are also three layers of muon chambers in the end-cap region, which are arranged vertically in disks. The end-caps are made up of MDT chambers as well as thin gap chambers (TGC) which are used for triggering. At large pseudorapidities, with considerably higher particle flux, cathode strip chambers (CSC) are used for tracking, which offer greater radiation tolerances. The combined components of the muon spectrometer cover a pseudorapidity of $|\eta| < 2.7$ with the exception of a 300 mm gap at $\eta = 0$, needed for the passage of services for interior detector systems.

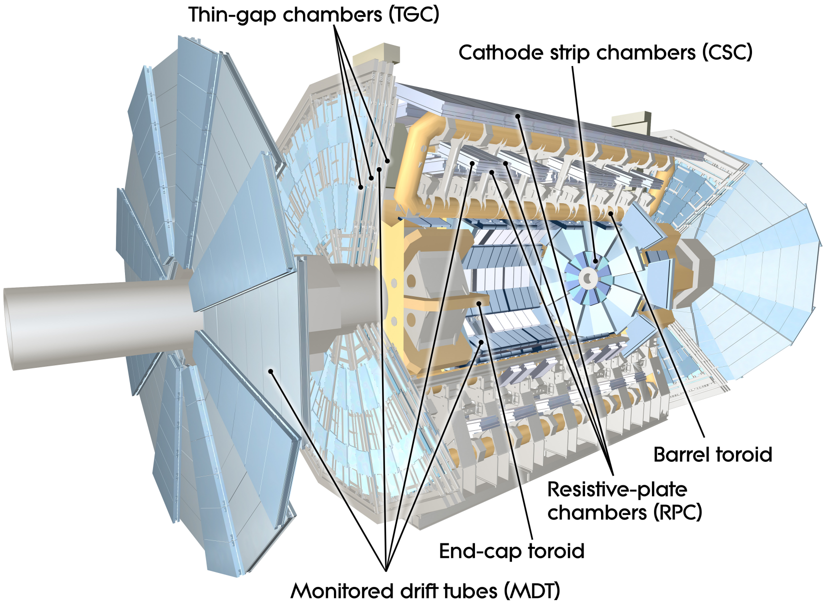


Figure 2.4: Cut-away view of the ATLAS muon spectrometer with labeled sub-detectors and components. Image taken from [38]

Chapter 3

Data collection

The ATLAS trigger and data acquisition system (DAQ) is based on three levels of event filters. The trigger applies selection criteria at each level with increasing refinement while the DAQ receives and buffers the event data. Figure 3.1 gives a conceptual overview of the trigger and DAQ system with labeled trigger levels and data recording rates.

The first trigger level (L1) uses custom-designed hardware that quickly accept or reject events based on low-level event characteristics from the calorimeter and muon sub-detectors. Calculations done at L1 use the raw data at low resolution.

After the initial trigger stage data is digitized and physics objects such as leptons and jets are identified in a process known as *event reconstruction.* During reconstruction tracks, pixel hits and calorimeter energy deposits are extrapolated, combined and identified as physical objects. This process serves two purposes: first to reduce the storage space needed to describe the event and second to allow for more sophisticated analysis of the event based on particles and energy flows. During reconstruction physics-level objects are calibrated based on their identified type. Section 3.2 discusses the reconstruction of electrons, muons, jets and missing transverse energy.

After the reconstruction stage the second level trigger (L2) and *event filter* (EF) stages are applied. These stages are used to calculate the same quantities as the L1 trigger but with higher resolution and calibrated detector responses. A brief outline of these systems, as well as L1, is given is section 3.1.

Section 3.3 describes the LHC operating conditions and ATLAS data collection periods for 2012. This section also gives a brief overview of the Monte Carlo generation, hadronization, showers and detector response simulations used.

3.1 Triggers

Interesting processes produced by the LHC will typically have small cross-sections making them rare to observe. For example, roughly only one in a trillion collision events will yield a Higgs boson. Nonetheless, the collision rate at the LHC is high enough to produce these types of events in statistically meaningful quantities. At full luminosity the LHC delivers a collision frequency of 40 MHz. In contrast, the data storage and offline computing facilities are only designed to handle an event frequency of roughly 200 Hz [49]. The trigger system is designed to bridge this gap. With costume-built hardware and speed-optimized algorithms the trigger quickly sorts through events keeping only those that are deemed to contain interesting physics, based on one of hundreds of predefined conditions called *triggers*.

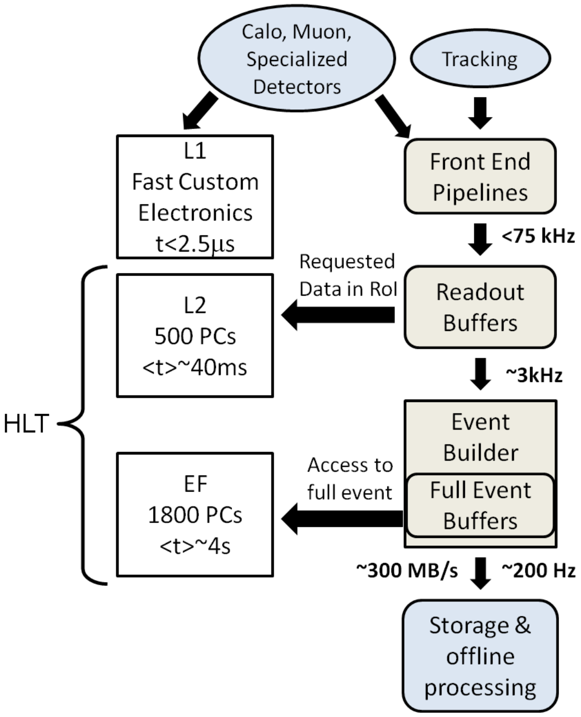


Figure 3.1: Block diagram of the ATLAS trigger and DAQ system. Taken from [49].

3.1.1 Level 1 trigger

The L1 trigger reduces the LHC bunch crossing rate from 40 MHz (at design luminosity) to 75 kHz by searching for high p_T leptons, photons, jets as well as large missing and total transverse energy. The L1 trigger uses information at reduced granularity from the RPC and TGC to identify high-p_T muons while electrons, photons, jets, τ-leptons, missing and total transverse energy are identified using the calorimeter. The selection criteria in the L1 trigger is often based on a p_T threshold, e.g. muon $p_T > 10$ GeV, but can also require any combination of physics objects.

The L1 calorimeter trigger decides to accept or reject events based on dedicated analogue signals from the ATLAS electromagnetic and hadronic calorimeters. The L1 decision is based on analogue sums of calorimeter elements within special 3-dimensional regions called trigger *towers*. Trigger towers have a size of roughly $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ per layer in the central part of the calorimeter ($|\eta| < 2.5$), and are less regularly sized in the forward regions. There are 7168 trigger towers in both the electromagnetic and hadronic calorimeters. Signals from the towers are digitized and matched up with their respective LHC bunch crossings. Then two separate processor systems work in parallel to form the trigger decision. One system, the cluster processor, uses all of the trigger towers in the central region to look for small localized clusters typical of electrons, photons or taus. The jet and energy-sum processor, uses 2×2 sums of the trigger towers, called jet elements, to identify jet candidates and form the missing energy and total energy sums. The magnitude of the trigger tower sums are then compared to the predefined trigger thresholds to form the trigger decision. The details of the algorithm can be found in Ref. [50].

Trigger towers that pass the threshold are identified as *Regions of Interest* (RoI) and are passed down the data pipeline for further analysis. The RoI data includes information on the types of features identified and the criteria passed. If the event is selected by the L1 trigger the RoI data is read out from front-end electronics in readout drivers (RODs) and then into readout buffers (ROBs) that feed event data to the L2 trigger.

3.1.2 High level triggers

The *high level trigger* (HLT) is composed of both the L2 trigger and EF. The HLT benefits from the additional information of inner detector hits, calorimeter readings at full granularity and high precision measurements from the muon detectors. The HLT trigger decision is based on reconstructed and calibrated physics objects (see section 3.2 for detail on reconstruction). The detailed calorimeter cell information available at the HLT allows for energy clusters to be calculated at higher precision than at L1. Additionally, shower shape variables are calculated which are used in particle identification.

The L2 calorimeter clustering algorithm is seeded by the RoIs from L1. The L2 trigger checks the L1 trigger decision with higher resolution and applies tighter selection criteria on the event data—reducing the trigger rate to approximately 3.5 kHz. Electrons and photons are identified by the depth and shapes of their energy deposits. The cluster position and transverse energy is calculated by taking an energy-weighted average of cell positions with a $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ window.

When an event passes the L2 trigger all of its event data is transferred by the DAQ to the EF. This last stage in the trigger systems reduces the event rate to roughly 200 Hz, as required by the offline computing environment. Starting in 2011 the EF jet selection is based on the offline anti-kt jet algorithm described in section 3.2. The E_T^{miss} is recalculated at the EF where data at full granularity from the whole event is available. Reconstructed objects at the EF use similar reconstruction and calibration schemes to the offline environment. Events selected by the EF are moved to permanent storage at the CERN computer centre where a typical event will occupy 1.3 MB of data. More information on the ATLAS trigger and DAQ systems can be found in Ref. [51].

3.2 Event reconstruction

In this section a brief outline is given on how physics objects, such as muons, jets and missing transverse energy are reconstructed. For a more detailed description of object reconstruction see Ref. [36].

3.2.1 Electron reconstruction

Electrons are reconstructed by identifying energy deposits in the EM calorimeter and associating them with reconstructed tracks in the inner detector. EM clusters are seeded from energy deposits with a transverse energy greater that 2.5 GeV using a sliding-window algorithm. Then tracks are associated with the cluster by extrapolating $p_T > 0.5$ GeV tracks from the inner detector to the EM calorimeter. Tracks and clusters are matched if the distance between the track impact point and cluster centre satisfies $|\Delta\eta| < 0.05$ and $\Delta\phi < 0.1$. Candidate electrons are considered reconstructed if at least one track is matched to its seed cluster. In the case where more than one track is matched tracks with hits in the pixel detector or the SCT are given priority, otherwise the track with the smallest ΔR is matched. If no tracks are matched then the cluster is reconstructed as a unconverted photon candidate. Electrons are distinguished from converted photons by the presence of one or two tracks pointing to a displaced vertex [52].

The energy of the reconstructed electron is corrected for deposits in the material in front of the EM calorimeter; energy deposited outside the cluster; and the estimated energy deposited beyond the EM calorimeter [53].

3.2.2 Muon reconstruction

Muons are reconstructed from a combination of ID and muon spectrometer tracks. The ID provides good measurements of muon tracks at low and intermediate momenta, while the muon spectrometer gives more accurate measurements at momenta over 30 GeV [36]. The central solenoid and toroidal magnets bend the trajectories of transversing muons to allow for momentum measurements. Muon tracks are found by combining track segments, which are defined as straight lines in a single MDT or CSC station. Track candidates are built from segments found in the outer and middle stations of the muon spectrometer and extrapolated back through the magnetic field. When segments are found within the proximity of the track extrapolation they are added to the track candidate. The final track fitting algorithm takes into account the geometry and composition of the material traversed as well as magnetic field inhomogeneities. Using parameters found in the inner stations the track candidates found in the muon spectrometer are propagated back to the interaction point, correcting for energy loss in the calorimeter. ID tracks are matched with these propagated tracks to form the combined tracks from which muon momenta are determined.

3.2.3 Topological cell clusters

Topological cell clusters attempt to reconstruct particle final states based on their 3-dimensional energy deposits in the calorimeter [36]. The cell clustering algorithm is a successive recombination algorithm, where cells that have an energy over a specific threshold are combined into a *topocluster* through an iterative process. Cells with an absolute energy threshold above four standard deviations of the total noise, $|E_{cell}| > 4\sigma$, (electronics and pile-up) act as seeds for the clustering procedure. Cells neighbouring these seeds are collected together with the seed cell into a cluster. Then if any of the neighbouring cells have an absolute energy above two standard deviations, $|E_{cell}| > 2\sigma$, they are taken as secondary seeds and added into the cluster. This process continues until all cluster perimeter cells have a $|E_{cell}| \leq 2\sigma$, at which point the process is halted with the perimeter cells added to the topocluster. Cell energies are added together in a weighted sum, where the weights are found by calibration on the electromagnetic scale. This weighted sum defines the final topocluster energy.

3.2.4 Jet reconstruction

As discussed in section 2.1.2 partons are not observed directly due to colour confinement, instead collimated sprays of hadrons called "jets" are observed. Jets are the result of the hadronization and fragmentation process from the hard scattered partons. To make accurate predictions from the parton-level to the hadron-level a well defined jet-finding procedure is essential.

The anti-k_t jet finding algorithm is a common jet finding procedure and is employed in the following analysis to reconstruct jets from topoclusters. The algorithm in its simplest form (i.e. not optimized for computation time) can be defined as in Ref. [54]:

1. The "k_T distance", $d_{ij} = \min(1/p_{Ti}^2, 1/p_{Tj}^2)\frac{\Delta R_{ij}^2}{R}$, between clusters i and j is calculated along with the distance between cluster i and the beam axis $d_{iB} = 1/p_{Ti}^2$. Here R is a parameter of the anti-k_t algorithm that sets the size of the search cone.

2. The minimum of all d_{ij} and d_{iB} is found. If d_{min} is a d_{ij} then clusters i and j are merged, summing their four-momentum. If d_{iB} is the minimum distance then the ith cluster is taken to be the reconstructed jet.

3. Steps 1 and 2 are repeated until all topoclusters are exhausted.

The ATLAS experiment has adopted the `FastJet` [55] algorithm, which exploits geometric relations between clusters to reduce computation time.

3.2.5 E_T^{miss} reconstruction

The E_T^{miss} is reconstructed primarily through calorimeter topocluster energy deposits and an additional muon term measured from the muon spectrometer and ID, which is not always included. The E_T^{miss} of an event is reconstructed with calorimeter topoclusters calibrated according to the reconstructed physics object with which they are associated [56]. Cells and topoclusters are associated with medium to high p_T objects in the order: electrons, photons, hadronic taus, jets and muons. There is also a non-negligible contribution from cells not associated with a reconstructed object referred to as the "cell out" term. After each term has been calibrated according to its associated object their components are summed:

$$E_{x,y}^{miss} = E_{x,y}^{miss,e} + E_{x,y}^{miss,\gamma} + E_{x,y}^{miss,\tau} + E_{x,y}^{miss,jets} + E_{x,y}^{miss,\mu} + E_{x,y}^{miss,CellOut}. \quad (3.1)$$

where $E_x^{miss,term} = -\sum E_i \sin\theta_i \cos\phi_i$ and $E_y^{miss,term} = -\sum E_i \sin\theta_i \sin\phi_i$ with the sum running over all associated topoclusters, or in the case of the muon term associated muon $p_{x,y}$. With these component terms the missing transverse energy in an event is given by the magnitude of the missing energy vector in the transverse plane

$$E_T^{miss} = |E_T| = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}. \quad (3.2)$$

3.3 Data samples

This section describes the data and Monte Carlo simulation samples currently being used in the monojet analysis. All datasets used are in D3PD format. D3PDs are structured n-tuples optimized for `ROOT` [57] which contain reconstructed physics objects.

3.3.1 2012 data periods

Runs between April 4th and December 6th, 2012 yielded a recorded integrated luminosity of 21.3 fb^{-1} with an impressive 95.5% detector uptime [58]. Figure 3.2 shows the cumulative luminosity versus time (green), luminosity recorded by ATLAS (yellow), and certified to be good quality data (blue) for *pp* collisions at 8 TeV.

The recorded luminosity includes data acquisition inefficiency, as well as the inefficiency of the "warm start" when the tracking detectors undergo a ramp of high-voltage and the pixel detectors turn on their preamplifiers. The "Good for physics" luminosity corresponds to all data that have passed minimum quality criteria, which removes luminosity blocks where certain sub-detectors may not be operating normally. All of the data that passes is added to a *good run list* (GRL). GRLs are typically used to

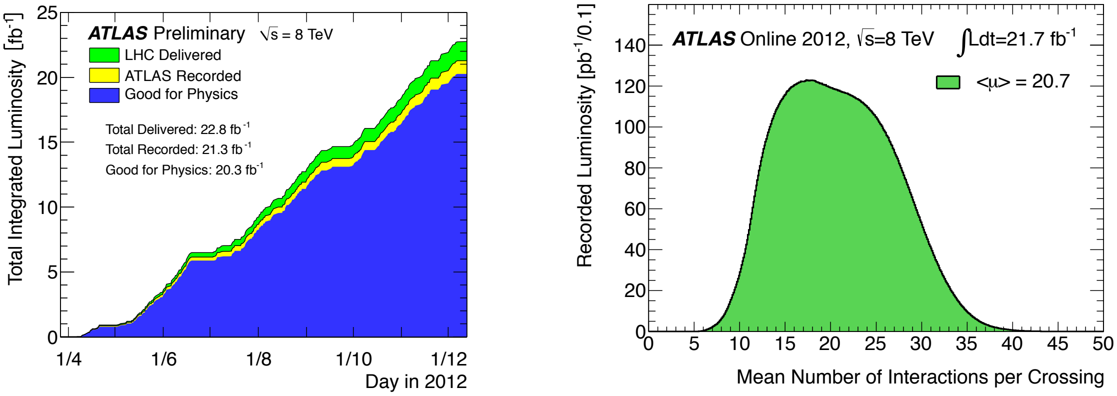


Figure 3.2: Integrated luminosity (left) and average pileup (right) during 2012 data collection period. Plots taken from [59].

filter data before an analysis is performed.

Also shown in figure 3.2 is the mean number of interactions, μ, per bunch crossing. For the 2012 data taking period the mean μ value, $< \mu >$, was found to be 20.7 [59]. Global quantities, such as E_T^{miss}, are affected by the number of interactions in the event. To reproduce this effect in Monte Carlo μ distributions must be weighted to reflect data.

3.3.2 Monte Carlo samples

Background processes with their respective detector signatures are simulated using Monte Carlo (MC) methods that model effectively all particle interactions in an event. Typically these programs simulate [60]:

- initial-state composition and substructure,
- the hard scatter process, i.e. desired process that leads to final state signature,
- resonance decay,
- particle showers,
- accompanying semi-hard process, i.e. “underlying event,”
- hadronization and further decay,
- detector responses.

Typically MC simulation programs will specialize in one or more of these aspects of the full simulation. For example *event generators* specialize in simulating the hard scatter process. The most common general event generators used by ATLAS for *pp* collisions are `SHERPA` [61], `PYTHIA` [62] and `HERWIG` [63]. Other MC simulation programs may specialize in other tasks such as: `MadGraph` [64], a multi-purpose parton level generator; `JIMMY` [65], for multiple parton processes; `AcerMC` [66], for simulating LHC background processes and `MC@NLO` [67], which calculates next-to-leading order QCD matrix elements (using `HERWIG` for parton showers).

For simulated samples of W/Z production in association with jets `SHERPA1.4.1` with the `CT10` [68] PDF set are used. Top processes are simulated using the `MC@NLO` generator using the `CT10` PDF interfaced to `JIMMY4.31` or `HERWIG6.520` for $t\bar{t}$ and single-t in the s-channel and Wt production. Single-t t-channel processes are generated with `AcerMC3.8` interfaced to `PYTHIA8.1` using the `CTEQ6L1` [69] PDF. Diboson samples are produced using `HERWIG6.52` with the `CTEQ6L1` PDF set.

The Higgs boson signal samples are generated in the $gg \rightarrow h$, $VV \rightarrow hV$ and $q\bar{q} \rightarrow hV$ production channels. The Higgs boson is assumed to be produced as predicted by the Standard Model with a mass of 125 GeV, however it may decay into invisible particles at a much higher rate[1]. The $gg \rightarrow h$ and $VV \rightarrow hV$ Higgs signals are generated with `POWHEG` [70], which is interfaced with `PYTHIA8.1` for showering and hadronization. For $q\bar{q} \rightarrow hV$ `POWHEG` is interfaced to `HERWIG` and the Z/W bosons are forced to decay to a pair of quarks. The invisible decay of the Higgs boson is simulated by forcing it to decay to two Z bosons, which in turn decay in their invisible mode to neutrinos.

Pile-up effects from non-diffractive, single-diffracting and double-diffractive collisions for all samples are overlaid using `PYTHIA8.1` assuming a $\mu = 20$. Samples are processed either with a full ATLAS detector simulation based on the `GEANT4` program [71] or a fast simulation based on the response of the EM and hadronic calorimeters [72] and the trigger system. The simulated events are reconstructed and analyzed in an identical manner to data.

---

[1] The Higgs boson may decay invisibly through Z bosons: $h \rightarrow ZZ \rightarrow 4\nu$.

Chapter 4

Event selection

After the trigger and reconstruction phases, data is stored to disk and distributed across the LHC computing grid. ATLAS analyses of all types use the LHC computing grid to access the same pool of data.

However, even at this stage the amount of data is on the scale of tens of Peta Bytes (PB), which is unmanageable for the types of sophisticated analyses conducted at ATLAS. The data is filtered further through as series of analysis-specific *cuts*, known as *event selection*, typically reducing the dataset, by 3-5 orders of magnitude. Section 4.2 describes the selection process used for object identification, while section 4.3 describes the preselection and baseline selection.

In addition to reducing the size of the dataset event selection is used to select events with a specific detector signature and to reduce background. The monojet signature and Standard Model backgrounds are discussed in section 4.1.

Often in high energy physics (HEP) analyses it is useful to split the dataset up into a number of different *regions*. The *signal region* is the one of interest that may contain the signal, in this case, jets plus E_T^{miss}. In addition *control regions* may also be defined; control regions are typically used for validation and estimating backgrounds in the signal region. This analysis makes use of four control regions: $Z\mu\mu$, Zee, $W\mu\nu$ and $We\nu$, which are used to estimate Z/W + jet backgrounds in the signal region, defined as events with large momentum imbalance and hadronic activity. Section 4.4 discusses the control region definitions and figure 4.3 shows some select distributions in these control regions for comparison with data.

4.1 Signature and background

Physics analysis at ATLAS are typically defined by a specific detector *signature*. A detector signature is a unique set of physics objects that identify the events. Signatures are often associated with a given physical process of interest. For example a signature of one high-p_T muon, one high-p_T jet and missing transverse energy might be associated with the process W($\rightarrow \mu\nu$)+ jet. However, there are other Standard Model processes that may lead to the same signature, such as W($\rightarrow \mu\nu$) + W($\rightarrow jj$) where one of the jets falls outside of the detector's acceptance. These additional processes that end up contaminating the signature are designated *background* processes.

4.1.1 Monojet signature

As discussed in section 1.4, WIMP models require a very small interaction cross-section with ordinary matter. This means any WIMPs created at the LHC would pass through the ATLAS detector completely undetected. Particles that cannot be detected by ATLAS are referred to as "invisible". Neutrinos, for example, are invisible to the ATLAS detector; their presence must be inferred from a large momentum imbalance in the transverse plane (E_T^{miss}).

If WIMPs are produced through a dark sector mediator particle, such as the Higgs, they will be produced in pairs through the decay of the mediator. Pure pair production events will result in 'back-to-back" topologies in the transverse plane, where the momentum vectors of the two decay products will have equal magnitude in opposite directions. If the decay products are invisible the event will not pass the trigger since no signature will be detected. For such events to pass the E_T^{miss} trigger there needs to be a least one additional object associated with the hard scatter to give something for the mediator particle to recoil off of. In this analysis the presence of at least one jet is required. For such events the mediator particle, and hence the jet it recoils off of, must have a significant amount of transverse momentum in order to be observable. Otherwise the event may not pass the E_T^{miss} trigger, or may be buried under the large multijet background. The mechanisms by which such jets may be produced in association with a Higgs boson are discussed in section 7.1.

The monojet signature is defined by events with a single high-p_T jet dominating the total hadronic activity and not containing any identified leptons. Figure 4.1 shows an event display of a typical monojet event. The figure shows a single high-p_T jet with nothing to balance its momentum in the transverse plane. The left panel shows the

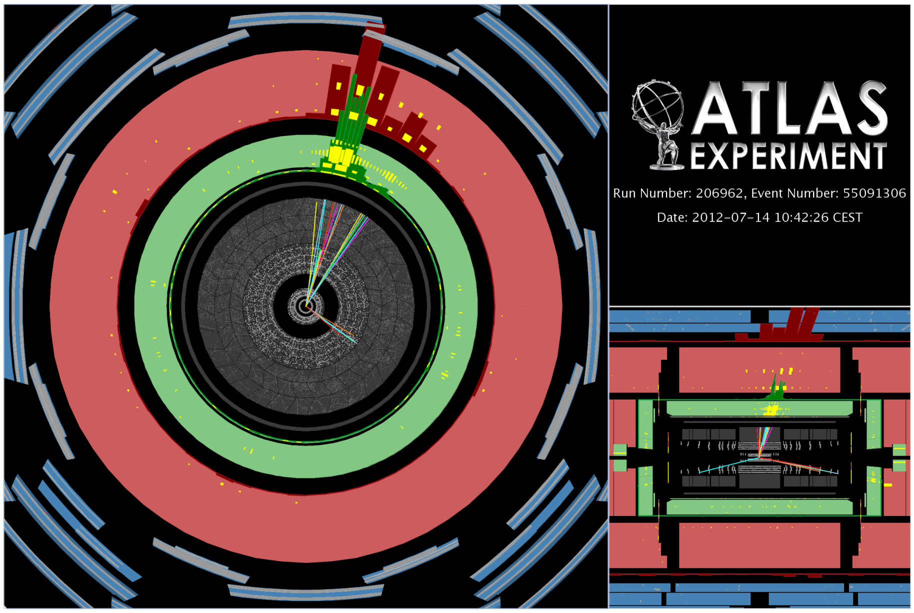


Figure 4.1: Event display of a typical monojet event. (Left) transverse plane; (bottom right) side-view of the detector.

event in the transverse plane, displaying calorimeter energy deposits as histograms and reconstructed tracks as line segments. The bottom right panel shows the same event from the side of the detector, i.e. parallel with the beamline.

4.1.2 Backgrounds

Since, like WIMPs, Standard Model neutrinos are invisible to the ATLAS detector they present a difficult background for the monojet signature. Final states of neutrinos and a single jet can occur in the Standard Model when a Z boson is produced in association with a jet (Fig. 4.2), where the Z boson decays in its invisible mode to neutrinos. Where most backgrounds can be reduced by using a series of selection cuts, the $Z(\rightarrow \nu\nu)$ + jets background is, to a large extent, irreducible[1]. Due to its irreducibility the $Z\nu\nu$ background tends to dominate, contributing more to the total background than all others combined.

Other backgrounds may occur when one or more of the particles and/or jets: falls outside of the main acceptance of the detector, is absorbed by inactive material, is misclassified, or has its energy miss-measured. The most common example of such an

---

[1] The extent to which this background can be reduced depends on the production mechanism of the jet(s) and the discriminating variables used in the analysis. See section 7.1.

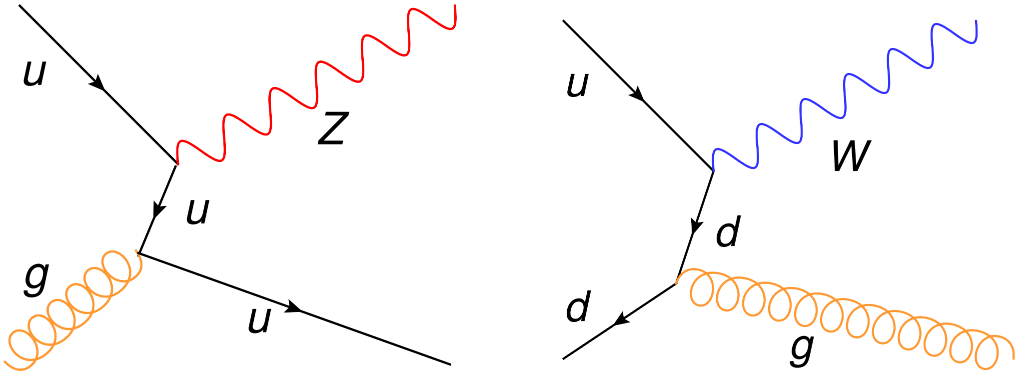


Figure 4.2: [Feynman diagrams of the mainain backgrounds to monojet signature. (left) Z boson produced in association with a light quark (u) jet. (right) W boson produced in association with a gluon jet.

event is a W boson produced in association with a jet (figure 4.2 right), where the W boson decays to a neutrino and a lepton and the lepton falls outside of the tracking detector's acceptance. The W background is the second largest background after $Z\nu\nu$. Other backgrounds can contribute through this mechanism such as $t\bar{t}$, single-t and diboson (WW, ZZ, WZ) events. In addition, multijet backgrounds can also arise from jet energy miss-measurements, causing a false missing energy signature aligned with the direction of the jet.

4.2 Object definitions

After the reconstruction stage identified objects are considered candidates for their respective physics object class (jet, muon, electron, etc). Physics object candidates are then subjected to further selection requirements that are tailored for the particular analysis being conducted. The definitions describe in this section define the final physics object that are used in selection and the classification model described in section 7.2.

4.2.1 Jets

Jet candidates are reconstructed with the anti-k_t clustering algorithm presented in section 3.2.4 with a distance parameter $R = 0.4$. Topoclusters are corrected under the locally calibrated (LC) scheme to the hadronic scale [73] and used as inputs. Once the jets are reconstructed further corrections are applied for pile-up conditions and jet energy scale (JES)[2]. For the events that pass preselection the jet four-vectors are further corrected offline with the help of the `ApplyJetCalibration` package [74].

All cuts on the transverse momentum of jets are done after the calibration. For jet candidates to be selected as jets they are required to have a $p_T > 30$ GeV and fall within $|\eta| < 4.5$. Since the anti-k_t algorithm often reconstructs electrons as jets it is necessary to perform jet-electron overlap removal, where if a jet is within $\Delta R < 0.2$ of any identified electrons it is removed from the event.

4.2.2 Missing transverse energy (E_T^{miss})

While the trigger system definition of E_T^{miss} is purely calorimeter based, the offline definition is based on reconstructed physics objects including jets, electrons, muons, taus, photons and soft terms. This definition, referred to as `RefFinal`, allows for a more refined energy calibration and consistent estimation of systematics uncertainties. In this analysis it is convenient not to use reconstructed taus, so a variant of `RefFinal` called `Egamma10NoTau` is used instead, where the tau terms have been left out of the E_T^{miss} definition.

In the control regions the E_T^{miss} is recalculated using the `MissingEtUtility` package [75] where selected leptons are removed from the calculation by setting the corresponding map weights to zero (see section 4.4 for more details).

4.2.3 Muons

Muon candidates are identified by an inner detector track that is confirmed by a segment in the muon spectrometer. Muons are required to fulfill a certain number of pixel, SCT and TRT hits as recommended by the muon combined performance group [76]. Only isolated muons are considered; the scalar sum of transverse momenta of tracks within $\Delta R < 0.2$ around the muon candidate must be less that 1.8 GeV.

In the signal, Zee and $We\nu$ control regions, the event is rejected if it contains any muons with $p_T > 7$ GeV that fall within $|\eta| < 2.5$. In the muon control regions, $Z\mu\mu$ and $W\mu\nu$, events are accepted based on the same muon criteria except the p_T cut is raised to 20 GeV.

---

[2] This is commonly referred to as the LC+JES calibration scheme

4.2.4 Electrons

Different definitions for electrons are used in this analysis depending on the signal/control region. In the signal region, where electrons are vetoed, a relatively loose selection is applied resulting in a tight veto. For the electron control regions (Zee and $We\nu$) electrons are explicitly selected, requiring tighter selection cuts.

The ATLAS electron performance group defines three baseline operating points with regard to electron identification efficiency: *loose++, medium++* and *tight++* [77].

Events in the signal region are rejected if they contain any electrons with $p_T > 7$ GeV that are within $|\eta| < 2.47$. No isolation requirement is imposed.

In the Zee control region electrons are required to have medium++ quality, $p_T >$ 20 GeV and fulfill $|\eta| < 2.47$ while excluding the transition region $1.37 < |\eta| < 1.56$. Again no isolation requirement is imposed.

The requirements for the $We\nu$ control region are more stringent, as the electrons must pass tight++ quality and $p_T > 25$ GeV. In addition electrons are required to be isolated by imposing that the scalar sum of the transverse momenta of tracks with $\Delta R < 0.3$ around the electron candidate must be less than 5% of its transverse momentum.

4.3 Event selection

To remove as much background as possible a series of cuts are made on reconstructed physics objects. The trigger is always the first selection cut, which is done online during the data taking process. The trigger cut is usually fairly general and may be shared among many different analyses.

Following the trigger is preselection, a set of cuts typically designed to select events that pass certain data quality criteria. The purpose of preselection is to filter out poorly reconstructed events, or occasionally data that's been contaminated by faulty sub-detectors and components[3].

The cuts that define the main signal region make up the baseline selection. The baseline selection is used to reduce background in the signal region for a general monojet signature. Further optimization of the signal region is discussed in section 7.2 for invisible Higgs signal models.

---

[3] Severely affected events are filtered out before preselection with the GRLs (see section 3.3).

4.3.1 Preselection

Trigger: The `EF_xe80_tclcw` trigger has a 80 GeV E_T^{miss} threshold calculated from LC topoclusters at the EF.

Vertex: At least one vertex with two or more associated tracks must be present in the event. This cut is designed to establish that the interaction point is well reconstructed in the event.

Data cleaning Events that suffer from noise bursts, dead tiles, power trips to sub-detectors, or severe amounts of calorimeter noise during the data taking are removed.

Jet cleaning Events that contain at least one poorly reconstructed jet with a p_T >20 GeV are removed. This cut is designed to reduce fake E_T^{miss} as poorly reconstructed jets will often suffer from energy miss-measurement.

4.3.2 Baseline selection

Lepton veto: Events with any identified electrons or muons are vetoed. This cut is to reduce backgrounds from electroweak processes that contain a W or Z boson that decays leptonically. This includes the W $\rightarrow l\nu$, Z $\rightarrow ll$, $t\bar{t}$, single-t and Diboson backgrounds already mentioned.

Track veto: Events with one or more isolated tracks are vetoed. This cut has been shown to reduce backgrounds containing one-prong leptonic tau, hadronic taus or isolated leptons that are not vetoed [78].

Lead jet: Events are required to have a leading jet with a $p_T > 120$ GeV and an $|\eta| < 2$. This cut is designed to select events containing a high p_T jet that is reconstructed within the high central section of the calorimeter.

Multijet The absolute angle between any of the jets and the E_T^{miss} in the transverse plane is required to be greater than 0.5 rads: $|\Delta\phi_{0j}| > 0.5$. This cut reduces fake E_T^{miss} that results from multijet events where one of the jets suffer from an energy miss-measurement.

E_T^{miss}**:** $E_T^{miss} > 350$ GeV. This stringent E_T^{miss} cut is placed to reduce the multijet background to a negligibly low rate.

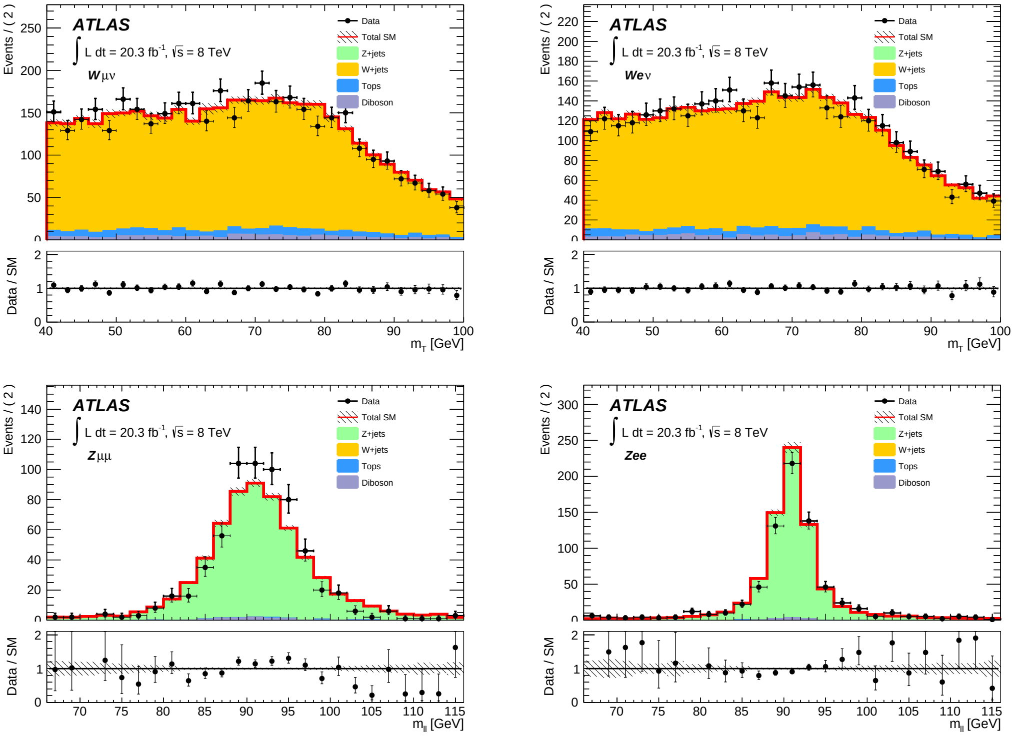


Figure 4.3: Histograms of m_T in the W$\mu\nu$ (top left) and W$e\nu$ (top right) control regions; m_{ll} in the Z$\mu\mu$ (bottom left) and Zee (bottom right) control regions.

Balance: The ratio of the E_T^{miss} to the jet-p_T scalar sum is required to be greater than 0.5. This cut is used to ensure the monojet topology.

4.4 Control regions

In addition to the signal region, control regions are defined, which are used for the purposes of estimating background (section 5.3) and validation of MC modelling.

There are four different control regions that can be used to estimate various backgrounds in the signal region: Z ($\rightarrow \mu\mu$) + jets, Z($\rightarrow ee$) + jets, W($\rightarrow \mu\nu$) + jets and W($\rightarrow e\nu$) + jets. In these control regions the identified leptons are removed from the event to mimic a monojet signature. Up to two identified leptons are removed from the reconstructed E_T^{miss} using the `MissingEtUtility` package to allow for the large missing energy spectrum required for the baseline selection[4]. To avoid confu-

---

[4] Conceptually this procedure is equivalent to the vector sum of E_T^{Ref} and the transverse momenta of the lepton(s). In practice, however, this does not properly take into account the `RefFinal`

sion, the missing transverse energy before the lepton removal shall be referred to as E_T^{Ref}. These leptons and tracks are also removed from the lepton and track veto cuts. To isolate each background requires some additional selection on top of the baseline selection, which define the control regions.

Figure 4.3 shows the transverse and invariant mass distributions for the four control regions after baseline and control region selection. Histograms such as these are used to validate MC modelling for different distributions. A full table of distributions for all control/signal regions can be found in Appendix C.

4.4.1 W$\mu\nu$ CR selection

Trigger: Same trigger as baseline selection, since `EF_xe80_tclcw` does not contain a muon term in the online E_T^{miss} definition.

Muon Exactly one good muon with a $p_T > 25$ and $|\eta| < 2.4$ must be present in the event.

E_T^{Ref} **cut:** Events are required to have a $E_T^{Ref} > 25$ GeV to select for the neutrino in the W boson decay.

m_T **cut:** Events are required to have a transverse mass, $m_T = \sqrt{2E_T^{Ref} p_T^{\mu}(1 - \cos\phi)}$, between 40 and 100 GeV. The distribution of m_T drops sharply, or has a “kinematic edge”, around the invariant mass of the W boson at 80 GeV.

4.4.2 W$e\nu$ CR selection

Trigger: Events must pass the `EF_e24vhi_medium1` or `EF_e60_medium` electron trigger. Using both triggers allows for greater efficiency.

Electron Exactly one good electron (tight++) with a $p_T > 25$ must be present in the event.

E_T^{Ref} **cut:** Events are required to have a $E_T^{Ref} > 25$.

m_T **cut:** Events are required to have a transverse mass, $m_T = \sqrt{2E_T^{Ref} p_T^{e}(1 - \cos\phi)}$, between 40 and 100 GeV.

---

calibration scheme.

4.4.3 Z$\mu\mu$ CR selection

Trigger: Same trigger as baseline selection, since `EF_xe80_tclcw` does not contain a muon terms in the online E_T^{miss} definition.

Muons: Exactly two good muons with a $p_T > 20$ must be present in the event.

Mass cut: Muons are required to have a invariant mass of $66 < m_{\mu\mu} \leq 116$ GeV to select for Z bosons with a mass of 90 GeV.

4.4.4 Zee CR selection

Trigger: Events must pass `EF_2e12Tvh_loose1`; a trigger for selecting dielectron events. Using this trigger instead of `EF_xe80_tclcw` allows for much greater efficiency.

Electrons: Exactly two good electrons (medium++) with $p_T > 25$ must be present in the event.

Mass cut: Electrons are required to have a invariant mass of $66 < m_{ee} \leq 116$ GeV.

Chapter 5

Statistical model

This chapter describes the statistical model used to estimate background contributions as well as effects of systematic uncertainties on event yields in the signal region. The chapter starts with section 5.1.1, which gives a brief theoretical overview of the mathematical tools and general likelihood model that have been developed at CERN and SLAC over the last few years—most notably from the works of Kyle Cranmer [79, 80] and Glen Cowan [81]. In section 5.1.2 a marked Poisson likelihood model is motivated by combining likelihood functions that arise from the overall event counts and discriminating variable distributions. Section 5.1.4 then expands this model to include constraint terms for nuisance parameters which are introduced as a way to parameterize systematic uncertainties. Then in section 5.1.5 the method of maximum likelihood for parameter estimation is introduced and motivated.

The second part of this chapter, section 5.2, gives a brief survey of the software implementation used to construct the likelihood model and perform the maximum likelihood fits. The software package `HistFitter` [82], which makes extensive use of the class structures provided by `RooStats` [79] and `RooFit` [83], is used to build the model and then the parameters are fit numerically with `MINUIT` [84].

The final part of this chapter starts off by motivating the data-driven background estimate procedure used to normalize the W/Z + jets processes—the dominant background in the signal region. Section 5.3.1 discusses how the control regions, as described in section 4.4, and associated normalization parameters enter the likelihood model. Then the background only fit procedure is presented in terms of the statistical formalism introduced earlier and propagation of parameter uncertainties is examined.

5.1 Likelihood model

The purpose of this section is to motivate and examine the statistical model that is used in this analysis. The model developed here is used in all aspects of the following analysis—from background estimates and systematic uncertainties to limit setting in both the model independent and dependant cases. The model is slowly derived by: introducing the concept of a likelihood function, modelling event rates as a Poisson process, parameterizing systematic uncertainties as nuisance parameters and adding constraint functions that summarize auxiliary measurements.

5.1.1 Likelihood functions

Likelihood functions are a useful tool in statistics as they allow one to model the parameters of a probability distribution given a fixed set of observations. The likelihood of a continuous probability distribution is given by

$$L(\boldsymbol{\eta}) = f(x|\boldsymbol{\eta})$$

where $\boldsymbol{\eta}$ are parameters that effect the probability distribution function (pdf)[1] $f(x|\boldsymbol{\eta})$ and x is the observed value (the data) for some chosen discriminating variable which shall be called X. It is important to note that likelihoods are not functions of the data, as a pdf is, but of the function parameters themselves; indeed, the likelihood is only calculable after the data has been collected.

Like probabilities, when multiple observations of x are made, individual likelihoods multiply

$$L(\boldsymbol{\eta}) = \prod_i f(x_i|\boldsymbol{\eta}) \tag{5.1}$$

where x_i is the ith observed value of the variable X.

One should note that likelihoods cannot be interpreted as probabilities, as they do not sum to one. Nonetheless, a likelihood can be interpreted as capturing the idea of something being likely to happen, or to have happened [85]. This property makes likelihoods a useful tool in parameter estimation, as the maximum value for a likelihood is achieved only for the most likely set of parameters for a given sequence of observations. This is referred to as the method of *Maximum Likelihood* (ML) in

---

[1] Here the lowercase pdf shall be used to avoid confusion with the Parton Distribution Function PDF.

statistical literature and is the basis for a large class of parameter estimation and machine learning techniques [86].

5.1.2 Marked Poisson model

The number of events observed that pass selection over a fixed time interval can be modelled as a Poisson process. Thus the probability of n collision events passing selection during a time interval of $[0, t]$ is given by the Poisson distribution

$$\mathrm{Pois}(n|\nu_{tot}) = \frac{\nu_{tot}^{n} \mathrm{e}^{-\nu_{tot}}}{n!} \tag{5.2}$$

where

$$\nu_{tot} = \epsilon \int_0^t R \mathrm{d}t = \epsilon \sigma_{pp} \int_0^t L \mathrm{d}t.$$

Here ν_{tot} is the total number of expected collision events, calculated from equation 2.4 and multiplied by a selection efficiency factor ϵ. Typically ν_{tot} is estimated using a combination of Monte Carlo simulations and control region measurements. It should be noted that n is a random variable, meaning if the experiment were conducted many times it would take on a distribution of values.[2]

The likelihood model can be built by holding n fixed and taking ν_{tot} as our variable

$$L(\nu_{tot}) = \mathrm{Pois}(n|\nu_{tot}). \tag{5.3}$$

For discrete distributions the likelihood is defined simply as a function of its parameters evaluated at the discrete value of the pdf variable. In this case it will be useful to define the observed event count as n_{obs}. Thus, the likelihood is evaluated at $n = n_{obs}$.

In a typical counting experiment at the LHC one is interested in estimating ν_{tot}, which is then compared with n_{obs}, where large deviations lead to discoveries while good agreement provides limits. In such experiments one is only interested in the total event yields—leaving out all of the information contained in the distributions of physical features, such as particle momentum, energy and angular distributions. The simple Poisson model is easily extended to take account of such information.

In any analysis the signal region is defined by a discriminating variable (or possibly a set of discriminating variables) such as X. Discriminating variables are chosen in

---

[2] This interpretation follows from the Frequentist school of thought, while Bayesian's interpret such variables to possess an inherent degree of uncertainty that leads to such distributions.

such a way as to enhance the differences between background and signal distributions. The likelihood of the discriminating variable can be added to the Poisson model by use of eq. 5.1;

$$L(\nu_{tot}) = \text{Pois}(n_{obs}|\nu_{tot}) \prod_{e}^{events} f(X = x_e), \tag{5.4}$$

where f is the pdf of X and the product is taken over all events e. $L(\nu_{tot})$ is an example of a mixture model; specifically, this is what is referred to by Statisticians as a *Marked Poisson Model* [80].

$f(X)$ is typically inferred from Monte Carlo simulation by nonparametric methods—usually histograms. In such cases the distribution of X is not continuous but discrete and one should interpret $f(x_e)$ as the area of the normalized histogram bin into which the observation x_e falls. This means that $f(x_e)$ is dependent on the background estimate of that bin, $\nu_b(x_e)$, as well as the bin size Δ;

$$f(x_e) \to \frac{\nu_b}{\nu_{tot}\Delta}.$$

In this case the bin size parameter Δ is a *hyperparameter* of the model—meaning that it is chosen before the fit, often through heuristic techniques. The product is then taken over all bins rather than the events

$$\prod_{e}^{events} f(x_e) \to \prod_{b}^{bins} \frac{\nu_b}{\nu_{tot}\Delta}.$$

In this analysis `HistFactory` is used to generate the pdfs over X—which uses discontinuous splines that look more like the unbinned expression [87]; however, all distributions will be presented as histograms.

5.1.3 Normalization factors

Often the overall rate estimates produced by Monte Carlo simulations is poor. If left uncorrected this may lead to an unreasonably large uncertainty in the total event count. The mismodeling of Monte Carlo rate estimates can be corrected by applying normalization factors to the effected Monte Carlo samples or processes. These normalization factors, μ, are incorporated into the likelihood model by attaching them to their respective event count estimates and distribution functions. In other words,

$$\nu_{tot} = \sum_{p}^{process} \mu_p \nu_p$$

and

$$f(X) = \frac{1}{\nu_{tot}} \sum_{p}^{process} \mu_p \nu_p f_p(X)$$

where p is the index of the background process (e.g. Z→ ee + jets) and μ_p is the normalization factor applied to that process. Likewise, f_p and ν_p correspond to the distribution of X and event count for process p. For processess that contribute relatively little to the signal region an event count uncertainty may be applied to that process instead of correcting it—in such cases $\mu_p = 1$. The values of μ_p are found using a set of control regions—at least one for each μ_p—as described in detail in section 5.3.1.

To accommodate these additional regions the likelihood model must be extended. Extending it to include an arbitrary number of regions is straight forward; it is simply the product of eq. 5.4 evaluated in each region

$$L(\boldsymbol{\eta}) = \prod_{r}^{region} \text{Pois}(n_r|\nu_r(\boldsymbol{\mu})) \prod_{e}^{events} f_r(x_{re}|\boldsymbol{\mu}) \tag{5.5}$$

where r labels the regions and $\boldsymbol{\mu} = \{\mu_p\}$ is the set of all normalization factors. To avoid cluttering the notation the subscript on n_{obs} is dropped, but it should be understood that n_r refers to the observed number of events and is not a model parameter. By including all of the control regions in the likelihood together the normalizations factors can be fit simultaneously. This allows all of the correlations that arise from background processes that enter multiple control regions to be properly modelled [82].

The control regions also play an important role in fitting the parameters that represent systematic uncertainties, as discussed in the next section.

5.1.4 Parameterization of systematic uncertainties

So far we have seen the parameters $\boldsymbol{\mu}$ enter the likelihood model as a consequence of possible mismodeling of the background distributions. Generally though, one is not interested in any of the values of $\boldsymbol{\mu}$—it is only ν_{tot} that is of value; however they are still inevitably part of our model and must be dealt with. For this reason such parameters are often called “nuisance parameters”. Systematic uncertainties which represent unknown properties of the detector’s response, such as the jet energy scale or electron identification efficiency, or theoretical uncertainty associated with a given model can also be parameterized as nuisance parameters.

Let the nuisance parameters of the model be defined as follows:

- $\boldsymbol{\mu}$ is the set of all normalization parameters,
- $\boldsymbol{\alpha}$ is the set of all parameters associated with systematic uncertainties,
- $\boldsymbol{\eta}$ is the total set of all nuisance parameters that enter the model, $\boldsymbol{\eta} = \{\boldsymbol{\mu}, \boldsymbol{\alpha}\}$.

Ideally we would have a full measurement for each α, which could correspond to an additional control regions entering the likelihood model in the typical way. However, there are often many unknown parameters, each requiring an independent study that could be just as complex as the one at hand. In practice the way this is handled within the ATLAS Collaboration is to have independent teams, or “performance groups”, that are in charge of making these types of auxiliary measurements in a way such that they are applicable to many different types of physics analyses. The performance group will typically supply a software package capable of calculating the nominal values of the parameter as well as ‘up’ and ‘down’ variations roughly corresponding to one standard deviation in each in direction.

In the case where $\boldsymbol{\alpha}$ represent parameters of the physical theory, such as corrections to the leading order term in a perturbative expansion of the scattering matrix or radiative corrections to charged particles arising from higher order loop diagrams, there are no possible auxiliary measurements that can be made. In such cases theorists usually provide an estimates of the nominal value, along with an estimate of its uncertainty which are taken as the up/down variations.

In either case, any deviation from nominal values will result in a change in the total expected event count ν_{tot}. Thus one must consider ν_{tot} as well as $f(X)$ as function of $\boldsymbol{\alpha}$ as well as $\boldsymbol{\mu}$:

$$\nu_{tot} = \nu_{tot}(\boldsymbol{\alpha}, \boldsymbol{\mu}) = \nu_{tot}(\boldsymbol{\eta})$$

and

$$f(X) = f(X|\boldsymbol{\alpha}, \boldsymbol{\mu}) = f(X|\boldsymbol{\eta}).$$

Using the up/down variations supplied by the performance groups one can see how much each systematic uncertainty independently effects the expected event count, where each deviation will contribute to the total uncertainty. Such a break down of the nuisance parameters is show in table 6.3. However, to find their best fit values one needs to treat $\boldsymbol{\alpha}$ as continuous parameters. The `HistFitter` package does this through a variety of interpolation methods—the exact implementations of which are thoroughly discussed in the `HistFitter` manual [82].

When incorporating systematic uncertainties into the likelihood model it is important that the possible values of $\boldsymbol{\alpha}$ in the ML fit are constrained to lie between their respective up/down variations, or very close to them. This can be achieved by including additional pdfs that correspond to the probabilities of the $\boldsymbol{\alpha}$ falling on their nominal values $\boldsymbol{\alpha}_0$.[3] Since these auxiliary measurements usually correspond to mean values they tend to a Gaussian distribution with large sample sizes thanks to the Central Limit Theorem. Here it will be convenient to standardize $\boldsymbol{\alpha}$ by translating to zero mean and scaling to unit variance

$$\boldsymbol{\alpha} \to \frac{\boldsymbol{\alpha} - \boldsymbol{\alpha}_0}{\boldsymbol{\alpha}_\pm} \tag{5.6}$$

where $\boldsymbol{\alpha}_\pm$ is the provided up/down variation. The up/down variations are not always symmetric, and in such cases they are symmetrized by taking the average variation. Now the constraint function for α can be written as

$$\mathrm{Gauss}(\alpha) = \frac{1}{\sqrt{2\pi}} \mathrm{exp}\left(-\frac{1}{2}\alpha^2\right) \tag{5.7}$$

where α is given by its standardized value.

Constraining parameters in this way is sometimes referred to a providing a ‘soft’ range for the parameters, i.e. they mostly lie between their up/down variations. However, `HistFitter` also provides the functionality to define a ‘hard’ range or limit for fit parameters where the fit values are not allowed to fall outside of the given range—such is equivalent to a truncated constraint function. This feature is used to avoid the possibility of the fit parameters taking on unphysical values, such as negative energy.

The normalization parameters $\boldsymbol{\mu}$ are also constrained, but since they are being

---

[3] This interpretation of the constraint functions is manifestly Bayesian, but should be tolerated for pedagogical reasons.

measured explicitly their constraints need not be so constrictive. Instead a uniform distribution is typically used;

$$\mathrm{Unif}(\mu) = \begin{cases} \frac{1}{b-a} & : a \leq \mu \leq b \\ 0 & : \text{otherwise} \end{cases}$$

where a and b are hyperparameters that represent the beginning and end points of some chosen range.

Incorporating these constraints into eq. 5.5 yields:

$$L(\boldsymbol{\eta}) = \underbrace{\prod_{r}^{\mathit{regions}} \mathrm{Pois}(n_r|\nu_r(\boldsymbol{\eta}))}_{\text{event rate}} \underbrace{\prod_{e}^{\mathit{events}} f_r(x_{re}|\boldsymbol{\eta})}_{\text{shape}} \times \underbrace{\prod_{s}^{\mathit{systematics}} \mathrm{Gauss}(\alpha_s)}_{\text{systematics}} \underbrace{\prod_{p}^{\mathit{process}} \mathrm{Unif}(\mu_p)}_{\text{normalizations}} \tag{5.8}$$

where s are the indices of the $\boldsymbol{\alpha}$ parameters relating to systematic uncertainties. The likelihood breaks up nicely into four distinct functions describing the event counts, distribution shapes, systematics constraints and normalization factors.

5.1.5 Maximum likelihood fitting procedure

The set of parameter values that maximize eq. 5.8 are referred to as the *maximum likelihood estimates* (MLEs) and are labeled with a hat: $\hat{\boldsymbol{\eta}}$. In practice it is easier, both analytically and computationally, to maximize the logarithm of the likelihood $\log L(\boldsymbol{\eta})$ since taking the logarithm results in a sum rather than a product and then each term can be differentiated independently. This does not affect the results since logarithms are monotonic functions and hence have the same extrema. It is also a matter of convention to minimize the negative of the logarithm. Thus the MLEs are given by

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \left(-\log L(\boldsymbol{\eta})\right). \tag{5.9}$$

In physics maximizing the likelihood is often referred to as a *fit* where the MLEs are referred to as the *best fit values.* In practice the minimum is usually found through numerical methods using `MINUIT`, which implements several different minimization algorithms [84].

The errors on the MLEs can be estimated easily when the number of events is large. In such cases the distribution of the MLE parameters approaches that of a Gaussian known as the maximum likelihood estimate asymptotic distribution [80]. In which case the inverse of the covariance matrix, V^{-1}, of the MLEs can be approximated as the Hessian matrix

$$V^{-1}_{i,j}(\boldsymbol{\eta}) = -\frac{\partial^2 \log L(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j}. \tag{5.10}$$

Note that the diagonal terms of the Hessian correspond to the variance of that parameter: $V_{ii} = \mathrm{Var}(\eta_i)$. `HistFitter` uses this method to estimate errors and correlations between the MLE parameters via the `MINUIT` `HESSE` subroutine [82] [84].

5.2 Statistical model implementation

The construction, fitting and evaluation of the likelihood model developed in the previous section is implemented with the `HistFitter` software package. `HistFitter` sits on top of an impressive tech-stack of statistical tools that have been developed at CERN and SLAC over the years to aid in the implementation and execution of statistical models over large datasets. This section provides a brief overview of this technology and gives references for more detailed descriptions.

`ROOT` A C++ library developed for high energy physics (HEP) data analysis, originally written by Rene Burn and Fons Rademakers in 1994 [57]. `ROOT` contains special class structures for HEP data analysis such as histograms, graphs, parametric functions, matrix algebra and four-vectors. A C++ interpreter called `CINT` is often used with `ROOT` to run script and macros. `ROOT` was designed with data intensive high performance computing in mind; it supports parallel processing for batch systems as well as multi-threading. `ROOT` also provides an efficient tree structure for persistent data storage which has become the default data format for the LHC experiments at CERN [88]. Over the years since its initial adoption `ROOT` has been extended by many libraries with class structures that inherit from `ROOT` primitives and objects. Many of these libraries have been incorporated into `ROOT` such as: libraries for detector simulation, multivariate analysis, graphical user interfaces and statistical modelling.

`MINUIT` A numerical minimization software package originally written in FORTRAN by Fred James at CERN in 1970s [84]. The program searches for minima in a user-

	defined function with respect to one or more parameters using several different minimization methods. The original FORTRAN code was later ported to C++ and incorporated into the `ROOT` library. `MINUIT` has become the standard tool used in HEP for MLE and curve fitting in general.
`RooFit`	A C++ library built on top of `ROOT` developed for the BaBar collaboration at Stanford Linear Accelerator Centre (SLAC) in 2003 [83]. `RooFit` introduced new data type primitives for building complex statistical models from data, such as real variables, multidimensional probability distribution functions, space points and real-valued integration. `RooFit` also supports useful statistical operations such as convolution of pdfs and likelihood profiling. `RooFit` now comes standard as part of the `ROOT` library.
`RooStats`	A C++ library expanding upon `RooFit` developed at CERN [79] around 2010. `RooStats` adds additional statistical subroutines for hypothesis testing, toy Monte Carlo methods and popular HEP test statistics used by the LHC experiments. `RooStats` also introduces RooWorkSpace, which is a persistent container for `RooFit` projects.
`HistFactory`	A C++ program designed to build template pdfs base upon `ROOT` histograms developed at CERN starting in 2012 [87]. `HistFactory` builds on top of the `RooFit` and `RooStat` statistical libraries. It provides binding for both C++ and Python with the capability to store configuration files in XML format. `HistFactory` is specifically designed to use the likelihood model described in section 5.1.1 as a basic template for general HEP data analysis.
`HistFitter`	A C++ program built on top of `ROOT`, `RooFit`, `RooStat` and `HistFactory` designed to add an additional layer of abstraction for data analysis [82]. `HistFitter` is currently still in development at CERN and SLAC but had its initial release in 2014. `HistFitter` has been quickly adopted by the ATLAS collaboration as it greatly simplifies the use of the above statistical packages and tools. `HistFitter` is implemented in the C++ programming language, however the user-defined configuration files are written in Python. `HistFitter` introduces the concept of control regions and signal regions as well as many different types of systematics uncertainties and provides command line macros for creating useful plot and tables.

5.3 Background fits

After the selection phase of the analysis the data has been partitioned into of several orthogonal regions, which are classified as either control regions or signal region(s). Control regions are used to make measurements which determine background rates or constrain systematic uncertainties. Then the signal region(s) are used to look for some excess of events over the expected background. The sensitivity to an excess in the signal region is highly dependent on the precision of the background estimate. This highlights the importance of the background fitting procedure outlined in this section.

Study of the control regions shows that the overall normalization of the W/Z + jets Monte Carlo simulation (Sherpa) over estimates data by 5-10%. This suggests that the Monte Carlo W/Z + jets background in the signal region suffer from a similar normalization offset. This discrepancy motivates the "data-driven" background estimate procedure outlined here.

Other backgrounds such as the Top, Diboson and Multijet backgrounds contribute on the order of 5%. Therefore these backgrounds do not need to be known to the same level of precision. They are modelled by Monte Carlo alone but are assigned a conservative uncertainty on their overall normalizations—as much as 50–100%.

5.3.1 Likelihood model for background only fit

Section 4.1.2 outlines the main background processes that contribute to the total background estimate in the signal region. The largest background arises from $\mathrm{Z}(\rightarrow \nu\nu) + \mathrm{jets}$, which contributes about 70% of the total background process in the signal region. Thus it is important to have an accurate estimate of it's overall contributions. However, it is not possible to directly measure $\mathrm{Z}(\rightarrow \nu\nu) + \mathrm{jets}$ as it gives exactly the same signature as the signal process. Hence there is no control region which can be defined to isolate it. Nevertheless, the related processes $\mathrm{Z}(\rightarrow \mu\mu) + \mathrm{jets}$ and $\mathrm{Z}(\rightarrow ee) + \mathrm{jets}$ can be easily measured since both electrons and muons can be identified with high efficiency. Theoretically the only differences in rates between these processes and $\mathrm{Z}(\rightarrow \nu\nu) + \mathrm{jets}$ is the branching ratios, which is usually absorbed into the cross sections. Experimentally, however, selection efficiencies and detector acceptances also need to be taken into account—typically estimated by MC. However trigger and lepton identification efficiencies are not well modelled and need to be corrected explicitly in the form of weights.

Mismodelling of the overall normalization of the event count predicted by a given MC simulation is fairly common. This mismodelling can derive from a number of different sources such as the event generator, momentum transfer, parton distribution functions, detector simulation, parton showers or hadronization choice. However, for two different Monte Carlo samples generated in the same way, i.e. same choice of generator, momentum transfer etc., one would expect the same bias in the event counts. Since all of the Z + jets MC samples were generated with Sherpa—with the same settings and parameters—any corrections applied to Z($\rightarrow \mu\mu$) + jets or Z($\rightarrow ee$) + jets can also be applied to Z($\rightarrow \nu\nu$) + jets. A similar argument can be made for the W + jet backgrounds that contaminate the signal region.

The strategy taken here is to treat all of the Z and W backgrounds the same; that is, to use two normalization parameters—one for all the Z and one for the W backgrounds. With this method one does not have to worry about propagating from the control regions to the signal region as it is the same set of parameters used in both.

Thus we can expand ν_{tot} into its individual components as

$$\nu_{tot} = \mu_Z \nu_Z + \mu_W \nu_W + \sum_p^{MC} \nu_p. \tag{5.11}$$

where

$$\nu_Z = \nu_{Zee} + \nu_{Z\mu\mu} + \nu_{Z\tau\tau} + \nu_{Z\nu\nu}$$

$$\nu_W = \nu_{We\nu} + \nu_{W\mu\nu} + \nu_{W\tau\nu}$$

and $\sum_p^{MC} \nu_p$ is simply the sum of the remaining (uncorrected) background contributions. A similar break-down can be performed on $f(X|\boldsymbol{\eta})$, which is composed of the same background processes.

The signal region is not used in the background only fit that determines the normalization factors $\hat{\boldsymbol{\mu}}$. Thus only the control regions should enter the likelihood function. This requires that the background fit be done independently, using a likelihood model that does not include the signal region. It's straight forward enough to construct a likelihood model with just the control regions but one needs to be careful in how the nuisance parameters are treated, as most of these parameters are shared among both the control and signal regions. The full set of nuisance parameters can be broken up into four subsets: $\boldsymbol{\eta} = \{\boldsymbol{\alpha}^{CR}, \boldsymbol{\alpha}^{shared}, \boldsymbol{\alpha}^{SR}, \boldsymbol{\mu}\}$, where

- $\boldsymbol{\alpha}^{CR}$ is the set of all systematic uncertainties contributing to the control regions only, e.g. the electron trigger efficiency;
- $\boldsymbol{\alpha}^{SR}$ is the set of all systematic uncertainties contributing to the signal region only, e.g. theoretical uncertainties on the signal simulation;
- $\boldsymbol{\alpha}^{shared}$ is the set of all the systematic uncertainties contributing to both the control and signal regions, e.g. the jet energy scale and resolution;
- $\boldsymbol{\mu}$ is the set of all the background normalization factors, which belong to the control regions.

For the following it will be convenient to define the additional two sets: $\boldsymbol{\eta}^{CR} = \{\boldsymbol{\alpha}^{CR}, \boldsymbol{\alpha}^{shared}, \boldsymbol{\mu}\}$ for the control regions and $\boldsymbol{\eta}^{SR} = \{\boldsymbol{\alpha}^{SR}, \boldsymbol{\alpha}^{shared}\}$ for the signal region.

Using only the control region parameter set $\boldsymbol{\eta}^{CR}$, one can construct a similar model as eq. 5.8,

$$L^{CR}(\boldsymbol{\eta}^{CR}) = \prod_{r}^{control} \text{Pois}(n_r|\nu_r(\boldsymbol{\eta}^{CR})) \prod_{e}^{events} f_r(x_{re}|\boldsymbol{\eta}^{CR}) \times \\ \text{Unif}(\mu_Z)\text{Unif}(\mu_W) \prod_{\alpha\in\{\boldsymbol{\alpha}^{CR},\boldsymbol{\alpha}^{shared}\}}^{systematics} \text{Gauss}(\alpha) \quad (5.12)$$

except now the product is limited to the control regions only.

Now the background-only fit is performed by maximizing L^{CR} with respect to $\boldsymbol{\eta}^{CR}$, or equivalently minimizing the negative log likelihood

$$\boldsymbol{\eta}^{CR} = \arg\min_{\boldsymbol{\eta}^{CR}}(-\log L^{CR}(\boldsymbol{\eta}^{CR})).$$

The total background estimate, ν_{tot} is then give by the results of the fitted values

$$\nu_{tot} = \nu_{tot}(\hat{\boldsymbol{\eta}}^{CR}).$$

The uncertainty on the total background estimate, σ_{tot}, is then calculated using the standard error propagation formula

$$\sigma^2_{tot} = \underbrace{\sum_i^n V_{ii} \left(\frac{\partial \nu_{tot}}{\partial \eta_i^{CR}}\right)^2_{\eta_i=\hat{\eta_i}}}_{\text{variance terms}} + \underbrace{\sum_i^n \sum_{j\neq i}^n V_{ij} \left(\frac{\partial \nu_{tot}}{\partial \eta_i^{CR}}\right)_{\eta_i=\hat{\eta_i}} \left(\frac{\partial \nu_{tot}}{\partial \eta_j^{CR}}\right)_{\eta_j=\hat{\eta_j}}}_{\text{covariance terms}} \tag{5.13}$$

where V_{ij} are the covariance matrix elements from eq. 5.10. Here the first set of terms correspond to the variances of the nuisance parameters in terms of the event counts, which the second set of terms gives their correlations. These terms are useful for breaking down the different contributions of the nuisance parameters to the total background estimate uncertainty. This can be done by setting all on the nuisance parameters to their nominal values, with the exception of the parameter of interest which is left to float. Then the variance term in 5.13 is calculated and used as an estimate on the parameter's uncertainty contribution to the total background yield.

The ML method also allows for the likelihood function to be profiled, which is a technique used to fix nuisance parameters and reduce the dimensionality of the likelihood to only a few parameters of interest. For example if $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)$ and one was only interested in one of these parameters, say α_0, then α_1 and α_2 may be replaced with their respective MLEs $\hat{\alpha_1}$ and $\hat{\alpha_2}$ where α_0 is allowed to float during the fit. This leads to a likelihood function $L(\alpha_0|\hat{\alpha_1}, \hat{\alpha_2})$ which is now only a single valued function over α_0. This profiling technique is useful for measuring only certain *parameters of interest* (POI), and will be used to set limits (one sided confidence intervals) on signal strength parameters.

Once the parameters $\hat{\boldsymbol{\eta}}^{CR}$ have been determined, the signal region can be added to the likelihood model and a second fit can be performed to find $\hat{\boldsymbol{\eta}}^{SR}$ by profiling over $\hat{\boldsymbol{\eta}}^{CR}$.

Chapter 6

Model independent results

This chapter presents the background-only fit results which are used to set model independent limits on BSM processes.

The chapter begins with section 6.1 which presents the MLE values with their respective uncertainties (table 6.1) and covariances (figure 6.3) after the background-only fit has been preformed. MLE uncertainties are propagated to the signal region, where each individual contribution is broken down by nuisance parameter (table 6.3). Total event yields in the the four control regions and signal region are shown (6.2) and split up by background processes. In addition, effects of the background-only fit on the W$\mu\nu$ and Zee control regions are displayed (figure 6.1) and select signal region distributions (figure 6.2) are shown for shape comparisons between Monte Carlo and data.

Section 6.2 examines and motivates the model independent limit setting procedure. A likelihood ratio test statistic is adopted using the likelihood model developed in section 5.1.1 to test against the background-only hypothesis. The distribution of the test statistic in its asymptotic approximation is then used to calculate the 95% CL limit. The results are then presented in table 6.4 which are interpreted in terms of upper limits on signal event counts and BSM effective cross sections.

6.1 Fit results

In this section the results of the background-only fit are presented. The fit is performed by maximizing eq. 5.12 over the parameter set $\boldsymbol{\eta}^{CR}$. All of the fit parameters, with their associated uncertainties, are propagated to the signal region. The best fit

Parameter	initial value and error	fitted value and error
mu_Z	1.00 ± 1.00	0.86 ± 0.03
mu_W	1.00 ± 1.00	0.88 ± 0.04
alpha_VVSys	0.00 ± 1.00	-0.25 ± 0.96
alpha_TopSys	0.00 ± 1.00	0.20 ± 0.91
alpha_DijetSys	0.00 ± 1.00	0.00 ± 1.00
alpha_RadWSys	0.00 ± 1.00	0.43 ± 0.38
alpha_PUSys	0.00 ± 1.00	0.37 ± 0.85
alpha_JetScale	0.00 ± 1.00	0.00 ± 0.99
alpha_JetRes	0.00 ± 1.00	-0.02 ± 0.99
alpha_MetScale	0.00 ± 1.00	-0.58 ± 0.92
alpha_MetRes	0.00 ± 1.00	0.80 ± 0.77
alpha_ElScale	0.00 ± 1.00	0.01 ± 0.99
alpha_ElRes	0.00 ± 1.00	-0.02 ± 1.05
alpha_MuScale	0.00 ± 1.00	-0.01 ± 1.02
alpha_MuRes	0.00 ± 1.00	0.00 ± 0.99

Table 6.1: Floating fit parameters for the analysis involving control regions Zee, Z$\mu\mu$, W$e\nu$, W$\mu\nu$, before (left) and after (right) the background-only fit. The fit is performed by minimizing eq. 5.12 using the numerical minimizer `MINUIT`. The quoted fit errors come from the Hessian matrix eq. 5.10 called numerically using the `MINUIT HESSE` subroutine.

values, uncertainties, event yields and distributions in both the control and signal regions are presented and discussed.

6.1.1 Parameters and best fit values

The fit parameters parameterize various corrections applied to MC simulations all deriving from unique sources. Yet, their effects on the signal region can be coarsely categorized into three groups. First there are the parameters that effect the overall event counts by scaling total contributions of a background process. These parameters reflect theoretical uncertainties on the background cross sections. Second are parameters that affect event counts by scaling contributions according to a dependent variable or parameter. These parameters are weight-based and derive from a number of different sources such as radiative corrections and pileup. The third and last type of parameters affect event counts by changing the events that pass selection. These parameters are event-based deriving from energy scale and resolution corrections to detector objects such as jets, missing energy and leptons. All of the background-only fit parameters are collected and categorized into one of the mentioned groups below.

Normalization parameters

- mu_W and mu_Z are the W/Z + jets normalization factor parameters μ_W and μ_Z introduced in section 5.3.
- alpha_VVSys, alpha_TopSys and alpha_DijetSys parameterize corrections on the overall event count for the Diboson, Top and Multijet background processes. These parameters differ from the normalization factors by the fact that they lack a control region to be accurately estimated; instead they are assigned conservative uncertainties with Gaussian constraint functions.

Weight-based parameters

- alpha_RadWSys parameterizes the radiative corrections applied to W + jets processes; weights are applied as a function of the W boson p_T.
- alpha_PUSys parameterizes corrections associated with the pileup; weights are applied as a function of the number of interactions in the collision event.

Event-based parameters

- alpha_ElScale and alpha_ElRes parameterize the electron energy scale and resolution corrections.
- alpha_MuScale and alpha_MuRes parameterize the muon energy scale and resolution corrections.
- alpha_JetScale and alpha_JetRes parameterize the jet energy scale and resolution corrections.
- alpha_MetScale and alpha_MetRes parameterize the missing transverse energy scale and resolution corrections.

The event-based corrections are applied to all detector objects identified after reconstruction but before selection. These corrections are also applied in the missing energy calculation before selection.

The parameters with the alpha prefix are associated with systematic uncertainties, while those with the mu prefix correspond to the normalization factors for the W/Z + jets backgrounds. All of the alpha parameters are accompanied by Gaussian constraint functions in the likelihood 5.12. These values are all standardized by translating the means to zero and scaling to unit variance, based on their up/down variations according to eq. 5.6. Uncertainties on all fit parameters are symmetric by construction.

Table 6.1 shows the best fit values, as well as their associated uncertainties, before and after the background-only fit has been performed.The quoted fit errors come from the Hessian matrix eq. 5.10 called numerically using the `MINUIT HESSE` subroutine. All of the fitted values fall well within their respective up/down auxiliary measurement variations. However, one may wonder why some of the fitted values, such as alpha_MuRes, seem to have changed very little from their nominal values. Little to no change will most likely reflect the fact that these parameters have negligible impact on the resulting likelihood value and overall event yields. In contrast, movements away from nominal values indicate that the likelihood is more sensitive to these parameters.

These best fit values along with their fitted uncertainties are used to estimate event yields and confidence intervals.

Unlike the alpha parameters, the mu parameters are initialized to one—corresponding to the MC expected backgrounds—and are allowed to vary unrestricted by 100% in either direction during the fit. Figure 6.1 shows the effects of the mu parameters on the (lepton-removed) E_T^{miss} distributions before and after the background-only fit. It can be seen that the resulting fit values scale down the MC expected background to better accommodate the data. One should also note that the systematic uncertainties (grey hatched bars) are only defined after the fit.

6.1.2 Event yields and signal region distributions

With the best fit values the background estimates are calculated in the signal region with propagated uncertainties. Then the total background estimates and discriminating variable distributions are compared against data.

Table 6.2 shows the observed and expected (fitted) event yields in the control and signal regions after the background-only fit. For comparison purposes the nominal Monte Carlo predictions are also shown; however, they are not used in the model or

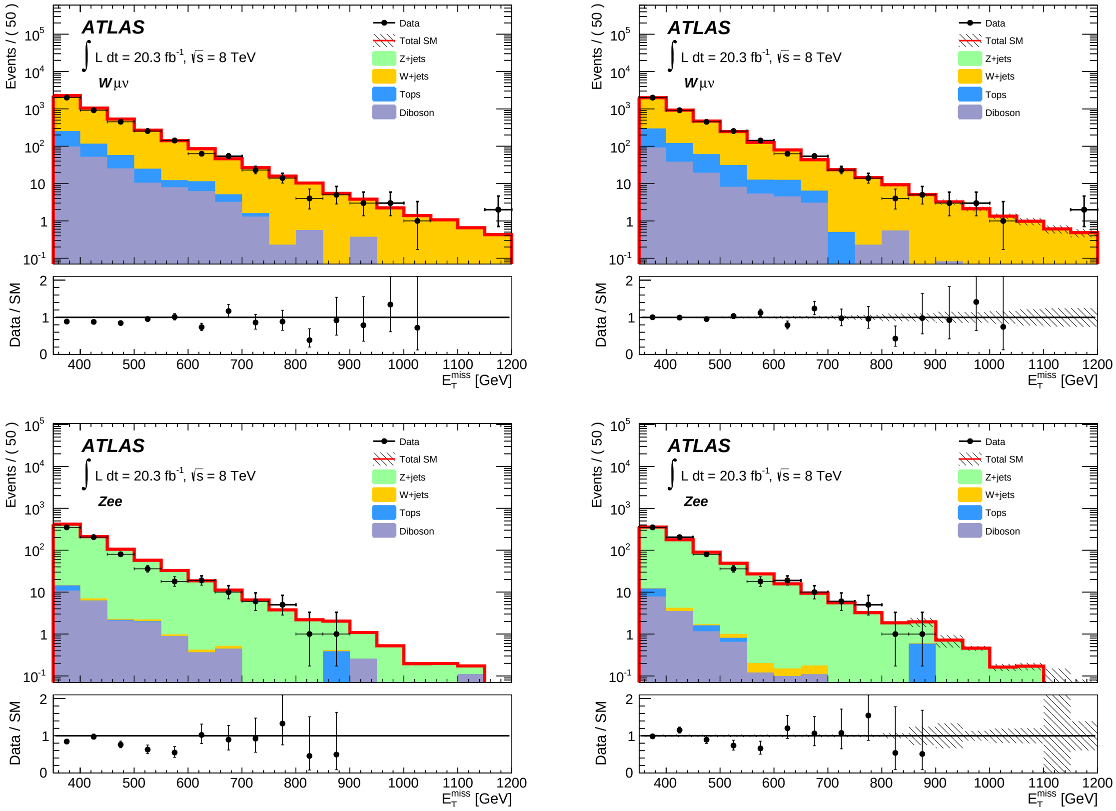


Figure 6.1: Comparison of (lepton-removed) E_T^{miss} distributions in the W$\mu\nu$ (top) and Zee (bottom) control regions before (left) and after (right) background-only fit. Systematic uncertainties (grey hatched bars) are only shown for after-fit results.

Regions	Zee	Z$\mu\mu$	W$e\nu$	W$\mu\nu$	SR
Observed events	732	636	3418	3940	8925
Fitted bkg events	748.1 ± 20.7	621.2 ± 17.0	3424.1 ± 44.0	3932.4 ± 50.6	8801.8 ± 211.3
Fitted W events	1.5 ± 0.1	0.14 ± 0.0	3005.3 ± 152.4	3464.6 ± 170.2	2034.8 ± 97.1
Fitted Z events	724.7 ± 23.4	597.0 ± 20.3	13.3 ± 0.6	25.3 ± 0.8	6315.1 ± 211.3
Fitted Top events	3.9 ± 2.3	7.4 ± 3.5	258.1 ± 122.4	279.9 ± 130.3	90.0 ± 42.7
Fitted Diboson events	17.9 ± 9.4	16.6 ± 8.8	147.3 ± 76.7	162.5 ± 86.2	346.0 ± 180.7
Fitted Multijet events	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	$15.8^{+15.9}_{-15.8}$
MC exp. SM events	868.1	718.0	3866.1	4442.7	10153.1
MC exp. W events	1.7	0.1	3396.5	3925.5	2310.6
MC exp. Z events	838.9	689.0	15.4	29.2	7292.9
MC exp. Top events	4.3	7.7	267.1	281.2	92.7
MC exp. Diboson events	23.2	21.2	187.2	207.0	442.2
MC exp. Multijet events	0.0	0.0	0.0	0.0	14.7

Table 6.2: Fit results for the control and signal regions for an integrated luminosity of 20.3fb^{-1}. The results are obtained from the control regions using the background-only fit. Nominal MC expectations (normalized to MC cross-sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties, which are added in quadrature. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

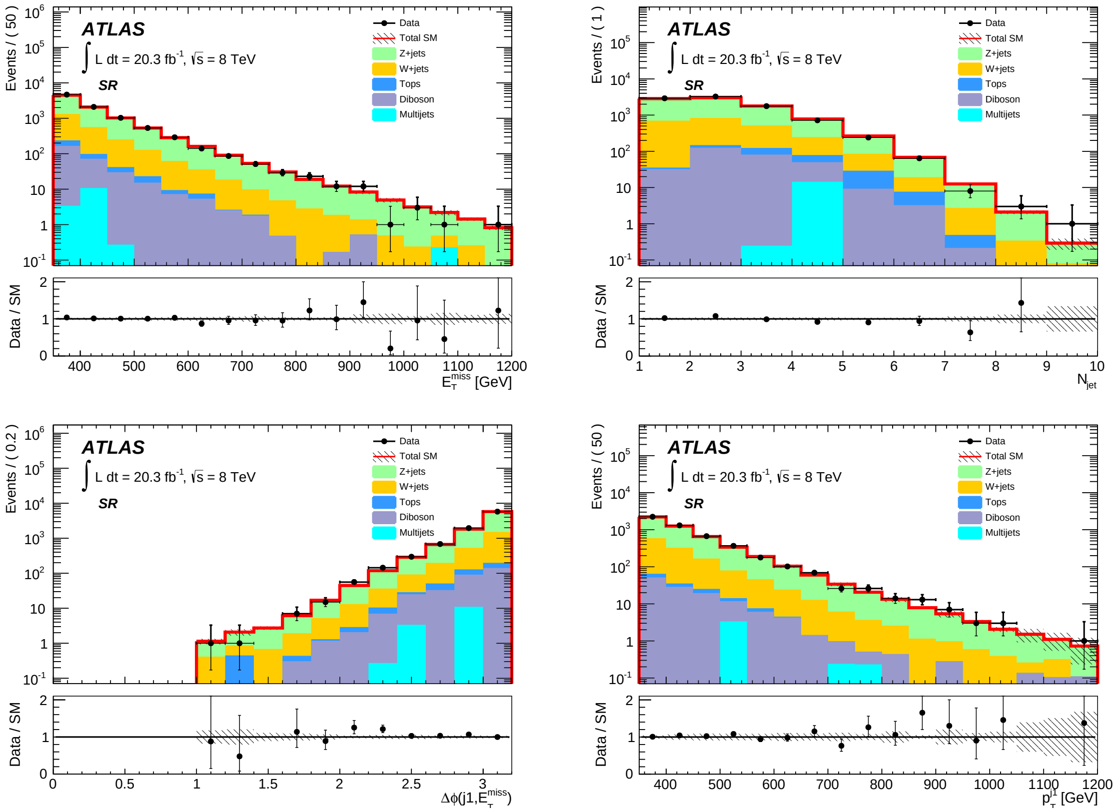


Figure 6.2: Signal region distributions of E_T^{miss}, N_{jets}, $p_T^{j_1}$ and $\Delta\phi(j_1, E_T^{miss})$ after background-only fit. Error bars on data are estimated as the square root of the number of events in a given bin, while grey hatched bars correspond to total uncertainty deriving from the fit parameters (systematic uncertainties).

limits. The event yield in the signal region is measured to be 8925, which is well within the expected value of 8801 ± 211.

The expected contributions from individual background processes, as well as their estimated absolute uncertainties, are shown. As expected Z + jets and W + jets give the largest contribution in the signal region, making up roughly 72% and 23% of the background, respectively. The remaining 5% is composed of Diboson (4%), Top (1%) and Multijet (0.1%) background processes. The Z + jets process contributes the largest uncertainty in the signal region, deriving from the μ_Z normalization factor. Diboson processes contribute the second largest uncertainty, which derive from a very similar source, the uncertainty on the Diboson cross sections. However, since these two sources of uncertainty have similar but competing effects they cancel to a large degree (see section 6.1.3 for discussion).

Figure 6.2 shows the E_T^{miss}, N_{jets}, $p_T^{j_1}$ and $\Delta\phi(j_1, E_T^{miss})$ distributions in the signal region after the background-only fit has been performed. E_T^{miss} and $p_T^{j_1}$ are expected to have similar distribution since the leading jet carries most of the visible energy as it recoils off of the invisible particle(s). In such cases the leading jet and missing energy will be back-to-back, leaving an open angle of $\Delta\phi(j_1, E_T^{miss}) \approx \pi$ between the two objects. However, this scenario breaks down as the jet multiplicity N_{jets} increases allowing for the visible energy to be more isotropically distributed in the transverse plane.

Even though no reference to a specific signal model is made, there are some qualitative features one can look for in such a distributions that would be indicative of BSM physics. In the E_T^{miss} and $p_T^{j_1}$ distributions one would expect to see a general excess of events in the tail of the distribution. These heavy-tailed distributions come from a harder E_T^{miss} spectrum due to the additional energy necessary to create high mass particles.

While the E_T^{miss} and $p_T^{j_1}$ distributions are sensitive to BSM physics the N_{jets} and $\Delta\phi(j_1, E_T^{miss})$ distributions are sensitive to MC mismodelling effects and fake jets contaminating the data. Poor MC modelling is often seen in angular distributions such as $\Delta\phi(j_1, E_T^{miss})$ since they rely on precise modelling of multiple objects, often with very different detector responses and calibration schemes. Large fake jet rates would be expected to show up as an excess in N_{jets}, the jet count per event.

The plot insets below the histograms in figure 6.2 show the ratio of data to expected values. Such plots are designed for easy comparison of data to expected values, especially when distributions are presented on a log scale. Perfect agreement corresponds to ratio values of one.

Overall agreement between data and Standard Model expectations is good, with no significant deviations observed.

6.1.3 Uncertainties

In section 5.1.2 the total observed event count n_{obs} is treated as a random variable sampled from a Poisson distribution. Since random variables are randomly sampled from their parent distributions they all experience what physicists refer to as *statistical uncertainty* (statisticians call it *sample error*). Statistical uncertainty is the amount by which n_{obs} is expected to deviate from its expected value ν_{tot}. For all Poisson distributed variables this is simply given by the square root of the expected value. However, as $n_{obs} \to \infty$ the relative difference between $\sqrt{\nu_{tot}}$ and $\sqrt{n_{obs}}$ shrinks—allowing one to approximate the statistical uncertainty with $\sqrt{n_{obs}}$, which is often more convenient. Sometimes one attributes this type of uncertainty to n_{obs} itself rather than to its expected value. This is customary for histograms where the statistical uncertainty is displayed as error bars on the data rather than Monte Carlo. Additionally, Monte Carlo simulations also introduce statistical uncertainty associated with ν_{tot}. This uncertainty is typically small, as the MC samples are produced at a much larger integrated luminosity and then scaled down to match that data. This uncertainty is absorbed into the final uncertainty for ν_{tot}.

In addition to statistical uncertainty, there are the systematic uncertainties associated with the estimate of ν_{tot}. These uncertainties derive from the best fit values and uncertainties of the nuisance parameters of the likelihood model. Since statistical and systematic uncertainty derive from independent sources they are always added in quadrature to obtain the total uncertainty. However it is useful to show both as it allows one to determine if the sensitivity of the analysis is limited by statistical or systematic uncertainty.

The total and individual contributions of all uncertainties to the expected background in both the control and signal regions are shown in table 6.3. The background event count uncertainties for a given fit parameter are calculated using the variance terms in eq. 5.13. The largest sources of uncertainty in the signal region derive from mu_Z (2.3%), alpha_VVSys (2.0%), mu_W (1.1%) and statistical uncertainty (1.0%). The relatively high contributions from mu_Z and mu_W reflect the importance of the control region measurements. Additional data—in the form of new control regions or future run periods—could be used to reduce this uncertainty. The uncertainty associated with alpha_VVSys is large due to the fact that it lacks a control region to constrain its overall normalization. Construction of such a control region could significantly reduce its uncertainty; however it should be noted that due to its negative correlations with mu_Z and mu_W its impact is not nearly as large as it may appear.

The smallest sources of uncertainty come from alpha_MuRes, alpha_MuScale and alpha_JetScale, which all contribute $\ll 1\%$ to the total background. While the relatively small contribution of the jet energy scale alpha_JetScale to the overall uncertainty may seem somewhat surprising, it can be attributed to the high cut placed on the E_T^{miss} compared to $p_T^{j_1}$. Since these two variables are highly correlated—due to the monojet topology—the $p_T^{j_1}$ distribution will be centred well above the lead jet

region	Zee	Z$\mu\mu$	W$e\nu$	W$\mu\nu$	SR
Total background expectation	748.0	621.3	3424.1	3932.4	8801.8
Total statistical ($\sqrt{N_{exp}}$)	±27.4	±24.9	±58.5	±62.7	±93.8
Total background systematic	±20.7 [2.77%]	±17.0 [2.74%]	±44.0 [1.29%]	±50.6 [1.29%]	±189.5 [2.15%]
mu_Z	±23.7	±19.5	±0.4	±0.8	±206.3
alpha_VVSys	±9.4	±8.7	±77.2	±85.7	±181.3
alpha_PUSys	±8.9	±2.4	±82.6	±104.5	±83.9
alpha_TopSys	±1.7	±3.2	±112.1	±121.6	±39.1
alpha_MetRes	±0.5	±1.0	±0.9	±0.5	±0.3
alpha_ElScale	±0.2	±0.2	±1.4	±2.2	±6.6
alpha_ElRes	±0.1	±0.1	±0.6	±0.8	±0.3
mu_W	±0.1	±0.0	±149.8	±172.7	±101.4
alpha_MuScale	±0.1	±0.1	±0.2	±0.2	±0.0
alpha_MetScale	±0.1	±0.9	±5.8	±3.5	±4.0
alpha_RadWSys	±0.0	±0.0	±0.2	±0.1	±3.3
alpha_JetRes	±0.0	±0.0	±0.6	±0.7	±1.6
alpha_JetScale	±0.0	±0.0	±0.8	±0.4	±0.1
alpha_MuRes	±0.0	±0.0	±0.0	±0.0	±0.0
alpha_DijetSys	±0.0	±0.0	±0.0	±0.0	±15.8

Table 6.3: Breakdown of the dominant systematic uncertainties on background estimates in the control and signal regions. Note that the individual uncertainties can be correlated, and do not necessarily add up quadratically to the total background uncertainty. The percentages show the size of the uncertainty relative to the total expected background.

cut and thus unaffected by small changes in energy scale.

Correlations between parameters provide important insight into interactions of uncertainties in the fit. Figure 6.3 shows the correlation matrix for all of the control region fit parameters. The matrix elements are the Pearson correlation coefficients, ρ_{ij}, calculated as

$$\rho_{ij} = \frac{V_{ij}}{\sqrt{V_{ii}V_{jj}}}$$

where V_{ij} is the covariance of the ith and jth fit parameter from eq. 5.10. Positive (negative) correlation coefficients close to one indicate that varying one parameter will cause the other to move in the same (opposite) direction with a linear relationship. It should be noted that non-linear interactions between parameters is also possible, but such relationships are not always captured by their correlation coefficients.

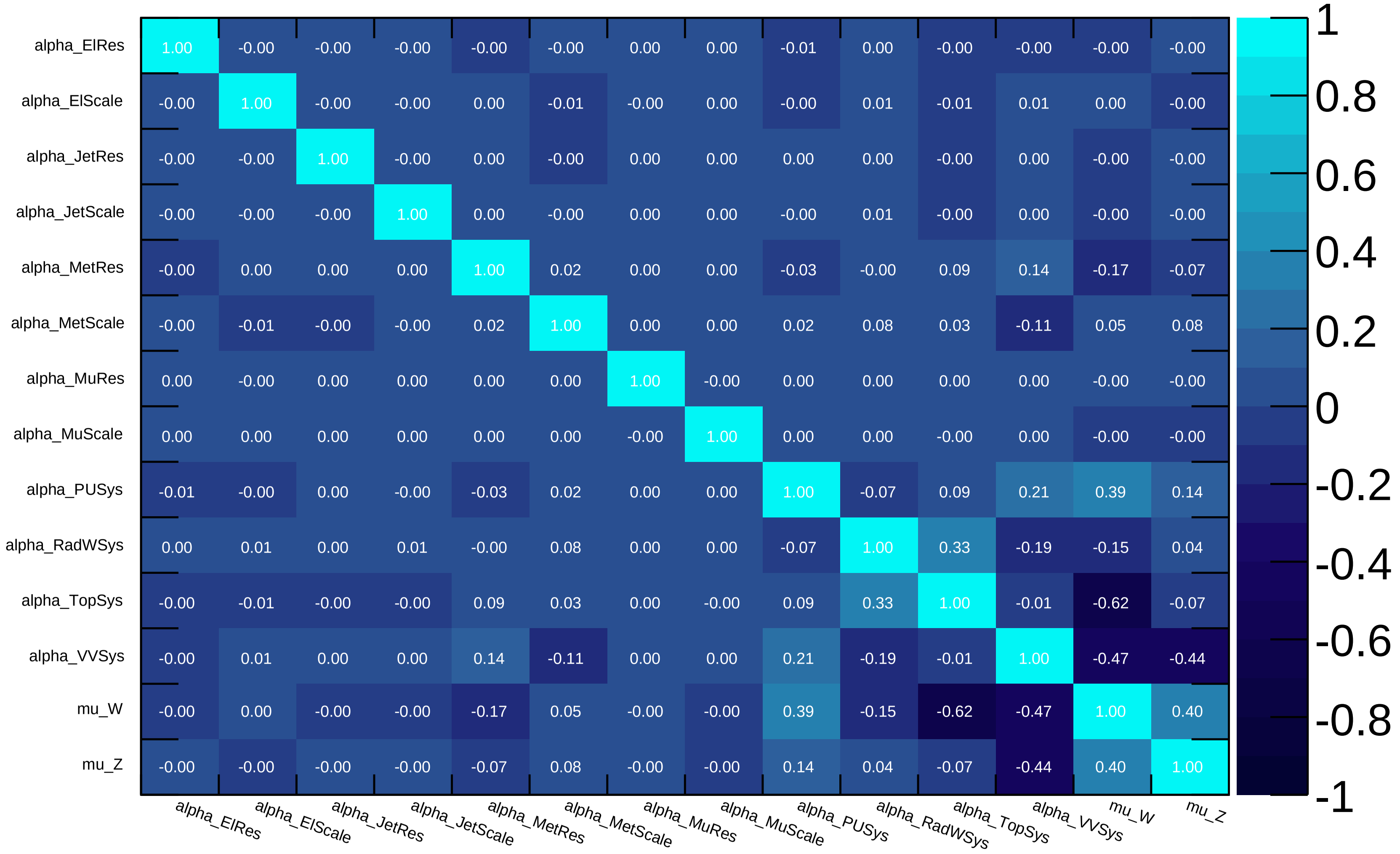
Some of the correlations between the fit parameters are quite large, especially for related processes. For example the correlation between mu_W and alpha_topSys is -0.64, indicating any movement in mu_W will be accompanied by a movement of alpha_topSys in the opposite direction. This effect is most likely due to the heavy contamination of the Top background ($t\bar{t}$ and single t) in the W control regions. Since alpha_topSys correspond to the uncertainty of the overall normalization of the Top backgrounds it's responsible for a very similar type of scaling effect as mu_W. If we neglect the effects of the Gaussian constraint function for alpha_topSys in eq. 5.12 the expected event count for the top background processes in the W control regions can be approximated by the first-order expansion

$$\nu_{Top}(\alpha_{Top}) \approx (a + b\alpha_{Top})\nu_{Top}$$

where a and b are constants. Putting this into eq. 5.11 gives

$$\nu_{tot} = \mu_W \nu_W + (a + b\alpha_{Top})\nu_{Top} + \nu_{other}$$

where ν_{other} is the remaining contributions from the other background processes. Holding all nuisance parameters expect μ_W and α_{Top} fixed (corresponding to 5.10), it is clear that any increase in μ_W will have to be accompanied by a decrease in ν_{Top} since the data constrains ν_{tot}. This effect leads to the negative correlation seen in fig. 6.3. Similar arguments can be made to explain the relatively large correlations between parameters that resulting in an overall scaling effect.



	alpha_ElRes	alpha_ElScale	alpha_JetRes	alpha_JetScale	alpha_MetRes	alpha_MetScale	alpha_MuRes	alpha_MuScale	alpha_PUSys	alpha_RadWSys	alpha_TopSys	alpha_VVSys	mu_W	mu_Z
alpha_ElRes	1.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.01	0.00	-0.00	-0.00	-0.00	-0.00
alpha_ElScale	-0.00	1.00	-0.00	-0.00	0.00	-0.01	-0.00	0.00	-0.00	0.01	-0.01	0.01	0.00	-0.00
alpha_JetRes	-0.00	-0.00	1.00	-0.00	0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00
alpha_JetScale	-0.00	-0.00	-0.00	1.00	0.00	-0.00	0.00	0.00	-0.00	0.01	-0.00	0.00	-0.00	-0.00
alpha_MetRes	-0.00	0.00	0.00	0.00	1.00	0.02	0.00	0.00	-0.03	-0.00	0.09	0.14	-0.17	-0.07
alpha_MetScale	-0.00	-0.01	-0.00	-0.00	0.02	1.00	0.00	0.00	0.02	0.08	0.03	-0.11	0.05	0.08
alpha_MuRes	0.00	-0.00	0.00	0.00	0.00	0.00	1.00	-0.00	0.00	0.00	0.00	0.00	-0.00	-0.00
alpha_MuScale	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	1.00	0.00	0.00	-0.00	0.00	-0.00	-0.00
alpha_PUSys	-0.01	-0.00	0.00	-0.00	-0.03	0.02	0.00	0.00	1.00	-0.07	0.09	0.21	0.39	0.14
alpha_RadWSys	0.00	0.01	0.00	0.01	-0.00	0.08	0.00	0.00	-0.07	1.00	0.33	-0.19	-0.15	0.04
alpha_TopSys	-0.00	-0.01	-0.00	-0.00	0.09	0.03	0.00	-0.00	0.09	0.33	1.00	-0.01	-0.62	-0.07
alpha_VVSys	-0.00	0.01	0.00	0.00	0.14	-0.11	0.00	0.00	0.21	-0.19	-0.01	1.00	-0.47	-0.44
mu_W	-0.00	0.00	-0.00	-0.00	-0.17	0.05	-0.00	-0.00	0.39	-0.15	-0.62	-0.47	1.00	0.40
mu_Z	-0.00	-0.00	-0.00	-0.00	-0.07	0.08	-0.00	-0.00	0.14	0.04	-0.07	-0.44	0.40	1.00

Figure 6.3: Correlation matrix of systematic uncertainties after background fits have been performed. The quoted correlations come from the Hessian matrix eq. 5.10 called numerically using the `MINUIT HESSE` subroutine.

6.2 Model independent limits

After the background-only fit has been performed it is natural to ask if the expected event count in the signal region is "compatible" with the observed value. However, defining exactly what compatible means in this sense is no simple task—indeed, the field of statistics is deeply divided on the issue. Statisticians divide into two main camps—employing either Frequentist or Bayesian methods. While a full discussion comparing the merits and shortcoming of the two methodologies is not necessary here it is worth noting that the Maximum Likelihood approach to parameter estimation developed in 5.1.1 is manifestly Frequentist, as will be the approach take here for hypothesis testing and setting limits.

In the Frequentist approach to hypothesis testing one must define a *test statistic*, which is used to either accept or reject the *null* hypothesis—as apposed to the *alternative* hypothesis. Test statistics are real-valued functions of the data, designed to optimally differentiate between the null and alternative hypotheses. A *critical* value is assigned to the test statistic according to an accepted *confidence level* (CL), which is used to categorize the outcome of a given experiment as belonging to either the null or alternative hypothesis. The accepted CL is somewhat dependent on the field and type of experiment being conducted, but the 95% CL is widely accepted as the minimal CL value for a result to be "statistically significant". For this reason the 95% CL is used to define a one-sided confidence interval—otherwise known in HEP as a *limit*.

6.2.1 The background-only hypothesis

For this analysis the null hypothesis corresponds to the background-only hypothesis, that is that the observed event count n_{obs} can be accounted for by background processes alone. This is tested against a general alternative hypothesis which simply requires that there be something in addition to the background, otherwise known as the signal-plus-background hypothesis.

To perform such a test the likelihood model must be extended to accommodate a possible signal. This can be achieved by adding a signal component to the expected event count, with a signal normalization parameter μ_{sig}. Thus the total event count in the signal region is given by

$$\nu_{tot} = \mu_{sig}\nu_{sig} + \nu_{bkg}.$$

Now the background-only hypothesis corresponds to the case where $\nu_{tot} = \nu_{bkg}$ or when $\mu_{sig} = 0$, while the alternative hypothesis is given by the situation where $\mu_{sig} > 0$.

Here μ_{sig} is a scaling factor associated with the signal, which can be interpret in a few different ways depending on the nature of the analysis. For example one could interpret μ_{sig} as the branching fraction of some BSM process, as is the case in section 8.2 for the decay of a Higgs boson into dark matter.

Since in the model independent case one is not testing against any model in particular one does not have any information on ν_{sig} or $f_{sig}(X)$, the signal distribution of X. Thus one must integrate over $f_{tot}(X)$ to remove any shape information from the likelihood model while ν_{sig} can be set to a dummy value, such as $\nu_{sig} = 1$. As a result μ_{sig} can be interpreted as the signal event count.

With the hypotheses clearly defined the next step is to construct a test statistic which will be used to assign one hypothesis over the other.

6.2.2 Log likelihood ratio test statistic

A good test statistic should summarize the data in such a way as to be able to best distinguish between the null and alternative hypotheses. The LHC experiments have adopted a profiled likelihood ratio test statistic, mainly motivated by the Neyman-Pearson lemma that states that the most powerful test statistic for two simple hypotheses is the likelihood ratio given that all parameters are known [89].[1]

In deciding whether to accept or reject the background-only hypothesis it seems reasonable to ask if the data is more likely with $\mu_{sig} = 0$ or $\mu_{sig} > 0$. This implies two separate fits for the nuisance parameters $\boldsymbol{\eta}$ will have to be performed. The first fit needs to be done with μ_{sig} fixed to zero, while the second lets μ_{sig} float. In general one should not expect the MLE parameters $\hat{\boldsymbol{\eta}}$ of both fits to be the same, since forcing μ_{sig} to take on a particular value will effect how $\boldsymbol{\eta}$ maximize the likelihood function. Fixing the value of one or more of the nuisance parameters is a type of profiling and the MLE parameters that maximize the profile likelihood are called *conditional maximum likelihood estimates* (CMLE) conventionally denoted with a double hat

$$\hat{\hat{\boldsymbol{\eta}}} = \arg\min_{\boldsymbol{\eta}} -\log L(\mu_{sig} = \text{const}, \boldsymbol{\eta}).$$

---

[1] Of course μ_{sig} is not known and the MLEs are only approximate values for the nuisance parameters. However, profiling is often regarded as a natural generalization of Neyman-Pearson procedure [80].

Since $\hat{\hat{\boldsymbol{\eta}}}$ depends on the value of μ_0 it can be considered a single-valued multidimensional function of μ_{sig}, i.e. $\hat{\hat{\boldsymbol{\eta}}} = \hat{\hat{\boldsymbol{\eta}}}(\mu_{sig})$.

The intuition behind using the profiled likelihood in constructing a test statistic is best demonstrated with an example. Say one performs two separate fits, one with $\mu_{sig} = 0$ and another where $\mu_{sig} = 1$. Which value of μ_{sig} is more likely given the data? Of course it's the one that produces the larger value of the likelihood. So if $L(0, \hat{\hat{\boldsymbol{\eta}}}(0)) > L(1, \hat{\hat{\boldsymbol{\eta}}}(1))$ one would say the background-only hypothesis is more likely than the $\mu_{sig} = 1$ signal-plus-background hypothesis and vice versa. This method only works for comparing two different values of μ_{sig}, but for the model independent limits one wants to compare $\mu_{sig} = 0$ against all other possible values. This can be achieved by letting the value of μ_{sig} float in the second fit. This line of reasoning leads to $\lambda(\mu_{sig})$, the likelihood ratio test statistic

$$\lambda(\mu_{sig}) = \frac{L(\mu_{sig}, \hat{\hat{\boldsymbol{\eta}}}(\mu_{sig}))}{L(\hat{\mu}_{sig}, \hat{\boldsymbol{\eta}})}. \tag{6.1}$$

One should note that in this ratio the denominator will always be greater than or equal to the numerator simply by the definitions of $\hat{\boldsymbol{\eta}}$ and $\hat{\mu}_{sig}$. So the largest possible value $\lambda(\mu_{sig})$ can take on is 1, which happens when $\mu_{sig} = \hat{\mu}_{sig}$.

In HEP it is conventional to use the negative logarithm of the likelihood ratio, called q_μ in stead of $\lambda(\mu_{sig})$. It is also conventional to allow μ_{sig} to take on negative values. While this is unphysical it is useful, as it allows for the common scenario where ν_{tot} over-estimates n_{obs}. Applying these conventions the test statistic q_0 is given as

$$q_0 = \begin{cases} -2\log\lambda(0) & : \hat{\mu}_{sig} > 0 \\ 0 & : \hat{\mu}_{sig} \leq 0 \end{cases}.$$

Thus large values of q_0 indicate an excess of events over expected while values close to zero indicate an event count equal to or less than expected.

6.2.3 Limit setting procedure

To measure the compatibility of the observed event count to the background-only hypothesis frequentists construct *confidence intervals*. A confidence interval is a range of values which is constructed in such a way as to have a fixed probability of containing the *parameter of interest* (POI) (such as μ_{sig}). The fixed probabilities are called *confidence levels* (CL); in HEP it is customary to calculate both the 68% (1σ) and

95% (2σ) confidence levels. Meaning, if the experiment and the confidence interval calculations were carried out multiple times 68% (95%) of them would contain the true value of the POI. In HEP confidence intervals are often referred to as "limits" and are usually one-sided since negative values are often unphysical. Limits are often interpreted as the largest/smallest values that the POI can have while still be supported by the data. However, it should be noted that such interpretations slip into the Bayesian way of thinking—in the Frequentist interpretation the POI only has one value; instead it is the limits that take on distributions.

In the previous section the test statistic q_0 was developed, which should be interpreted as a single value that represents the outcome of an experiment. However q_0 alone is not enough to build a confidence interval. Since for a given experiment one doesn't know how likely the observed value of q_0, this requires a distribution over q_0. Construction of such a distribution can be a difficult task, for typically the experiment is only performed once.[2]

There are two methods commonly used in HEP: "Toy" Monte Carlo and asymptotic formulas. The Toy Monte Carlo method uses the likelihood model to generate random dataset samples, where each sample leads to a different q_0 value, allowing a distribution to be built up through nonparametric methods for numerical integration [80]. The asymptotic approach uses Wilk's theorem [90] and Wald's theorem [91] to find a analytical form for the distribution of q_0 which can then be integrated directly [81]. This method works for sufficiently large data samples. Since the signal region selection in this analysis is fairly loose the data sample is relatively large and thus the asymptotic approach is most appropriate here.

Following [81] the distribution of q_0 in it's asymptotic approximation is

$$f(q_0|\mu_{sig} = 0) = \frac{1}{2}\delta(q_0) + \frac{1}{2\sqrt{2\pi q_0}}\mathrm{e}^{-\frac{1}{2}q_0} \tag{6.2}$$

where $\delta(\cdot)$ is the Dirac delta function.

The distribution of q_0 can now be integrated over to calculate probabilities in the typical way. Such probabilities are called a p-values, which can be seen as a function of q_0^{obs}, the observed value of q_0 given the data. The *significance* of a test is given by probability of obtaining q_0^{obs}, or greater, assuming the background-only hypothesis.

---

[2] Even if it were feasible to repeat the experiment multiple times to generate a distribution it would not be an efficient use of data.

This is given by

$$p_0 = \int_{q_0^{obs}}^{\infty} f(q_0|\mu_{sig} = 0)dq_0 \tag{6.3}$$

where small values of the p_0 are considered evidence for the alternative hypothesis and thus evidence against the background-only hypothesis. In this case the integral has the closed-form solution

$$p_0 = 1 - \Phi(\sqrt{q_0}) \tag{6.4}$$

where $\Phi(\cdot)$ is the Gaussian cumulative distribution function.

In HEP it is customary to convert the p-value obtained from an experiment into the quantile of a unit Gaussian distribution. This conversion is performed by applying the inverse of the Gaussian cumulative distribution function, Φ^{-1} as so

$$Z = \Phi^{-1}(1 - p_0)$$

where the significance is then quoted as $Z\sigma$. Plugging in eq. 6.4 yields the simple expression

$$Z = \sqrt{q_0}. \tag{6.5}$$

The 68% (95%) CL limit can be obtained by setting $Z = 1$ ($Z = 2$) which corresponds to $q_0 = 1$ ($q_0 = 4$).

Although the 95% CL is considered sufficient in HEP to reject the alternative hypothesis, the requirement to accept it is set much higher. The background-only hypothesis is only rejected when the very conservative critical value of 5σ—or $p_0 = 2.87 \times 10^{-7}$—is reached. Such high standards are set to limit the false positive rate of discoveries.[3]

6.2.4 Expected sensitivity

The expected upper limit on μ_{sig} is the upper limit one would expect to obtain if the background-only hypothesis were true. To find the expected limit at the 95% CL μ_{sig}^{95} one needs to know its distribution $f(\mu_{sig}^{95}|\mu_{sig} = 0, \boldsymbol{\eta})$. However, $\boldsymbol{\eta}$ are unknown so one must use the CMLEs $\hat{\hat{\boldsymbol{\eta}}}(0)$ instead. This means that the expected limits are somewhat dependent on the data since the CMLEs are. With this distribution one

---

[3] It may also be seen as Bonferroni correction to account for the look-elsewhere effect arising from the many discovery searches being conducted at any given time—and often over large parameter spaces.

Signal channel	$\langle\epsilon\sigma\rangle^{95}_{obs}$[fb]	ν^{95}_{obs}	ν^{95}_{exp}	p_0	q_0	Z
SR	16.61	337.3	$262.5^{+350.1}_{-102.1}$	0.31	0.26	0.51

Table 6.4: Left to right: 95% CL upper limits on the visible cross section ($\langle\epsilon\sigma\rangle^{95}_{obs}$) and on the number of signal events (ν^{95}_{obs}). The third column (ν^{95}_{exp}) shows the 95% CL upper limit on the number of signal events, given the expected number (and $\pm 1\sigma$ excursions on the expectation) of background events. The last two columns indicate the CL_b value, i.e. the confidence level observed for the background-only hypothesis.

can also define an uncertainty on the upper limits as well by finding the limit of integration where the p-value is equivalent to a significance of $Z = 1$

$$\int_0^{\mu^{\pm}_{sig}} f(\mu^{95}_{sig}|\mu_{sig} = 0, \hat{\hat{\boldsymbol{\eta}}}(0))d\mu^{95}_{sig} = \Phi^{-1}(\pm 1) \tag{6.6}$$

where the equation is solved for $\mu^{\pm}_{sig}$.

6.2.5 Results

Table 6.4 shows the results of the model-independent upper limit calculations. The observed upper limit on the number of signal events in the signal regions is 337.3 events at the 95% CL, well within the expected value of $262.5^{+350.1}_{-102.1}$. Also shown is the upper limit on the effective interaction cross section, which shows that this analysis is not sensitive to signals processes with an effective cross section less than 16.61 fb. The p-value of the observed results is $p_0 = 0.31$, corresponding to a significance of 0.51σ. Thus these results are consistent with the background-only hypothesis and the signal-plus-background hypothesis can be rejected.

Chapter 7

Invisible Higgs interpretation

In this chapter the detector signatures of an invisibly decaying Higgs are examined and a classification model is constructed based on Monte Carlo simulation.

In section 7.1 the three main Higgs production mechanisms are investigated with a focus on the kinematics and geometrical distributions in the relevant final states. Table 7.1 shows the discussed processes and their theoretical cross-sections calculated to NNLO.

Section 7.2 describes the classification model used in section 8.2 to define the model-dependent signal regions. A set of discriminating variables are chosen and their simulated signal distributions are compared to backgrounds (Figure 7.3). These variables are then used as inputs to two independent Random Forest classification models. The models are trained on a random sample of half of the simulated data while the other half is used to evaluate the performance of the classifiers.

7.1 Higgs production mechanisms

Higgs couplings are proportional to the particle masses entering the vertex. Therefore Higgs production mechanisms are dominated by the heaviest Standard Model particles; namely the top quark (174 GeV), Z boson (90 GeV) and the W boson (81 GeV). At the LHC this leads to three main production mechanisms depending on the initial state particles: gluon-gluon fusion (ggF) for gluon initial states, vector boson fusion (HVBF) for quark initial states and Higgs-strahlung (HV) for quark-antiquark initial states. Figure 7.1 shows the relevant Feynman diagrams. All three production mechanisms lead to different final states—although there is significant overlap in the

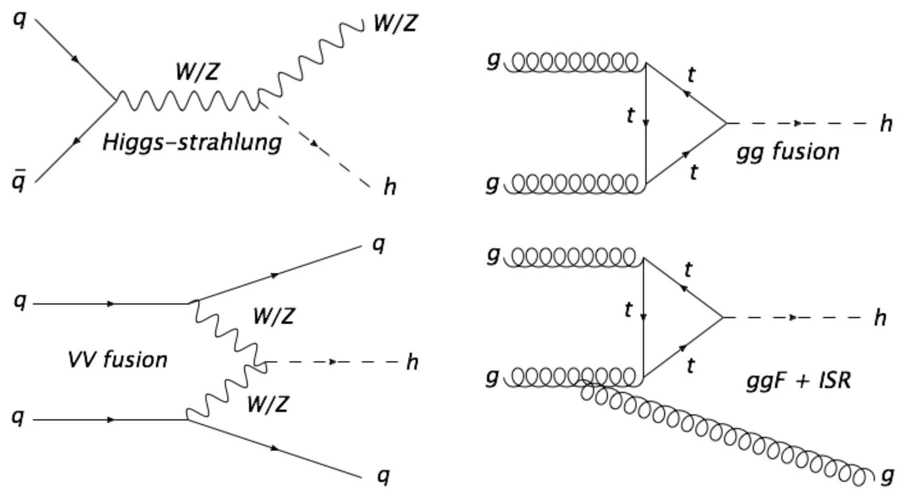


Figure 7.1: Feynman diagrams of the three dominant Higgs production mechanisms at the LHC.

detector signatures. This section is dedicated to examining these three production mechanisms and relevant final states which will be used in 8.2 to define orthogonal signal regions.

Table 7.1 shows the cross-sections and branching ratios with their respective uncertainties for the different production modes. All quoted cross-sections are taken from the most recent results of the LHC Higgs Cross Section Working Group for $\sqrt{s} = 8$ TeV [92]. Cross-sections are calculated to NNLO for QCD corrections and NLO for EW corrections. Two independent uncertainties are quoted for all cross-sections calculations; QCD uncertainties, which derive from QCD factorization and renormalization scales and PDF uncertainties which derive from the choice of Parton Distribution Function.

7.1.1 Gluon-gluon fusion

The dominant Higgs production mechanism at the LHC is that of gluon-gluon fusion $gg \to h$ via an intermediate top-quark loop. The quark loop is present since there is no fundamental ggH coupling in the Standard Model, due to the fact that gluons are massless particles and thus do not couple directly to the Higgs. Although in principle all quarks should be included in the loop, only the top quark is used in the cross-section calculation since it is roughly 35 times more massive than the bottom quark—the second heaviest of the quarks. With two quarks entering each vertex

Production mode	σ [pb]	± (QCD)	± (PDF)	± (combined)	BR [%]	± [%]
$gg \to h$	19.27	1.45	1.38	2.00	—	—
$gg \to hj$	6.177	—	—	0.221	—	—
$qq \to VVqq \to hjj$	1.578	0.003	0.043	0.043	—	—
$q\bar{q} \to Wh \to hjj$	0.696	0.007	0.016	0.017	67.60	0.27
$q\bar{q} \to Zh \to hjj$	0.394	0.013	0.010	0.016	69.91	0.06

Table 7.1: All quoted cross-sections are take from the most recent results of the LHC Higgs Cross Section Working Group for $\sqrt{s} = 8$ TeV [92]. All cross-sections area calculated to NNLO for QCD corrections and NLO for EW corrections. Two independent uncertainties are quoted for all cross-sections calculations. QCD uncertainties derives from QCD factorization and renormalization scales, while PDF uncertainties derive from the choice of Parton Distribution Function and uncertainty associated with the strong coupling constant. The last column shows the two uncertainties added in quadrature. The branching ratios are taken from the Particle Data Group [13]

this leads to a suppression scale of 35^2 for the bottom quark's contribution, which is considered negligible.

Any decay products of the Higgs that can be considered dark matter candidates must be invisible in the ATLAS detector; the presence of such particles can only be inferred by a large missing energy signature. Ordinarily in the $gg \to h \to \chi\bar{\chi}$ process the decay products are created centrally, with back-to-back momentum. In such situations no missing energy signature would be detected. However, it is possible for the Higgs to be produced with a gluon-initiated jet via a ggg vertex in the initial state—known as initial state radiation (ISR). An ISR jet in the reaction gives the Higgs, and its decay products, something to recoil off of, leaving a detectable missing energy signature. Since at least one jet is needed in the event to produce the missing energy signature it is only these types of events that will pass the missing energy trigger requirement.

The cross-section of $gg \to hj$ is suppressed by 70% compared to $gg \to h$ due to the additional vertex. Production in association with multiple jets is also possible, however such cross-sections would be expected to have a similar suppression scale and thus drop-off exponentially with jet multiplicity.

7.1.2 Vector boson fusion

After ggF the second most copious Higgs production mode at the LHC is that of vector boson fusion $qq \to VVqq \to hjj$. In vector boson fusion two W or Z bosons are radiated by quarks in the initial state which annihilate to produce a Higgs boson and two jets. Typically the two quarks continue in the forward and backward directions forming jets in close proximity to the beamline while the Higgs is produced centrally. Again, in such situations the Higgs has nothing to recoil off of and, if it decays invisibly, will not leave a missing energy signature in the detector. However, if the momentum fraction imparted to one of the vector bosons is large enough that the associated jet is centrally produced the Higgs will recoil in the opposite direction. Since the baseline signal region selection requires the leading jet to be central ($|\eta_1| < 2$) only these types of events pass selection.

Vector boson fusion events are best characterized by their forward/backward jets. The jets tend to be separated by a large polar angle and have opposite signs in pseudorapidity. The jets are also characterized by their large energy and invariant masses.

Although the Higgs vector boson fusion production mechanism proceeds through both the WWH and ZZH vertices a single cross-section is quoted since the vector bosons are virtual and thus the two processes are indistinguishable.

7.1.3 Higgs-strahlung

The third most common Higgs production mechanism at the LHC is Higgs-strahlung: $q\bar{q} \to hV$. In Higgs-strahlung a quark and anti-quark annihilate to form a vector boson, which in turn radiates a Higgs boson. In this scenario the Higgs can be considered final state radiation, sometimes referred to as bremsstrahlung radiation, hence the name "Higgs-strahlung".

Technically Higgs-strahlung refers to three different processes—one for each possible vertex: W^+W^+H, W^-W^-H and ZZH. In this analysis the W^+W^+H and W^-W^-H cross-sections are combined into one. The ZZH cross-section is treated independently due to its larger coupling constant.

For Higgs-strahlung events to pass signal region selection the W/Z boson must decay hadronically producing two quark-initiated jets. Therefore the production cross-sections need to be scaled by their respective branching ratios, which can be found in Table 7.1.

Hadronic Higgs-strahlung events can be characterized by their centrally produced jets. These jets will tend to have a relatively light invariant mass, on the same order of magnitude as their parent W/Z boson. Higgs-strahlung jets can also be expected to have a small open angle between one another and be back-to-back to the missing energy signature.

7.2 Classification model

In statistics and machine learning classification is the problem of identifying to which categories a new observation belongs. The task of classifying a given event as being "background" or "signal" based on a set of discriminating variables can be cast as a binary classification problem. Classification models are constructed based on a set of observations where the true class labels for the observations are known—a process known as *supervised training*. Training usually requires fitting a set of model parameters to the data [1] to construct a parameterized model. The model parameters are fit in such a way as to minimize a chosen metric called the *error function*. One example of an error function is the negative log of the likelihood function, such as used in section 5.1.1. In general, however, the most appropriate error function will depend on the classification model and ultimately the purpose of the classifier. Trained models can then be used to classify new observations as either background or signal to within the accuracy of the model. Typically the classification model will return a real-valued output where large positive (negative) values indicate a signal (background) event. It is then left to the user to decide what the best signal-acceptance threshold value should be. For such classification models there is always a trade-off between signal efficiency and background rejection. By adjusting the signal-acceptance threshold value one can choose the most appropriate operating point for the problem at hand.

7.2.1 Classification problem

In chapters 5 and 6 a likelihood model and test statistic were developed based on a discriminating variable X. However, in a typical HEP data analysis there will be many discriminating variables to choose from, where each may contribute additional categorical information. In principle the likelihood model can be extended to an arbitrary number of dimensions; however, in high dimensional spaces the joint probability distributions become difficult to estimate. With parametric methods the issue manifests in estimating the covariance structure of the data [80]; with non-parametric methods (such as histograms) the amount of data required scales exponentially with the dimensionality of the space[2]—a problem known as "the curse of dimensionality".

These issues motivate the use of a classifier function that maps a set of discriminating variables $\mathbf{X}$ onto a single real-valued output variable Y. Ideally this mapping is done in such a way as to maximize the discriminating power of Y between background and signal events. This is the canonical problem of supervised learning from statistical learning theory: presented with a set of tuples $(\mathbf{X}, Y)_i$ find the function f with parameters $\boldsymbol{\theta}$ that minimizes the error function $E(Y_i, f(\mathbf{X}_i; \boldsymbol{\theta}))$.

In HEP the class labels of events are fundamentally unobservable, so one must rely on simulated data for the training samples. This means any mismodeling of the

---

[1] Exceptions to this are so called "lazy learners" such as kth-nearest neighbour or Naive Bayes, which use the training data directly rather than fitting a parameterized model.

[2] If 100 is the minimum number of data points needed to populate a 1-D histogram, then for a 10-D histogram ones needs 100^{10} data points to reach the same data density.

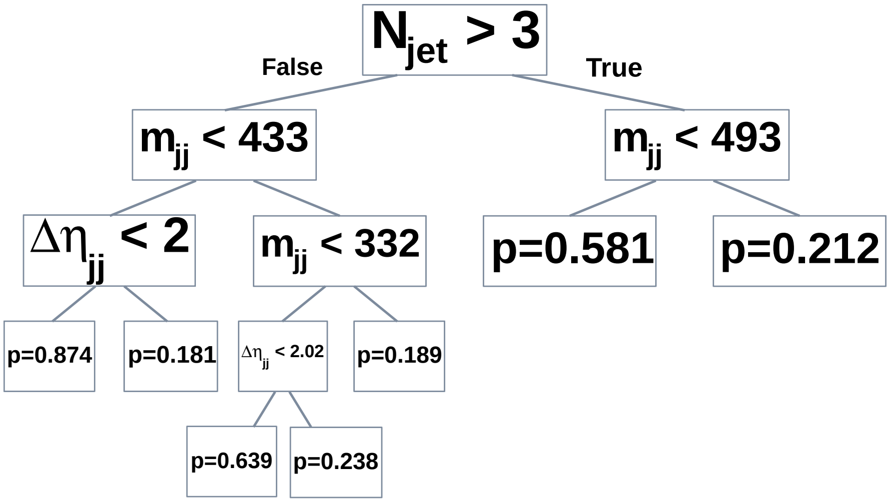


Figure 7.2: Diagram of a typical decision tree. Leaf nodes show the purity of the terminal splits.

input variables $\mathbf{X}$ will be propagated to Y. However, like $\mathbf{X}$, data/MC agreement for Y can be validated by choosing the appropriate control regions. In general the output of the classifier should be treated in the same manner as any of the other discriminating variables.

Today there are many different classification models to choose from. In HEP boosting methods using ensembles of decision trees (BDTs) are the most popular as they have proven to achieve high accuracy with little (typically zero) hyperparameter tuning [93]. Boosting methods, however, are prone to fitting statistical noise rather than underlying structure—an issue known as over-training—requiring large training sets to avoid this issue [94]. Here, due to a limited training sample in the signal region, a more conservative classifier based on ensemble bagging of decision trees called a *Random Forest* is employed instead.

7.2.2 Random Forests

The Random Forest algorithm was originally developed by Leo Breiman [95] in 2001. The algorithm combines the "bagging" method and randomization feature selection over an ensemble of decision trees.

Decision trees

A decision tree is a series of rectangular cuts on the input variables that divide the phasespace into orthogonal regions. Figure 7.2 shows a graphical representation of a typical decision tree used in this analysis. Each node can be represented by the tuple of the cut variable and threshold, often referred to as a *split*, as well as references to its children nodes. Terminal nodes are called *leaves* which are associated with the signal purity of the resulting partition

$$p = \frac{n_{sig}}{n_{bkg} + n_{sig}} \tag{7.1}$$

where n_{sig} (n_{bkg}) is the weighted number of simulated signal (background) events. The entire tree itself can be represented as a piecewise function of the input variables that encodes the tree structure and contains the splits $\boldsymbol{\theta}$ as parameters. Given a set of input values the tree $T(\mathbf{X}|\boldsymbol{\theta})$ will return the purity of the leaf it evaluates to

$$T(\mathbf{X}|\boldsymbol{\theta}) = p.$$

Trees are trained by minimizing the Gini index—which acts as the error function—given by

$$\text{Gini} = p(1 - p)$$

for every split. The optimal split for each node is found in a greedy layer-wise fashion [3] by performing a grid search over each input variable. The splitting procedure ends when the resulting split yields a pure partition (Gini = 0) or when the number of remaining data points is 5 or less.

A single decision tree by itself tends to experience high variance in its output given small changes to the inputs [94]. This is an undesirable feature as it reduced the accuracy of the classifier. The main innovation of Random Forests is to reduce this variance by averaging over many trees. For independent and identically distributed variables averaging over M decision trees reduces the variance by a factor of M^{-1}. Or in the scenario of correlated but identically distributed variables with common

---

[3] Greedy learning algorithms attempt to find global optima by making locally optimal choices at each stage in the training algorithm. This type of learning is generally preferred for its computational efficiency.

variances σ and correlations ρ the averaged variance is given by [94]

$$\rho\sigma^2 + \frac{1-\rho}{M}\sigma^2. \tag{7.2}$$

As M increases the second term disappears; however the first term can only be reduced by decorrelating the individual tree. Random Forests employ two different techniques to achieve this. First each tree is trained on a bootstrapped (sampled with replacement) sample of the data—a technique known as *bagging*. Second the variables available to perform splits at each node are restricted to a random subset of the discriminating variables. This randomization is key to the performance of the classifier since otherwise the splits tend to be dominated by just a few variables.

The Random Forest algorithm

The full Random Forest algorithm can be given as follows:

1. for $m = 1$ to M:
   (a) Draw a bootstrapped sample $\{\mathbf{X}\}^*$ from $\{\mathbf{X}\}$
   (b) Train tree $T_m(\mathbf{X};\boldsymbol{\theta})$ from $\{\mathbf{X}\}^*$ by repeating these steps until the resulting split is pure or contains 5 or less data points:
      i. randomly select the subset of variables $\mathbf{X}'$ from $\mathbf{X}$,
      ii. select the split that minimizes Gini from $\mathbf{X}'$,
      iii. use the split to define its two child nodes
2. Average the ensemble of trees $\{T_m\}$ to form the classifier

$$T_{RF}(\mathbf{X}|\{\boldsymbol{\theta}_m\}) = \frac{1}{M}\sum_{m}^{M} T_m(\mathbf{X}|\boldsymbol{\theta}_m).$$

Variable ranking

The Random Forest classifier also offers a method for ranking the importance of input variables to the classifier. A ranking of the Random Forest variables is derived by counting how often variables are used in splits, where each splits is weighted by its separation gain-squared [96]. This is a general ranking method that can be used for all decision tree based classifiers.

7.2.3 Input variables

The full set of input variables are shown in Figure 7.3 for background, HVBF and Higgs-strahlung processes. To avoid a over-loaded notation the subscripts $0, 1, 2$ will be used to label E_T^{miss}, the leading jet (highest p_T jet) and second leading jet respectively.

Descriptions of the input variables are given below.

$Balance$: The logarithm of the ratio of E_T^{miss} and the p_T of the leading jet

$$Balance = \log\left(\frac{E_T^{miss}}{p_{T1}}\right).$$

m_{12} : The combined invariant mass of jets 1 and 2 given by

$$m_{12} = \sqrt{(E_1 + E_2)^2 - (|\vec{p}_1| + |\vec{p}_2|)^2}.$$

η_i : The pseudorapidity of the ith jet.

$\Delta\eta_{12}$: The absolute difference between the pseudorapidities of jets 1 and 2

$$\Delta\eta_{12} = |\eta_1 - \eta_2|.$$

$\eta_1\eta_2$: The product of the pseudorapidities of jets 1 and 2.

$\Delta\phi_{ij}$: The azimuthal angle between objects i and j.

$\Delta\phi_{0,min}$: The smallest azimuthal angle between E_T^{miss} and any of the jets.

N_{jets} : The number of jets in the event, referred to as the jet multiplicity.

This variable set has been studied in [78] with regards to the the three signal regions and found to give good signal discrimination. Some physical motivations are given below.

E_T^{miss} and the leading jet p_T are not used directly since these variables are already used in the baseline selection, with a relatively high cut, and thus offer little additional signal/background discrimination in the signal regions. However, the ratio of the two is useful for distinguishing monojet-like topologies since an ideal monojet event will have a ratio of unity. Then the log-transform of this distribution is used which allows the classifier to better distinguish differences in the tails of the distribution.

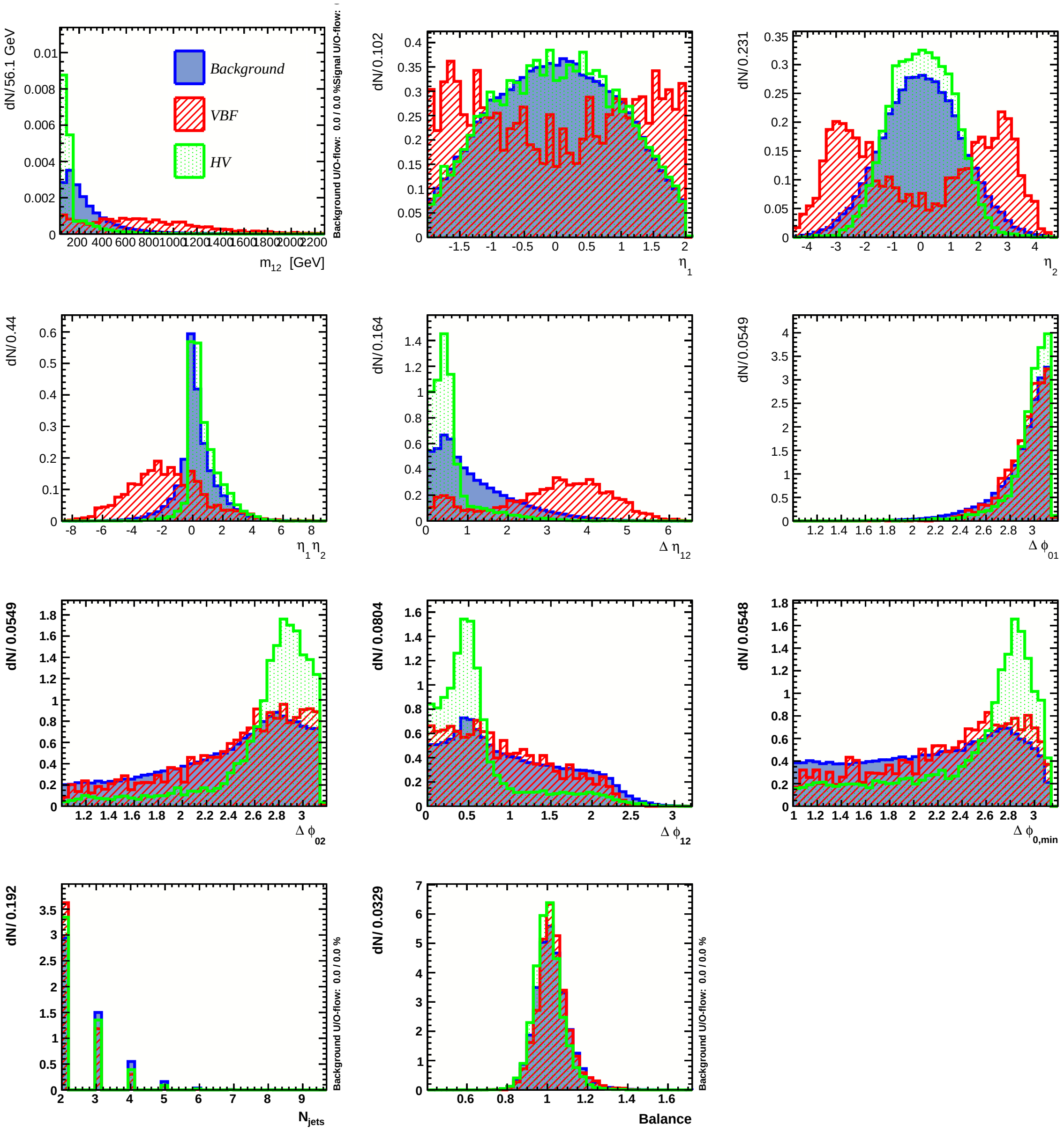
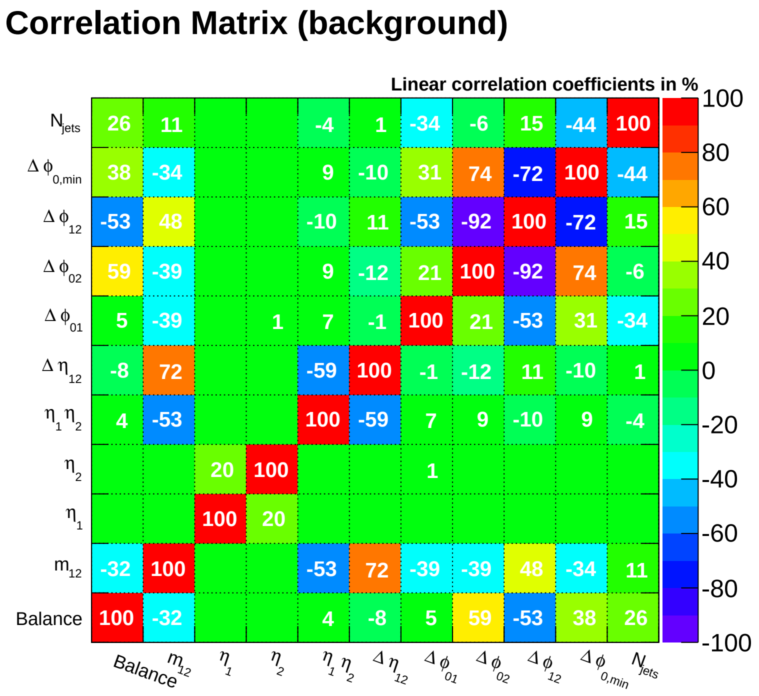


Figure 7.3: Discriminating variables used as input to the Random Forest classifiers with HVBF and Higgs-strahlung as signal classes. All distributions have been normalized to unit area. Background contains the weighted sum of the Z+jets, W+jets, Diboson and Top processes.



	Balance	m_{12}	η_1	η_2	$\eta_1\eta_2$	$\Delta\eta_{12}$	$\Delta\phi_{01}$	$\Delta\phi_{02}$	$\Delta\phi_{12}$	$\Delta\phi_{0,min}$	N_{jets}
N_{jets}	26	11			-4	1	-34	-6	15	-44	100
$\Delta\phi_{0,min}$	38	-34			9	-10	31	74	-72	100	-44
$\Delta\phi_{12}$	-53	48			-10	11	-53	-92	100	-72	15
$\Delta\phi_{02}$	59	-39			9	-12	21	100	-92	74	-6
$\Delta\phi_{01}$	5	-39		1	7	-1	100	21	-53	31	-34
$\Delta\eta_{12}$	-8	72			-59	100	-1	-12	11	-10	1
$\eta_1\eta_2$	4	-53			100	-59	7	9	-10	9	-4
η_2			20	100			1				
η_1			100	20							
m_{12}	-32	100			-53	72	-39	-39	48	-34	11
Balance	100	-32			4	-8	5	59	-53	38	26

Figure 7.4: Correlation matrix for input variables calculated with the background training set.

The unique dynamics of HVBF suggest certain discriminating features that may distinguish it from background and the other signal processes. For example the second-leading jet tends to have a large pseudorapidity as it carries on in the forward/backward direction. As can be seen in figure 7.3 this gives the second-jet pseudorapidity η_2 a unique bimodal distribution. Since the jets tend to have opposite directions along the beamline the $\eta_1\eta_2$ distribution is skewed towards negative values while $\Delta\eta_{12}$ takes on larger positive values compared to the background and HV distributions. Additionally, the m_{12} spectrum is much more uniformly distributed due to virtuality of the W/Z boson radiated from the initial state quarks.

In contrast the HV $\Delta\eta_{12}$ and $\Delta\phi_{ij}$ distributions are much more peaked than background and HVBF. This feature can be attributed to the boosted W/Z boson which then decays hadronically resulting in two collinear jets. In addition, since the jets result from on-shell W/Z bosons, the m_{12} distribution is highly peaked around 100 GeV.

Ideally all of the input variables would have little to no correlation since corre-

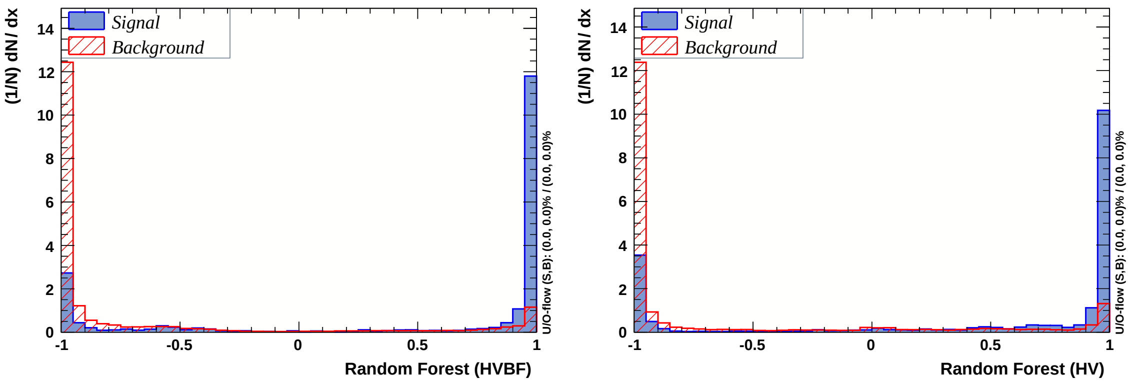


Figure 7.5: Comparison of output distribution of the Random Forest classifier for background and signal.

lated variables indicate a redundancy in the information provided to the classifier[4]. Nonetheless, large correlations between some variables are expected due to their geometric properties. Figure 7.4 shows the correlation matrix of the input variables, as calculated from the background training sample. Unsurprisingly the largest correlations are found between the $\Delta\phi_{ij}$ variables. The large correlation between m_{12} and $\Delta\phi_{12}$ is somewhat unexpected, but may be caused by glancing-collision events where the jets carry a high momentum in the forward and backward directions. It should be noted that the correlation matrix does not measure non-linear interactions between variables—which may explain why η_i appear to be independent of $\eta_1\eta_2$ and $\Delta\eta_{12}$ when they are clearly related.

7.2.4 Classifier response

Using the input variables two independent Random Forest classification models were trained for HVBF and Higgs-strahlung signal events against the weighted sum of the background processes Z+jets, W+jets, Diboson and $t\bar{t}$. Unlike the HVBF and Higgs-strahlung processes, ggF does not contain structure in the jet dynamics that can be exploited to build a classifier. Indeed, its signature is essentially identical to that of the main background Z$\rightarrow \nu\nu +$ jets making it very difficult to discriminate against. For this reason ggF signal will be left out of the classification models.

---

[4] Highly correlated variables can also lead to collinear instabilities in the weight parameters when training linear models—which can lead to large uncertainties associated with variable rankings. However, this should not be an issue for Random Forest classifiers since they are neither linear nor do they use weight parameters on the input variables.

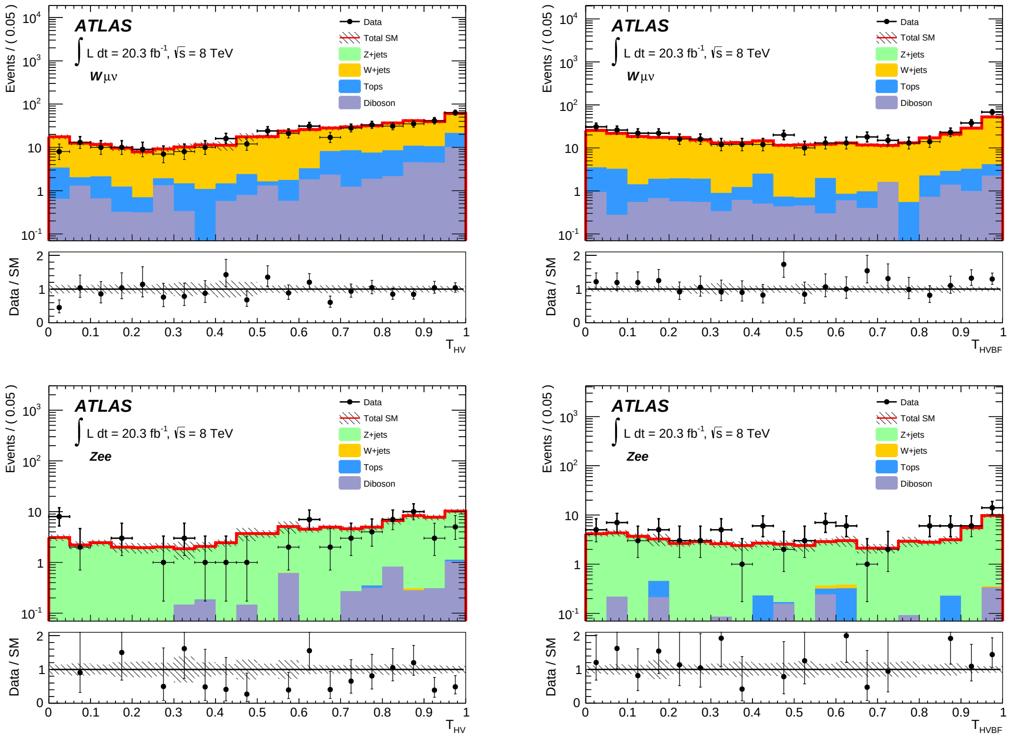


Figure 7.6: Random Forest outputs for HV (left) and HVBF (right) classifiers shown in the $W\mu\nu$ (top) and Zee (bottom) control regions.

Rank	**HVBF** Variable	score [%]	**HV** Variable	score [%]
1	$\Delta\eta_{12}$	32.4	m_{12}	29.6
2	m_{12}	23.7	$\Delta\eta_{12}$	23.8
3	$\eta_1\eta_2$	21.6	$\Delta\phi_{12}$	14.1
4	η_2	9.92	$\Delta\phi_{0,min}$	9.55
5	η_1	3.58	$\Delta\phi_{02}$	9.29
6	$\Delta\phi_{01}$	1.67	$\eta_1\eta_2$	4.06
7	$\Delta\phi_{12}$	1.66	$\Delta\phi_{01}$	2.54
8	N_{jets}	1.62	*Balance*	2.45
9	$\Delta\phi_{0,min}$	1.44	N_{jets}	2.03
10	*Balance*	1.24	η_1	1.34
11	$\Delta\phi_{02}$	1.23	η_2	1.27

Table 7.2: Variable ranking according to importance in the Random Forest algorithm.

All training and evaluation was done using the `ROOT` multivariate analysis and machine learning library `TMVA` [97]. Samples were filtered for 2 or more jets and then split into two equal sets for training and validation. For both models 3000 decision trees were trained, each grown with a maximum depth of 20. Node splits were chosen from one of 4 randomly selected variables using a 20 point grid search for the best cut value. The exact `TMVA` options and hyperparameter settings can be found in appendix D.

Figure 7.5 shows the output distributions of the Random Forest classifier models. The two distributions are normalized to unit area and scaled to cover the range $(-1, 1)$ for easy comparison. The distributions are sharply peaked for background and signal, with most of the events falling in either the first or last bin. Bins between these two peaks contain roughly the same amount of signal events and background events, indicating that the classifiers does not carry any class information in these regions.

Table 7.2 shows the Random Forest variable importance rankings for the HVBF and HV classification models. For both classifiers m_{12} and $\Delta\eta_{12}$ can be seen to contain the most discrimination power. As expected the pseudorapidity-derived variables η_i and $\eta_1\eta_2$ play an important role in the HVBF classifier while the $\Delta\phi_{ij}$ variables tend to be more important for the HV classifier. In both cases N_{jets} and *Balance* are ranked quite low, but still seem to offer some discrimination against background.

Chapter 8

Invisible Higgs results

This chapter presents the results of the model dependent search for invisible decays of the Higgs boson. In section 8.1 signal regions are defined for each of the three production mechanisms (HggF, HVBF and HV) based on the responses of the T_{HVBF} and T_{HV} Random Forest classifiers trained in section 7.2.

In section 8.2 the likelihood model is extended to include the invisible Higgs signals. The three additional signal regions are combined with the four control regions in the likelihood, for a total of seven fit regions. The discriminating variables E_T^{miss}, T_{HVBF} and T_{HV} are used for the HggF, HVBF and HV signal regions respectively, for a binned likelihood fit. A likelihood ratio test statistic is built with regards to the invisible Higgs alternative hypothesis and the CL_s upper limit setting procedure is discussed.

Section 8.3 presents and discusses the results of the model dependent fit. Table 8.1 shows the observed and expected events yields for the invisible Higgs signal regions with fitted background contributions. The 95% CL upper limit results are presented in terms of the parameter of interest μ_{sig} which is interpreted as the upper limit on the branching ratio of the Higgs boson decaying to invisible particles.

8.1 Signal Regions

As discussed in section 7.1 the three dominant Higgs production mechanisms lead to unique detector signatures in the final state. Thus the model independent signal region described in section 4.3 can be split up into three orthogonal signal regions, one for each production mechanism: HggF, HV and HVBF. To achieve this, additional selection is applied on top of the baseline selection defined in section 4.3.2.

In section 7.2 two Random Forest classifiers are trained to distinguish HVBF and HV events from background. The two correlated jets present in HVBF and HV allow for such discrimination. However, in the case of HggF no such correlation between jets exists. In fact, due to the similarity of the HggF signature to the Z($\rightarrow 4\nu$) + jets irreducible background additional selection does little—if anything—to improve the purity of the signal. Nonetheless, the HggF signal region can be used for events that pass the baseline selection but yet do not fulfill the selection requirements for the HVBF or HV signal regions. This includes all events with only a single jet and those that have negative T_{HVBF} and T_{HV} values.

These considerations motivate the following signal region selection criteria:

HggF: $N_{jets} = 1$ or ($T_{HVBF} \leq 0$ and $T_{HV} \leq 0$)

HV: $N_{jets} \geq 2$ and $T_{HV} \geq T_{HVBF} > 0$

HVBF: $N_{jets} \geq 2$ and $T_{HVBF} > T_{HV} > 0$

Typically in HEP analyses the output of a classifier is used as a final cut variable, where the cut is chosen to maximize the signal-to-background ratio. In cut-based analyses this is the only sensible approach since distributions are integrated over and thus all shape information is lost. However, if the likelihood incorporates these distributions directly one should be able to achieve better results by using all of the data since each event is weighted by the likelihood function.

In practice, `HistFitter` converts the fitted pdfs for each process to histograms to form a binned likelihood function that is used for limit setting. Thus the contribution of a given event to the likelihood is weighted by its probability of falling in the observed bin. Therefore events that fall into bins with a low expected bin count will contribute more discrimination power than those that fall into ones with high expected bin counts. When all events are taken together the likelihood function is effectively comparing the shapes of the distributions in addition to the total event counts.

This technique is most effective when shapes of the signal and background distributions are dissimilar. One way to interpret what a classifier does is that it maps a high-dimensional space (the input variables) down to a one-dimensional space (the Random Forest output) in such a way as to maximize separation between signal and background. This forces the classifier to push most of the probability mass of signal and background pdfs to opposite ends of the output spectrum, resulting in distributions such as seen in figure 7.5. This consideration motivates the use of the Random Forest classifiers as discriminating variables in the signal regions.

8.2 Limit setting procedure

In section 6.2 a likelihood test statistic is constructed to test the background-only hypothesis against a general alternative hypothesis (background plus some unknown signal). This section develops in a similar manner, except now the background-only hypothesis is to be tested against an excess of Higgs boson events that decay invisibly. By tailoring the test statistic to a specific model one can achieve greater discriminating power than with q_0. This can be achieved by comparing the signal and background distribution shapes (or equivalently histogram bin counts). Such is possible if one knows—or has a good estimate of—the signal distribution. In HEP this type of analysis in known as a *model dependent search* and a *shape analysis*.

8.2.1 Likelihood model

To construct a test statistic to test for an excess of invisible Higgs events the likelihood model must be extended to accommodate the three new signal regions. The same likelihood template from eq. 5.8 can be used where the regions index r now extends over the four control regions and three signal regions. The pdfs $f_r(X)$ in the shape factor can now be specified as the sum of the signal and background contribution in a given region. Now the total pdf of discriminating variable X in a given signal region is

$$f(X) = \frac{\mu_{sig}\nu_{sig}f_{sig}(X) + \nu_{bkg}f_{bkg}(X)}{\mu_{sig}\nu_{sig} + \nu_{bkg}}$$

where

$$\nu_{sig}f_{sig}(X) = \nu_{HV}f_{HV}(X) + \nu_{HVBF}f_{HVBF}(X) + \nu_{HggF}f_{HggF}(X).$$

Here $f_{bkg}(X)$ is the weighted sum of the background pdfs with ν_{bkg} as the expected background event count and ν_{sig} is the total expected signal count with $Br(h \to inv.) = 1$. Note that all three signal distributions are scaled by the common factor μ_{sig}.

As in section 6.2 μ_{sig} is the POI—however μ_{sig} takes on a very different interpretation here. Since μ_{sig} scales the signal rate it can be interpreted as the branching fraction of the Higgs boson decaying to invisible particles $Br(h \to inv.)$. This branching fraction places direct limits on Higgs-portal models as discussed in sec 8.4.

8.2.2 Test statistic

In section 6.2.2 the background-only test statistic q_0 is derived by taking the logarithm of the likelihood ratio $\lambda(\mu)$ (eq. 6.1) and accounting for negative fluctuations. The model-dependent test statistic q_μ can be derived in the same fashion. The test statistic q_μ is used to differentiate the hypothesis of signal events being produced at a rate μ from the alternative hypothesis of a lesser rate $\mu' < \mu$.

$$q_\mu = \begin{cases} -2\log\lambda(\mu_{sig}) & : \hat{\mu}_{sig} \leq \mu_{sig} \\ 0 & : \hat{\mu}_{sig} > \mu_{sig} \end{cases} .$$

Note that q_μ is a test statistic for a one-sided alternative, appropriate for upper limits.

Now the p-value p_{s+b} can be defined as

$$p_{s+b} = \int_{q_\mu^{obs}}^{\infty} f(q_\mu|\mu_{sig})dq_\mu \tag{8.1}$$

where the asymptotic distributions for $f(q_\mu|\mu_{sig})$ is not as simple as eq. 6.2 but can still be expressed in closed form [81].

8.2.3 The CL_s method

In HEP it is conventional to use the CL_s method for setting upper limits on model parameters that only take non-negative values, rather than the more common confidence interval procedure followed in section 6.2. The CL_s method was fist introduced by physicists working at the LEP experiments at CERN and has since been used in numerous publications from the Tevatron and LHC experiments [98].

Upper limits set with the CL_s method differ from confidence intervals in that the standard confidence level of the interval is not necessarily equal to its coverage probability α. The CL_s method is motivated by the fact that for downward fluctuations the upper limit of the confidence interval can be arbitrarily small, even zero. This feature is undesirable since one should not claim sensitivity to an arbitrarily small signal rate. The CL_s method confronts this problem by scaling the p-value obtained from the standard confidence interval procedure p_{s+b} with one over the p-value obtained under the background-only hypothesis p_b [99]:

$$p_s = \frac{p_{s+b}}{p_b}. \tag{8.2}$$

With p_s the upper limit μ_{up} can be extracted in the usual way, by solving $p_s = \alpha$.

Although limits produced with the CL_s method tend to have a higher coverage probability than α, a value of μ is regarded as excluded at the 95% confidence level ($\alpha = 0.05$) if $\mu < \mu_{up}$.

In the asymptotic limit of large sample size the CL_s median upper limit and μ_{up+N} bands are given by [81]

$$\mu_{up+N} = \sigma_{\hat{\mu}_{sig}} \Phi^{-1}(1 - \alpha\Phi(N)) + \sigma_{\hat{\mu}_{sig}} N \tag{8.3}$$

where $\sigma_{\hat{\mu}_{sig}}$ is the standard deviation of $\hat{\mu}_{sig}$ which can be estimated with the appropriate element of V given by 5.10. For the median upper limit $N = 0$ yielding the simple result

$$\mu_{up}^{med} = \sigma_{\hat{\mu}_{sig}} \Phi^{-1}(1 - 0.5\alpha). \tag{8.4}$$

8.3 Fit results

Using the control region specific CMLEs from the background-only fit as in section 6.2 the fitted backgrounds can be calculated and compared to observed values. Table 8.1 shows the results of the background-only fit in the invisible Higgs signal regions. The HggF signal region contains the most events with 6432 observed and 6305.44 ± 154.43 expected. The HVBF and HV signal regions have comparable yields with 1462 observed with 1404 ± 63.04 expected and 1031 observed with 1091.48 ± 66.36 expected respectively. Although HggF and HVBF contain slightly more events than expected, and HV slightly fewer, all observed yields fall within uncertainties.

The distributions for E_T^{miss}, T_{HVBF} and T_{HV} in their respective signal regions are show in figure 8.3. Data and MC are in good agreement, with no significant deviations in any of the bins.

The model dependent fit is performed over the four control regions and three signal regions simultaneously. Once the fit is performed the best-fit values can then

Region	HggF	HVBF	HV
Observed events	6432	1462	1031
Fitted bkg events	6305.4 ± 154.4	1404.9 ± 63.0	1091.5 ± 66.4
Fitted W events	1422.5 ± 73.8	324.3 ± 16.5	244.8 ± 13.4
Fitted Z events	4596.0 ± 145.6	971.4 ± 35.8	791.0 ± 45.4
Fitted Top events	56.2 ± 26.5	11.1 ± 5.8	22.8 ± 13.4
Fitted Diboson events	215.0 ± 121.5	98.1 ± 54.7	32.9 ± 16.3
Fitted Multijet events	$15.8^{+15.9}_{-15.8}$	0.0 ± 0.0	0.0 ± 0.0
MC exp. SM events	7274.9	1620.9	1259.3
MC exp. W events	1641.2	374.1	282.4
MC exp. Z events	5302.6	1120.7	912.6
MC exp. Top events	64.8	12.8	26.3
MC exp. Diboson events	248.0	113.1	37.9
MC exp. Multijet events	18.2	0.0	0.0

Table 8.1: Fit results for the control and signal regions for an integrated luminosity of 20.3fb^{-1}. The results are obtained from the control regions using the background-only fit. Nominal MC expectations (normalized to MC cross-sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties, which are added in quadrature. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

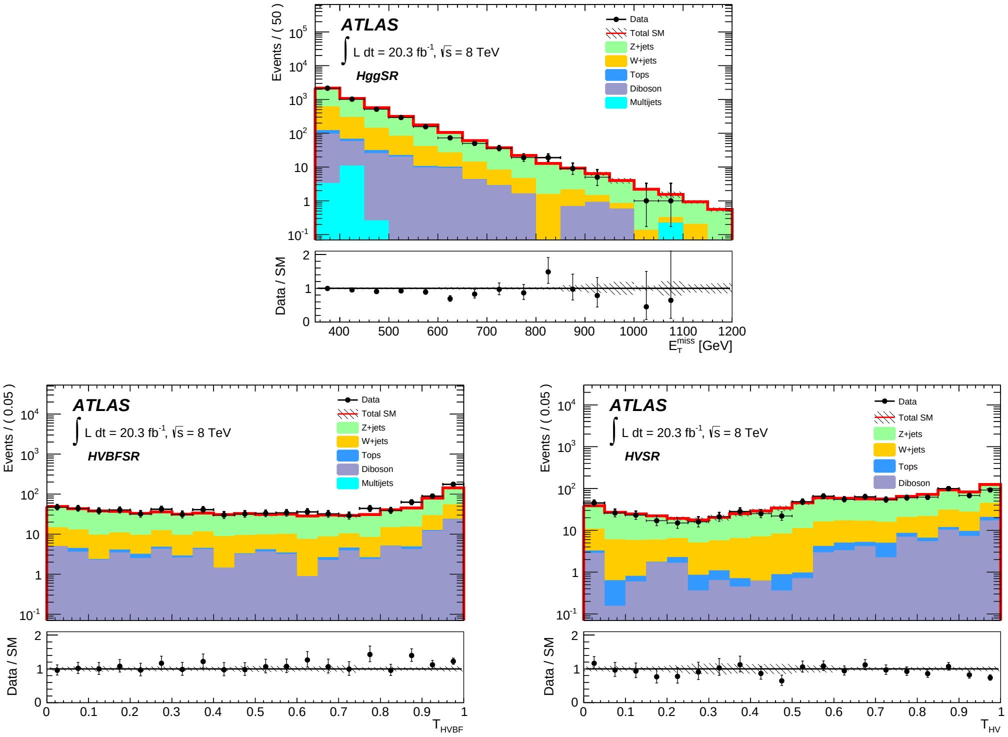


Figure 8.1: Invisible Higgs signal region distributions for E_T^{miss} in the HggF (top), T_{HVBF} for HVBF signal region (bottom left) and T_{HV} for HV signal region (bottom right).

CL	μ_{up}^{obs}	μ_{up}^{med}	$\mu_{up}^{med} \pm \sigma$	$\mu_{up}^{med} \pm 2\sigma$
90%	0.66	0.58	$^{+0.91}_{-0.42}$	$^{+1.4}_{-0.61}$
95%	0.84	0.73	$^{+1.1}_{-0.51}$	$^{+1.6}_{-0.71}$
98%	1.0	0.95	$^{+1.5}_{-0.74}$	$^{+2.0}_{-1.3}$

Table 8.2: Invisible Higgs branching ratio upper limits.

used to calculate the test statistic q_{μ}^{obs} along with its median, $\pm\sigma$ and $\pm 2\sigma$ values. `HistFitter` uses a grid-search approach in finding the μ_{up} values at a given CL by calculating the p_s values for 20 grid points of $\hat{\mu}_{sig}$ and interpolating the results. Table 8.2 shows the upper limit results at the 90%, 95% and 98% CL. The table shows that values of μ_{sig}, which is interpreted as the branching ratio of invisible Higgs decays, greater 84.3% are excluded at the 95% CL.

8.4 Higgs-Portal interpretation

Results from the previous section can be interpreted in terms of the Higgs-portal model discussed in section 1.4.1. If the dark matter particles are light enough, $M_{\chi} \leq \frac{1}{2} m_h$, the Standard Model Higgs boson may decay into WIMP pairs $h \to \chi\chi$. The upper limits on the branching ratio to invisible particles define the maximum allowed decay width to the dark matter particles $\Gamma_{h \to \chi\chi} = Br(h \to inv) \times \Gamma_h$. The three different scenarios for the dark matter fields are considered: real scalar S, vectors V and Majorana fermions f. Decay widths can be calculated from equations 1.13–1.15. Adopting the equations from Ref. [8] the decay widths for the three scenarios are

$$\begin{aligned}
\Gamma_{h \to SS}^{inv} &= \frac{\lambda_S^2 \nu^2 \beta_S}{64\pi m_h} \,, \\
\Gamma_{h \to VV}^{inv} &= \frac{\lambda_V^2 \nu^2 m_h^3 \beta_V}{256\pi M_V^4} \left(1 - 4\frac{M_V^2}{m_h^2} + 12\frac{M_V^4}{m_h^4}\right) , \\
\Gamma_{h \to \chi\chi}^{inv} &= \frac{\lambda_F^2 \nu^2 m_h \beta_F^3}{32\pi \Lambda^2} \,, \qquad (8.5)
\end{aligned}$$

where $\beta_X = \sqrt{1 - 4M_X^2/m_h^2}$, λ_X are WIMP-Higgs couplings, Λ is the energy cut-off scale for the fermion scenario and ν is the vacuum expectation value. It should be noted that the fermion model is non-renormalizable and that in constructing a UV complete theory one may have to introduce additional fields and dynamics that could significantly alter the limits quoted here [100]. Nonetheless, the model is included for completeness.

Since some of the most stringent limits on dark matter come from direct detection experiments it is desirable to compare the finding of this analysis with those of the leading direct detection experiments. This is possible since in the Higgs-portal model WIMPs can interact elastically with nuclei through the exchange of a Higgs boson. The resulting nuclear recoil is then interpreted in terms of the WIMP mass and WIMP-nucleon scattering cross-section. Cross-sections can be derived using the effective field theory equations 1.17–1.19. Adopting the results from Refs. [8] and [101] the spin-independent WIMP-nucleon interactions can be expressed as

$$\begin{aligned}
\sigma^{SI}_{S-N} &= \frac{\lambda_S^2}{16\pi m_h^4}\frac{m_N^4 f_N^2}{(M_S+m_N)^2}\ ,\\
\sigma^{SI}_{V-N} &= \frac{\lambda_V^2}{16\pi m_h^4}\frac{m_N^4 f_N^2}{(M_V+m_N)^2}\ ,\\
\sigma^{SI}_{\chi-N} &= \frac{\lambda_F^2}{4\pi\Lambda^2 m_h^4}\frac{m_N^4 M_F^2 f_N^2}{(M_F+m_N)^2}\ ,
\end{aligned} \tag{8.6}$$

where m_N is the nucleon mass and f_N parameterizes the Higgs-nucleon coupling. Since the WIMP-nucleon cross-sections are proportional the Higgs-WIMP couplings they too depend on the invisible Higgs decay width. Thus with equations 8.5 and 8.6 the invisible Higgs branching ratio can be placed in the same phase space as the direct detection limits via the Higgs-portal model.

Figure 8.2 shows the 95% CL limit from section 8.3 with some of the leading results from the direct detection experiments XENON100 [24], CRESST-II [25] and CoGeNT [26]. Also shown is the expected neutrino background that ultimately limits the sensitivity of direct detection experiments [102]. For all three scenarios the Higgs-portal limits are stronger than the exclusion limits by the direct detection experiments for WIMP masses less than about 10 GeV. This is expected as the LHC has no limitations for the production of low mass particles, whereas the recoil energies produced in the interactions of WIMPs with nuclei in direct detection experiments are below the sensitivity threshold for light WIMP masses. However, for the heavier mass range where $m_\chi > \frac{1}{2}m_h$ decay of the Higgs boson is kinematically forbidden, thus only direct detection experiments are sensitive in this range.

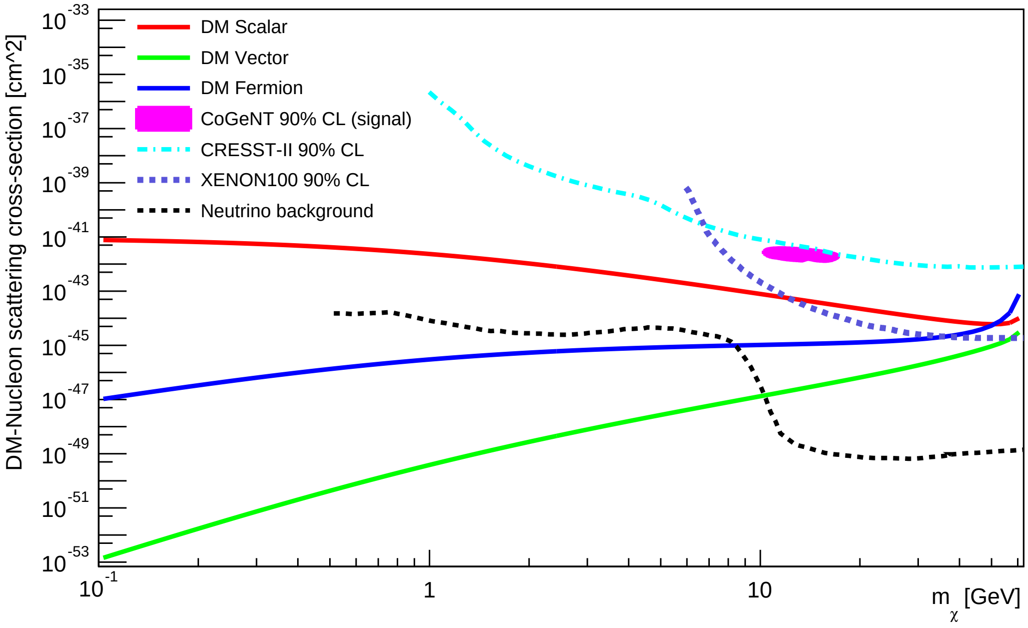


Figure 8.2: Spin-independent DM-nucleon cross-section as a function of Dark Matter mass. Results from the direct detection experiments XENON100 [24], CoGeNT [26], CRESST-II [25] and expected neutrino direct-detection background [102] are compared to the results from the $Br(h \rightarrow inv.)$ 95% CL upper limits in the Higgs-portal scenario.

Chapter 9

Conclusion

9.1 Conclusion

A search for new phenomena in events with a monojet signature is conducted at $\sqrt{s} = 8$ TeV with the ATLAS 2012 data set. Results are presented in term of model-independent limits as well as a limit on the invisible Higgs branching ratio. No significant deviations from Standard Model expectations are observed.

Expected signal and background levels in the signal regions are determined from Monte Carlo predictions using a background-only fit. The total event counts of the dominant background processes W/Z + jets are estimated using four control regions with similar signatures. The results of the event yields for the four control regions and total signal regions for the background-only fit are shown in table 6.2, and those for the three invisible Higgs signal regions in table 8.1. In all signal regions good agreement is seem between data and the background-only predictions.

Systematic uncertainties are included as Gaussian-constrained nuisance parameters, which are shared among the signal and control regions so that their correlated effects are taken into account. The fitted effects of the systematics in the control and signal regions are shown in table 6.3 with their correlations shown in figure 6.3.

Limits on the model-independent effective cross-section and invisible Higgs branching ratio are computed using two independent likelihood test statistics. A model-independent test statistic based on the background-only hypothesis places a 95% CL upper limit on the effective cross-section of 16.61 fb corresponding to an upper limit of 337 observed signal events. A second test statistic based on an invisible Higgs alternative hypotheses and a Random Forest multivariate classifier places a 95% CL observed upper limit of 0.84 on $Br(h \rightarrow inv.)$.

The observed limits on the invisible branching fraction of the Higgs boson imply a constraint on the WIMP-nucleon scattering cross-sections mediated by a Higgs boson through the effective field theory operators in section 1.4.2. WIMP-nucleon scattering limits in terms of three Higgs-portal models are shown and compared with leading direct detection experimental limits in figure 8.2. LHC and direct detection limits are shown to be complementary with LHC limits most sensitive for $m_{DM} \leq 10$ GeV and direct detection experiments most sensitive in the range $m_{DM} > 60$ GeV.

Appendix A

Author's contribution to Monojet paper

This section briefly summarizes the author's contributions to Ref. [103]. The background estimate strategy adopted by the monojet group differs significantly from the one presented in this dissertation. Below briefly describes the background estimate approach used in the 2012 monojet paper [78] and some of the author's results.

A.1 W/Z + jets background estimate

The data-driven W/Z + jets background estimate relies on the assumption of the relationship

$$\frac{N^{Data}_{SR,p}}{N^{MC}_{SR,p}} = \frac{N^{Data}_{CR,p}}{N^{MC}_{CR,p}}, \tag{A.1}$$

where N denotes the number of events for a given process p (such as Z$(\rightarrow \nu\nu)+$jets, etc.) in either a signal (SR) or control (CR) region derived from data or Monte Carlo (MC). Here N can be interpreted as the number of events in a bin of a given variable (e.g. E_T^{miss}), as would be required to compare distribution shapes, or simply as the total number of events, i.e the sum off all bins. If $N^{MC}_{SR,p}$ and $N^{MC}_{CR,p}$ both come from the same Monte Carlo generator, as is the case in the monojet analysis, one would expect this ratio to hold true. Thus the number of events in a given signal region for a given process can be estimated by

$$N^{est.}_{SR,p} = N^{Data}_{CR,p} \times \left(\frac{N^{MC}_{SR,p}}{N^{MC}_{CR,p}} \right). \tag{A.2}$$

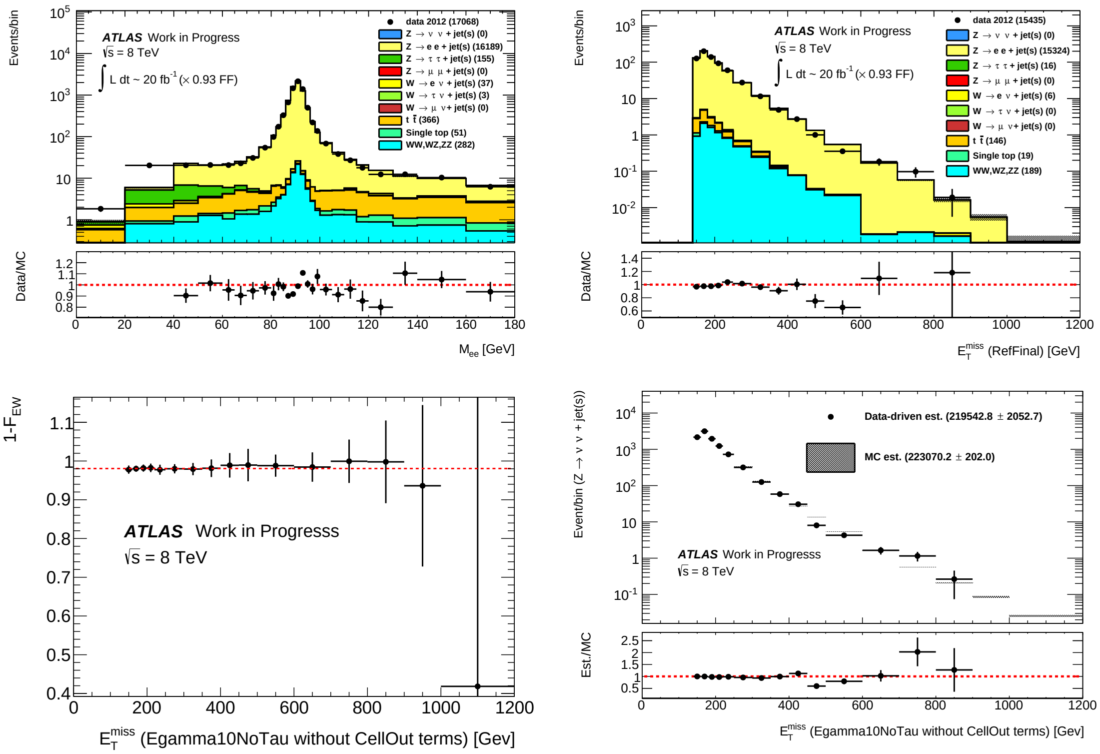


Figure A.1: (top left) Invariant mass distribution of dielectrons in the Z($\rightarrow ee$) + jets control region with Z invariant mass window cut removed. (top right) E_T^{miss} distribution with good electrons removed in the Z($\rightarrow ee$)+jets control region. (bottom left) $1 - F_{EW}$ factor, i.e. efficiency of Z($\rightarrow ee$) + jets process in the Z($\rightarrow ee$) + jets control region as a function of E_T^{miss}. Red dashed line corresponds to best fit value of $1 - F_{EW}$, which evaluates to $\approx 98\%$. (bottom right) Z($\rightarrow \nu\nu$) + jets background in signal region estimate as a function of E_T^{miss}. Points correspond to data-driven estimate while hatched histograms to Monte Carlo simulations. In all plots data has been scaled by a factor of 0.93 to account for a 7% normalization offset in simulation.

However, often the process p will not have an adequate control region that can be selected in data. For example one cannot define an efficient control region for the $\mathrm{Z}(\rightarrow \nu\nu) + \mathrm{jets}$ process. Nevertheless, one can select $\mathrm{Z}(\rightarrow ee) + \mathrm{jets}$ events with $\approx 98\%$ efficiency, where the electrons can be removed from the event to mimic the $\mathrm{Z}(\rightarrow \nu\nu) + \mathrm{jets}$ signature. However, residual differences between the two selection criteria, such as different backgrounds and object identification efficiencies, must be taken into account. The list below outlines the procedure as a series of steps, where the subscript p for the process has been dropped for brevity.

1. Select data events in CR: N_{CR}^{Data}
    - Control region may be any of the four defined in section 4.4
2. Remove multijet background in CR: $(N_{CR}^{Data} - N_{CR}^{multijet})$
    - Multijet background is negligible in Z + jets control regions, but not W + jets.
3. Factor out EW backgrounds in CR: $1 - F_{EW} = \frac{N_{CR}^{MC}}{\sum^{All\ EW} N_{CR}^{MC}}$
    - Here $1 - F_{EW}$ is the efficiency of the control region process in the control region as determined by Monte Carlo.
4. Transfer from the control region to the signal region: $TF = \frac{N_{SR}^{MC}}{N_{CR}^{MC}}$
    - Here TF is usually referred to as the “transfer factor.”
    - For electron channel control regions events need to be weighted by electron trigger scale factors to account for differences in trigger, reconstruction and identification efficiency.

Finally, putting steps 1-4 together, one arrives at the equation:

$$N_{SR}^{est} = (N_{CR}^{Data} - N_{CR}^{multijet}) \times (1 - F_{EW}) \times TF. \tag{A.3}$$

By conducting background estimates using this procedure, differences in overall normalizations between data and Monte Carlo can be corrected for. In addition if the procedure is applied in each bin of the signal region E_T^{miss} distribution, differences in shape can also be accounted for.

A.2 Zee control region analysis results

During 2012-2013 the author conducted a background estimate analysis using the Z($\rightarrow$ ee) $+$ jets control region for the ATLAS monojet group. The definition of the Zee control region is very similar to that given in section 4.4.4. Figure A.1 (top left and right) shows the validation distributions for E_T^{miss} (leptons removed) and the dielectron invariant mass, which indicate that Monte Carlo systematically underestimates data (Monte Carlo as been scaled by a factor of 0.93). The bottom left plot in the figure shows the calculated, bin-wise values for the $1 - F_{EW}$ factors, with a global best fit value of 0.98. Using equation A.3 the data-driven background estimate for the Z($\rightarrow$ $\nu\nu$) $+$ jets E_T^{miss} distribution is calculated, applying bin-wise corrections. The bottom right plot of figure A.1 compares the data-driven estimate of the Z($\rightarrow$ $\nu\nu$) $+$ jets background in the signal region with the uncorrected Monte Carlo estimate.

Appendix B

Authors contribution to Missing E_T trigger group

This section briefly summarizes the main contributions the author made to the ATLAS Missing ET (MET) trigger group from 2012-2014.

B.1 On-call trigger expert and monitor duties

While based at CERN from 2012-2013 the author regularly took on on-call MET trigger expert shifts. The trigger expert shifter is the person on-call to deal with any problems with the trigger that might come up during LHC operations. In addition the expert trigger shifter's responsibilities include:

- being at, or very near CERN and available 24/7;
- being aware of any ongoing issues with regards to the MET trigger;
- attending daily run coordinator meetings and weekly JET and MET meetings;
- marking the data quality for runs covered;
- checking new trigger menus/production cache when requested and reporting findings;
- documenting any relevant information about runs covered.

B.2 Contributions to the 2011 MET performance note

During 2012 the author made substantial contributions to the 2011 MET performance note [104]. 2011 LHC beam conditions varied quite dramatically—going from an average interaction per bunch crossing of 3, up to 15 and one high luminosity run at 30. The main focus of the 2011 performance note was to study how these changes affected the MET trigger. Specifically, the author's contributions to the 2011 performance note included (figures refer to [104]):

- comparing data and MC distributions for E_T^{miss} and ΣE_T at the L1 and EF triggers levels (figure 1);
- studying the distributions of empty bunch crossing for E_T^{miss} and ΣE_T at EF trigger level (figure 2);
- studying the distributions of randomly triggered events for E_T^{miss} and ΣE_T at EF trigger level (figure 3);
- plotting the 2D distributions of $\not{E}_{x,y}$ and $\sqrt{\Sigma E_T}$ of randomly triggered events at the EF trigger level (figure 8);
- studying XS fit parameters at the L1 and EF trigger level (figures 9 & 10);
- studying XS, E_T^{miss} and ΣE_T distributions for various levels of pile-up (figures 12, 19, 20, 23, 24).

B.3 Sum-E_T pile-up suppression studies

In 2010 it had been observed that, in both data and Monte Carlo, there was a strong correlation between sum-E_T and sum-E for minimum bias (minbias) events independent of the level of pile-up. This relationship is clearly demonstrated in figure B.1. In 2010 studies were conducted within the MET trigger group that indicated that is correlation could be used to help discriminate minbias events from high-p_T multijet events. It was hoped that this relationship could be exploited to design a new trigger chain based on sum-E_T. This trigger variable was dubbed $ETXS$ and defined as:

$$ETXS = \sum E_T - f\left(\sum E\right) \tag{B.1}$$

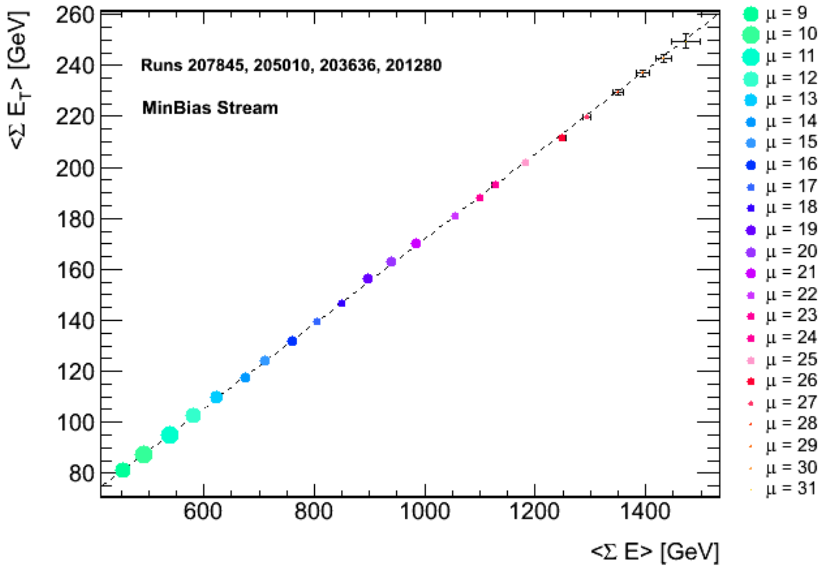


Figure B.1: Average SumEt vs average SumE. Data has been binned by average interaction per beam crosses, μ. The size of the points is proportional to the amount of data in its given μ bin.

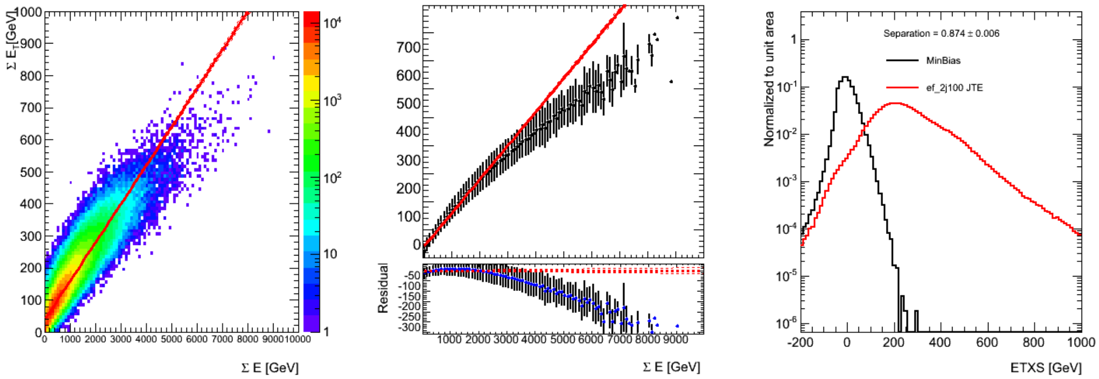


Figure B.2: (left) SumEt vs. SumE with best fit line; (middle) fit in profile with residuals; (right) ETXS distributions for minibus events and high energy multijet events for purposes of comparison.

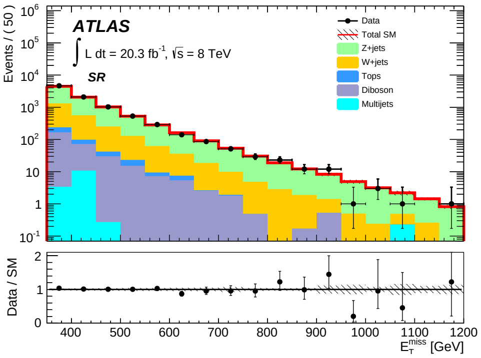
where $f(\sum E)$ is an empirical function of the expected sum-E_T contribution from minbias events.

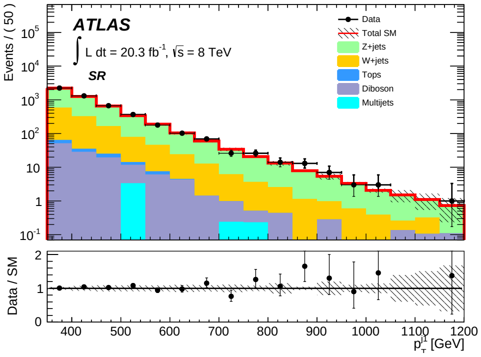
Previous studies has shown that there was a linear relationship between sum-E_T and sum-E and that the fit parameters were stable across runs. However, the author's studies based on 2011and 2012 data indicated that this linear relationship did not hold for larger values of sum-E_T and sum-E, as seen in figure B.2. Some nonlinear parametric function for f showed to have good fit results, however the fit parameters were unstable across different runs and pile-up conditions. It was also shown that, even with the proper choice of fit function and fit parameters $ETXS$ only offered marginal gains in separation between minbias events and high-p_T multijet events.

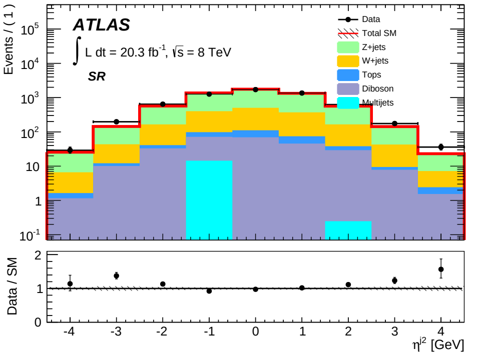
Appendix C

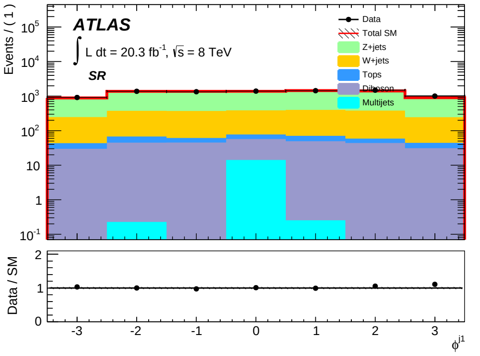
Validation histograms

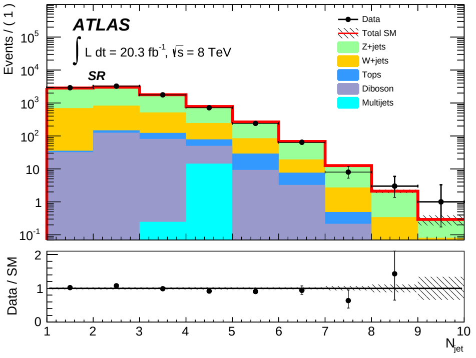
C.1 Signal region histograms

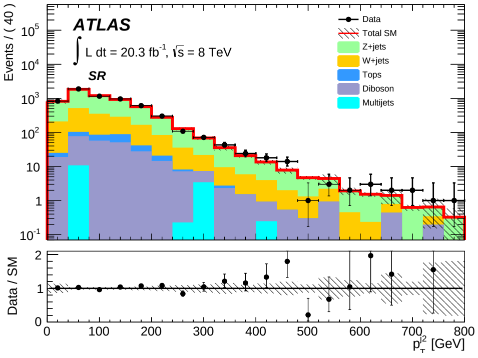


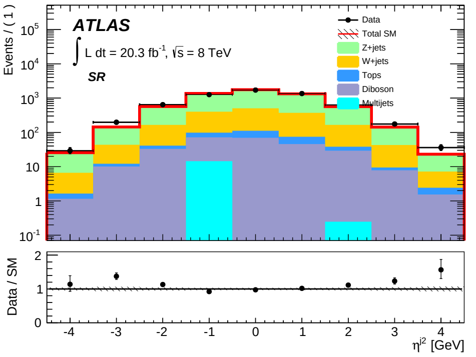


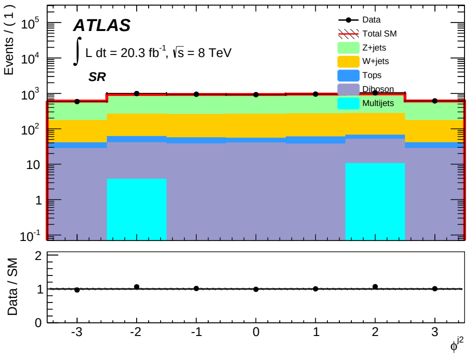


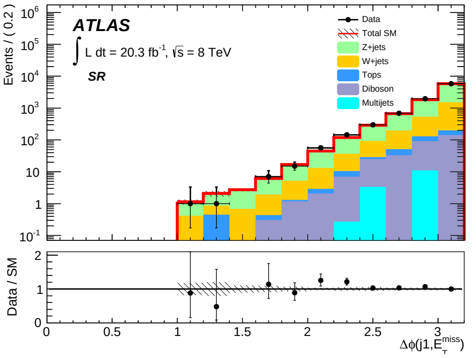


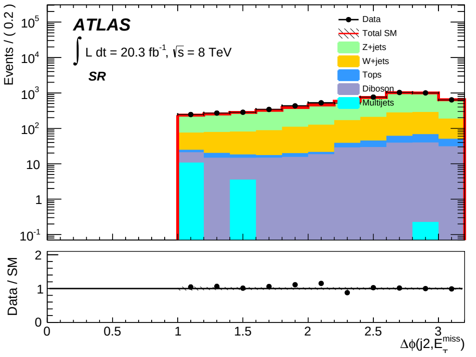

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
SR
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Multijets
Events / (1)
Data / SM
N$_{jet}$

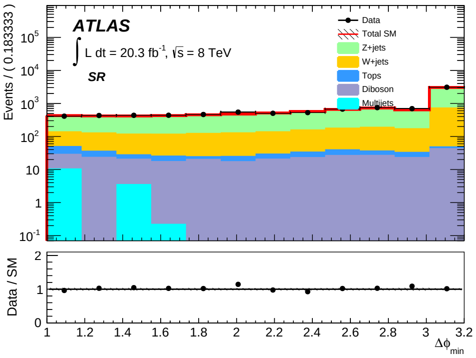

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
SR
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Multijets
Events / (40)
Data / SM
p$_T^{j2}$ [GeV]

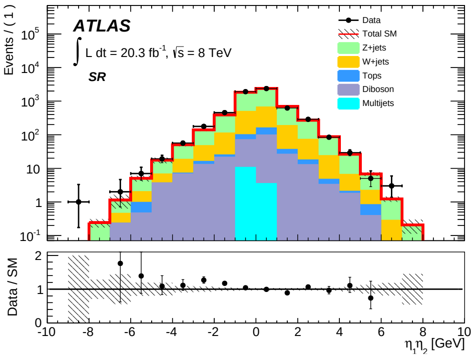

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
SR
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Multijets
Events / (1)
Data / SM
η^{j2} [GeV]

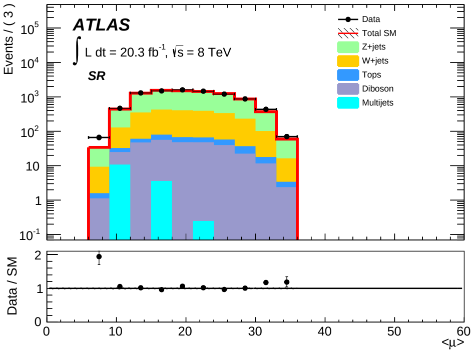

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
SR
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Multijets
Events / (1)
Data / SM
φ^{j2}

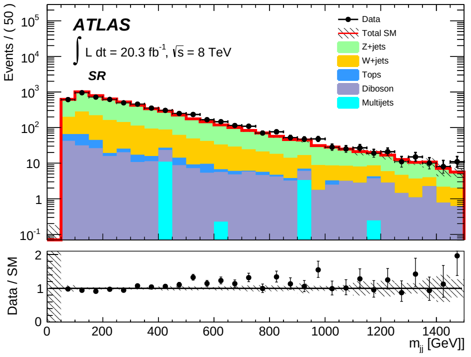

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
SR
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Multijets
Events / (0.2)
Data / SM
Δφ(j1,E$_T^{miss}$)

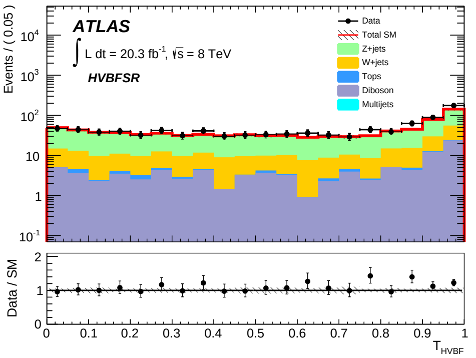

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
SR
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Multijets
Events / (0.2)
Data / SM
Δφ(j2,E$_T^{miss}$)

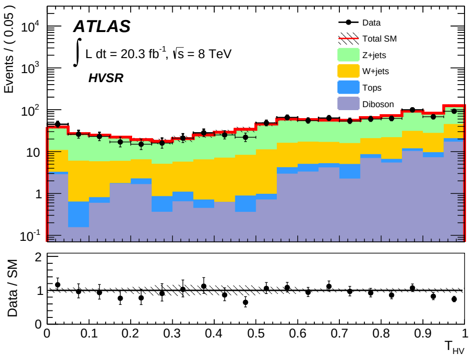

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
SR
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Multijets
Events / (0.183333)
Data / SM
Δφ$_{min}$


ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
SR
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Multijets
Events / (1)
Data / SM
η$_1$η$_2$ [GeV]

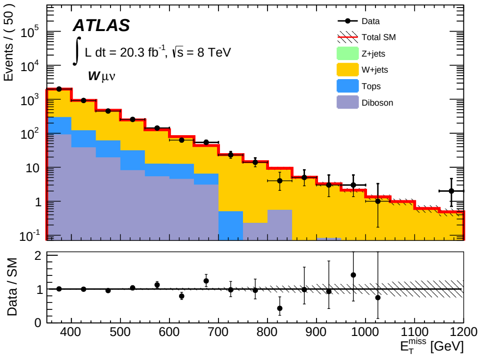


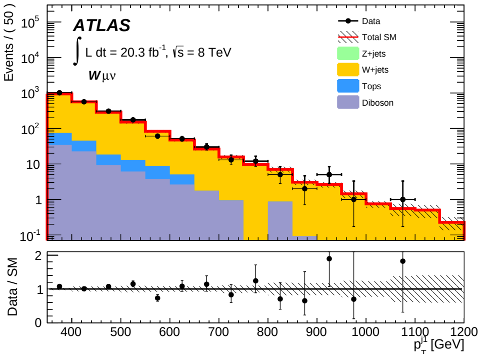


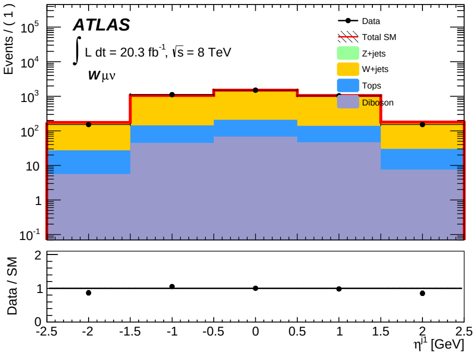


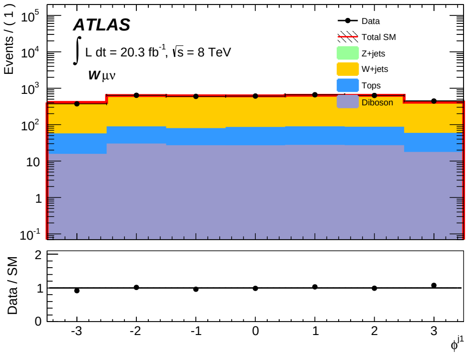


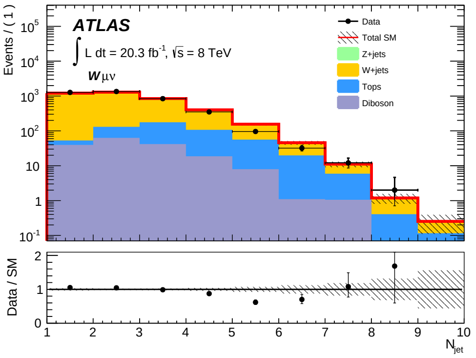
C.2 $W\mu\nu$ control region

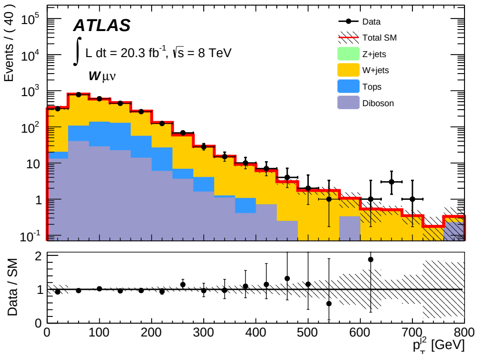


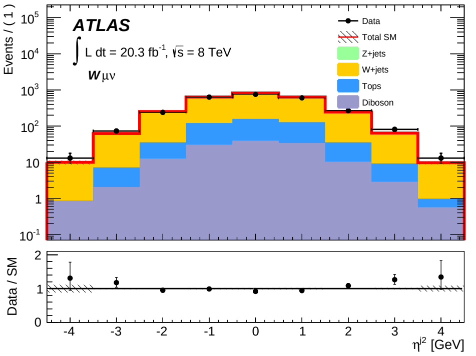


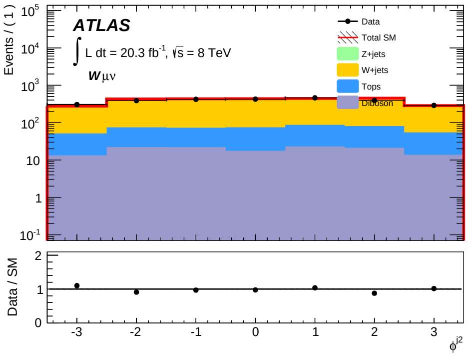

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
Data / SM
η^j1 [GeV]

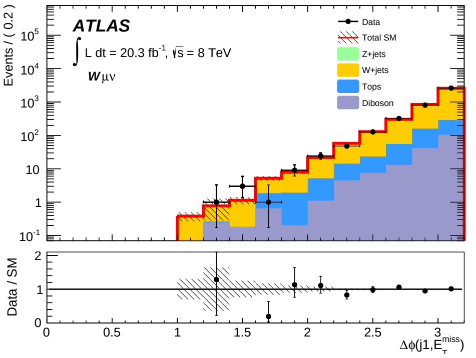

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
Data / SM
φ^j1

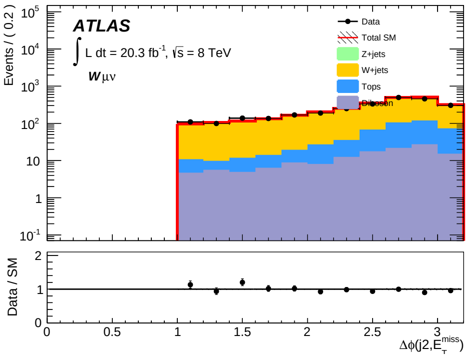

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
Data / SM
N_jet

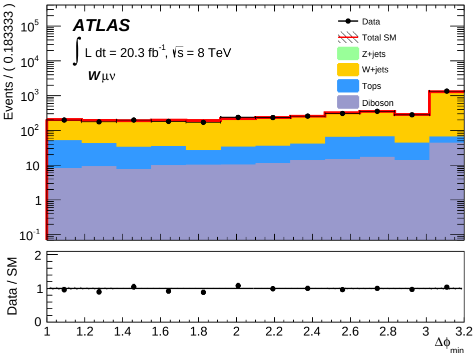

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (40)
Data / SM
p_T^j2 [GeV]

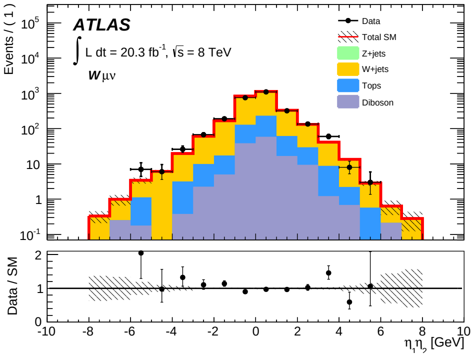

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
Data / SM
η^j2 [GeV]

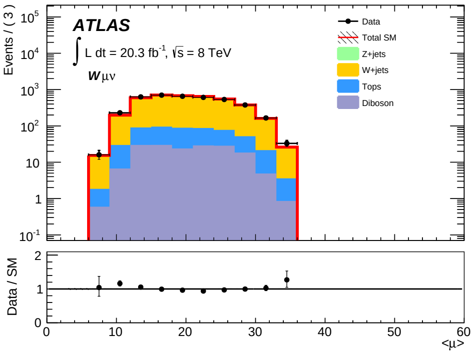

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
Data / SM
φ^j2

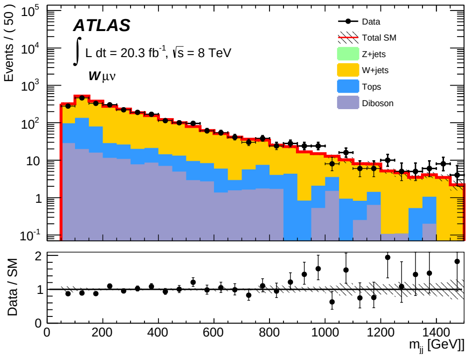

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.2)
Data / SM
Δφ(j1,E_T^miss)

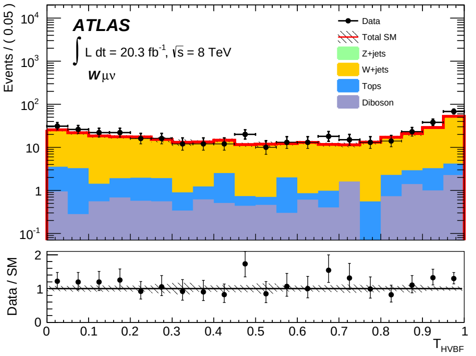

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.2)
Data / SM
Δφ(j2,E_T^miss)

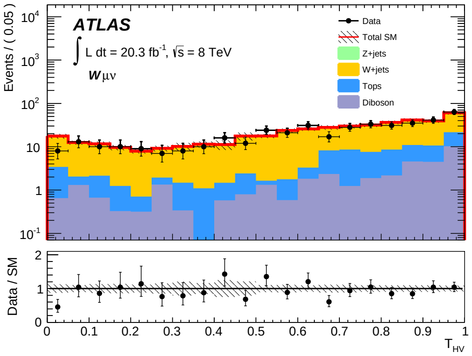

ATLAS
$\int$ L dt = 20.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.183333)
Data / SM
$\Delta\phi_{min}$
1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3 3.2


ATLAS
$\int$ L dt = 20.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
Data / SM
$\eta_1\eta_2$ [GeV]
-10 -8 -6 -4 -2 0 2 4 6 8 10

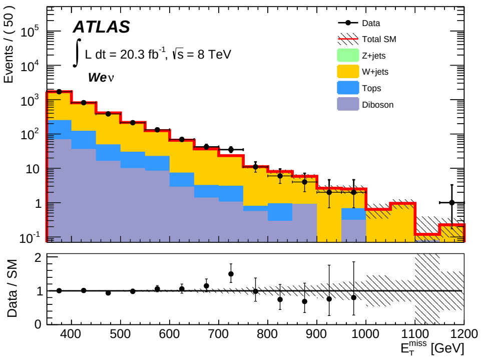

ATLAS
$\int$ L dt = 20.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (3)
Data / SM
<μ>
0 10 20 30 40 50 60

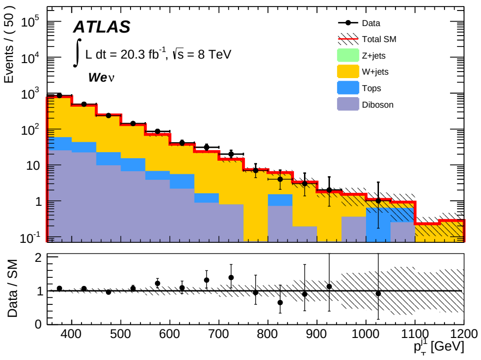

ATLAS
$\int$ L dt = 20.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (50)
Data / SM
m_{jj} [GeV]]
0 200 400 600 800 1000 1200 1400

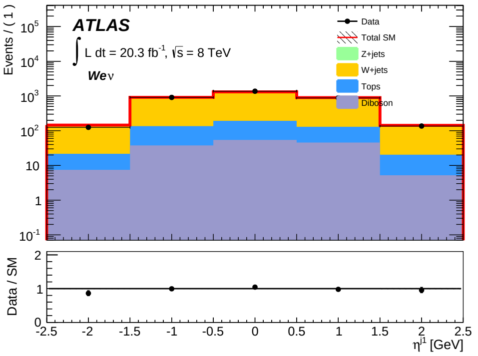

ATLAS
$\int$ L dt = 20.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.05)
Data / SM
T_{HVBF}
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

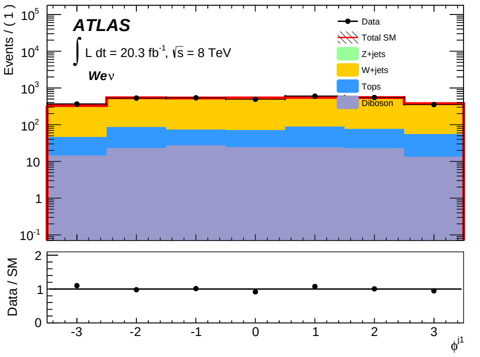

ATLAS
$\int$ L dt = 20.3 fb^{-1}, $\sqrt{s}$ = 8 TeV
Wμν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.05)
Data / SM
T_{HV}
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

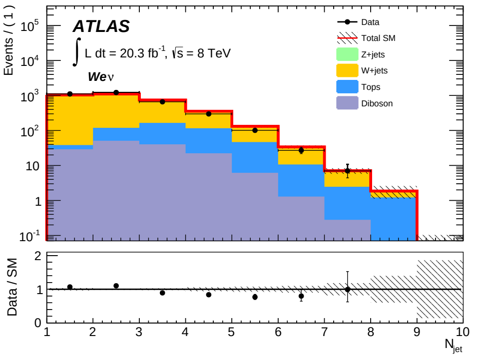
C.3 $We\nu$ control region

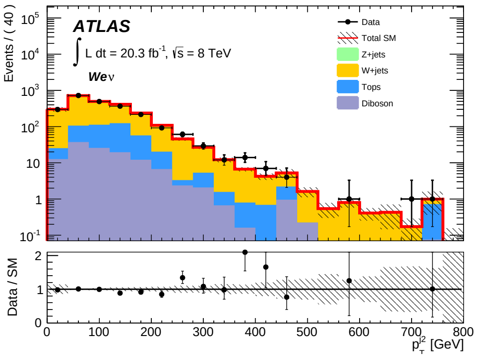


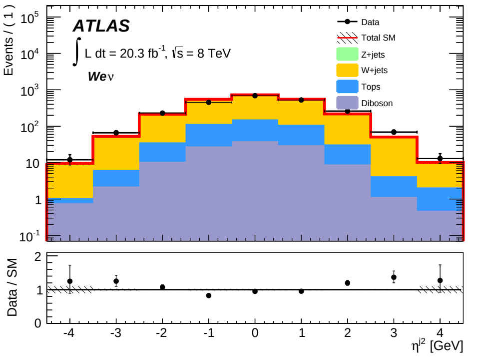


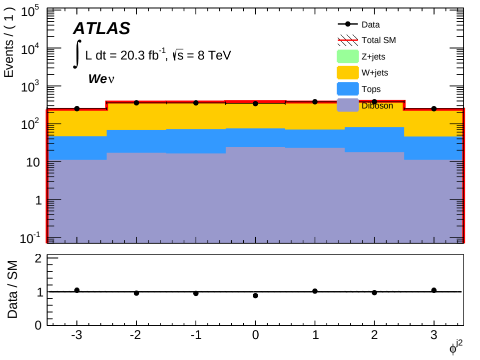


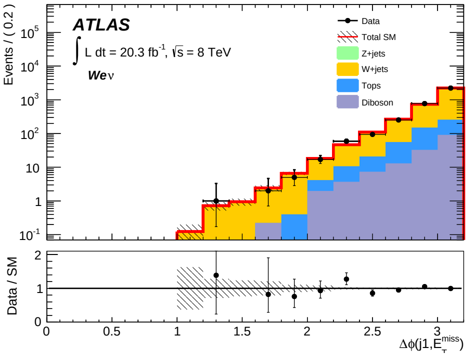


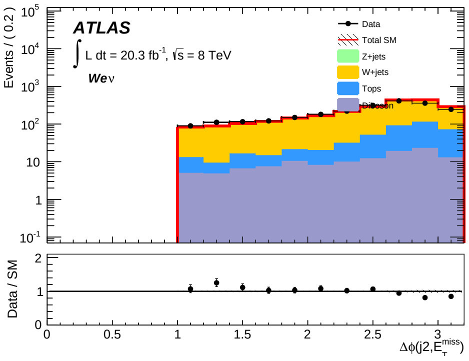


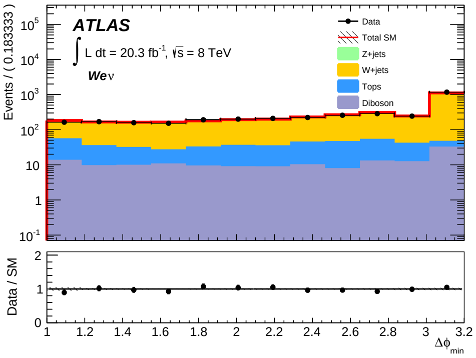


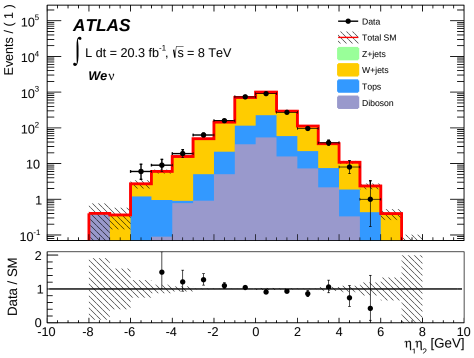


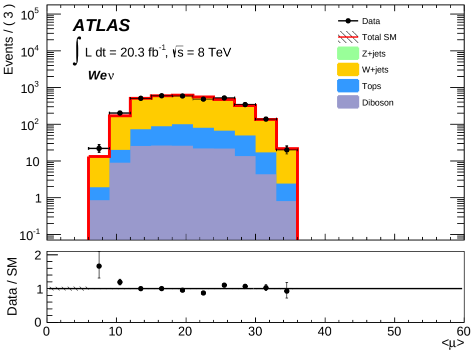


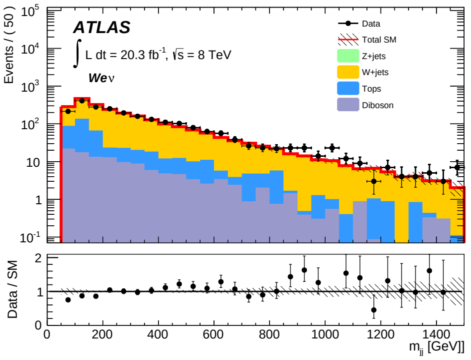

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
Weν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.2)
Data / SM
$\Delta\phi(j1,E_T^{miss})$

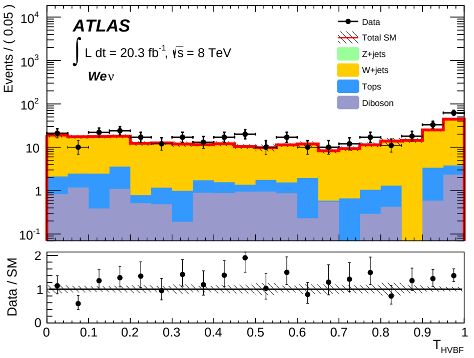

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
Weν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.2)
Data / SM
$\Delta\phi(j2,E_T^{miss})$

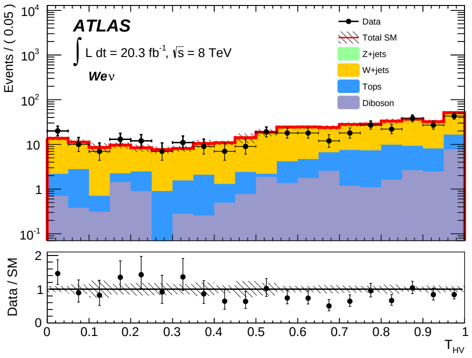

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
Weν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.183333)
Data / SM
$\Delta\phi_{min}$


ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
Weν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
Data / SM
$\eta_{j_1}\eta_{j_2}$ [GeV]

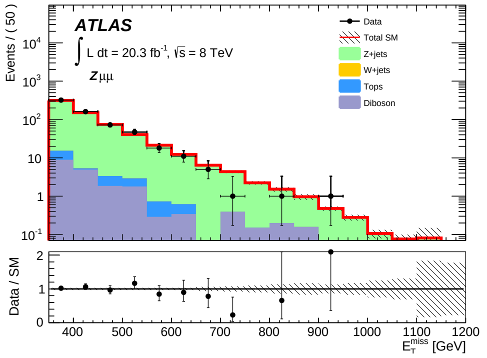

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
Weν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (3)
Data / SM
<μ>

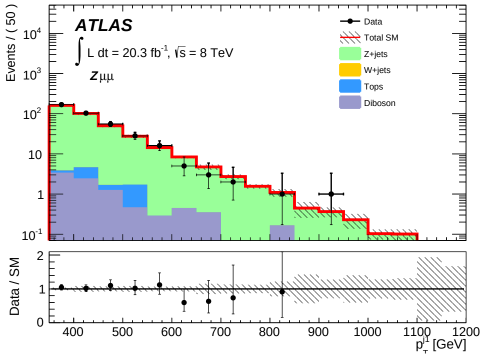

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
Weν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (50)
Data / SM
m_{jj} [GeV]]

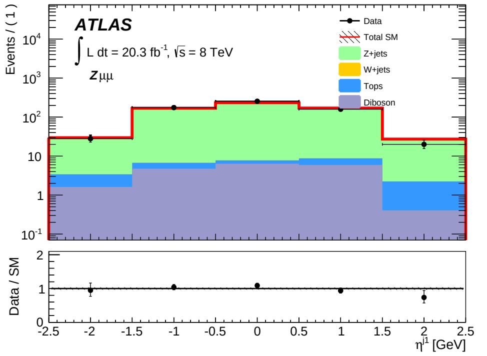

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
Weν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.05)
Data / SM
T_{HVBF}

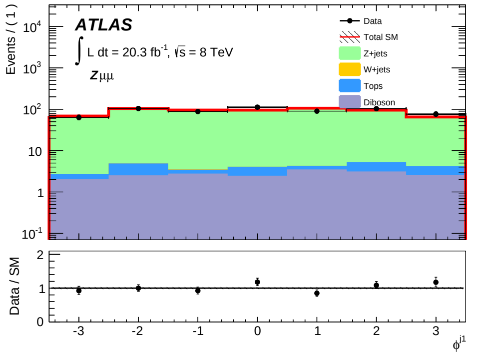

ATLAS
∫ L dt = 20.3 fb^{-1}, √s = 8 TeV
Weν
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.05)
Data / SM
T_{HV}

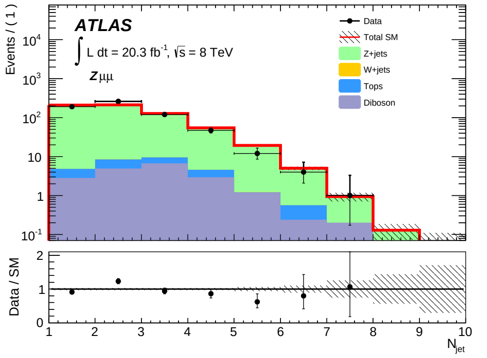
C.4 Z$\mu\mu$ control region

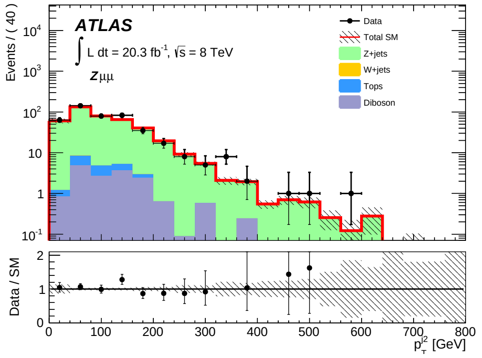


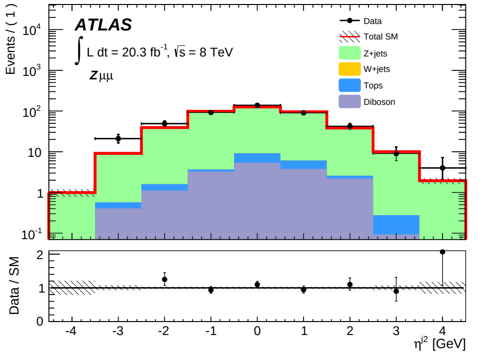


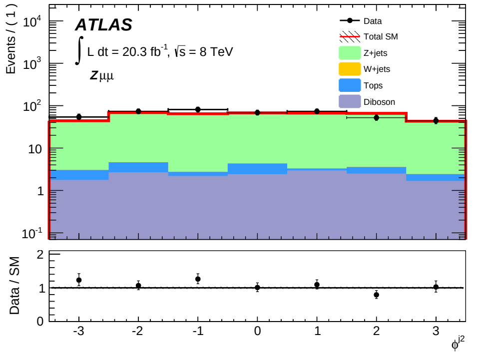


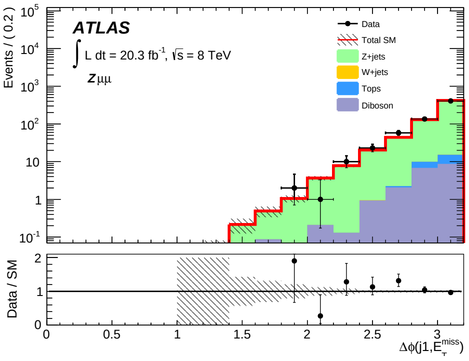


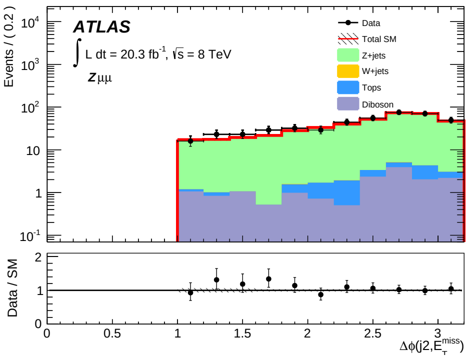


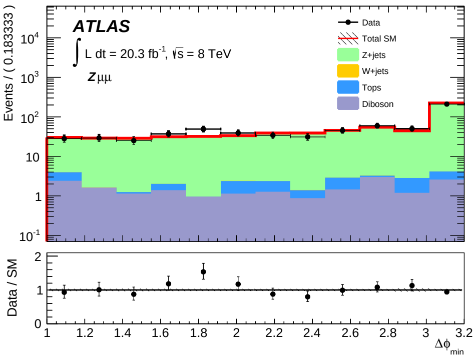


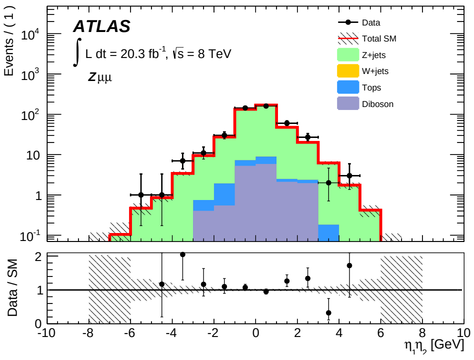


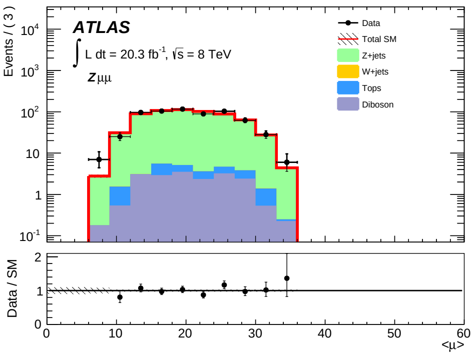


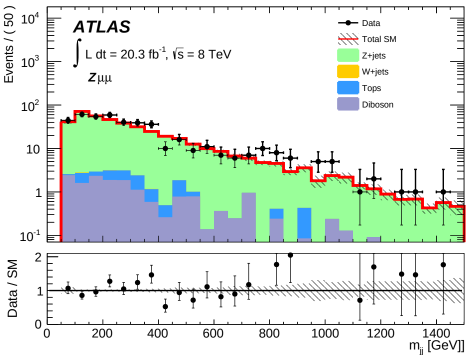

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Zμμ
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.2)
10-1
1
10
102
103
104
105
Data / SM
0
1
2
0
0.5
1
1.5
2
2.5
3
Δφ(j1,E_T^miss)

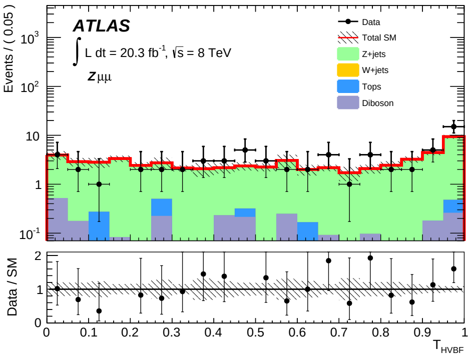

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Zμμ
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.2)
10-1
1
10
102
103
104
Data / SM
0
1
2
0
0.5
1
1.5
2
2.5
3
Δφ(j2,E_T^miss)

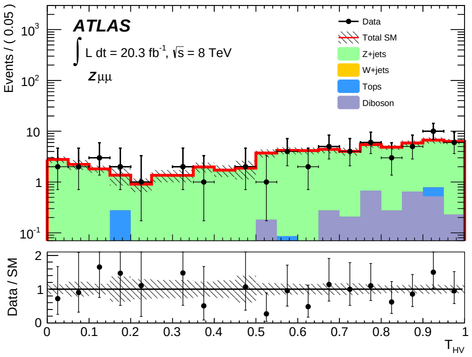

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Zμμ
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.183333)
10-1
1
10
102
103
104
Data / SM
0
1
2
1
1.2
1.4
1.6
1.8
2
2.2
2.4
2.6
2.8
3
3.2
Δφ_min


ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Zμμ
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
10-1
1
10
102
103
104
Data / SM
0
1
2
-10
-8
-6
-4
-2
0
2
4
6
8
10
η_1η_2 [GeV]

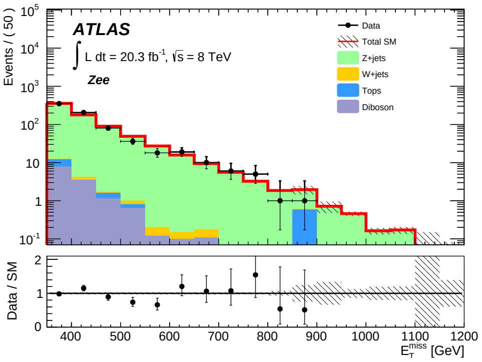

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Zμμ
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (3)
10-1
1
10
102
103
104
Data / SM
0
1
2
0
10
20
30
40
50
60
<μ>

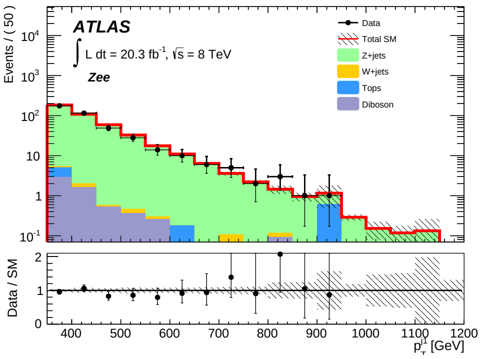

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Zμμ
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (50)
10-1
1
10
102
103
104
Data / SM
0
1
2
0
200
400
600
800
1000
1200
1400
m_jj [GeV]]

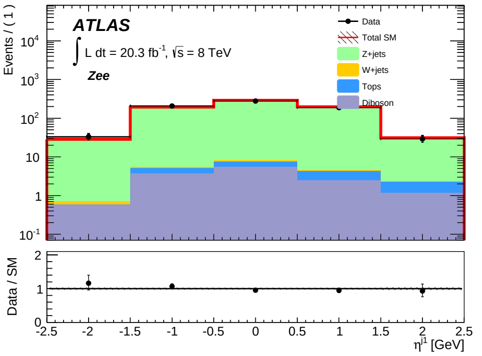

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Zμμ
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.05)
10-1
1
10
102
103
Data / SM
0
1
2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
T_HVBF

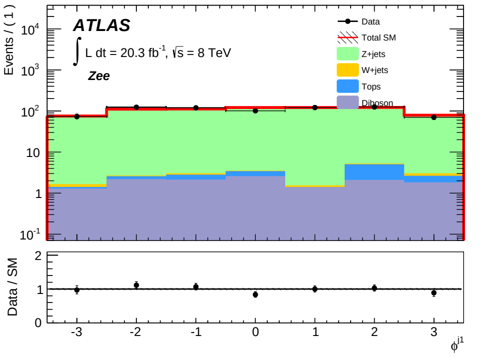

ATLAS
∫ L dt = 20.3 fb-1, √s = 8 TeV
Zμμ
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.05)
10-1
1
10
102
103
Data / SM
0
1
2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
T_HV

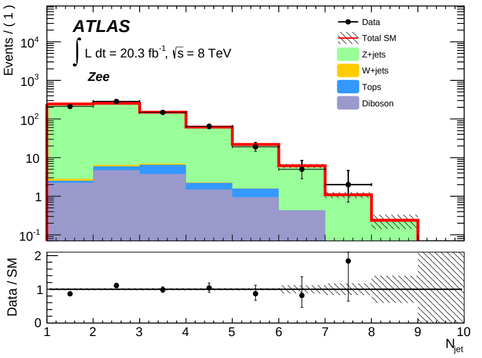
C.5 Zee control region

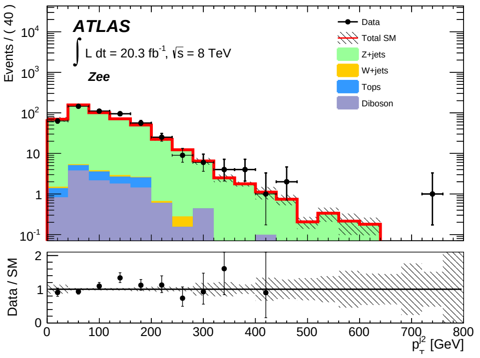


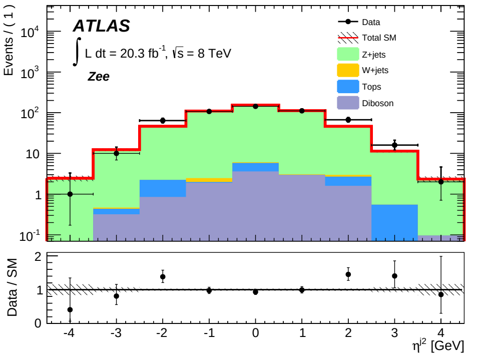


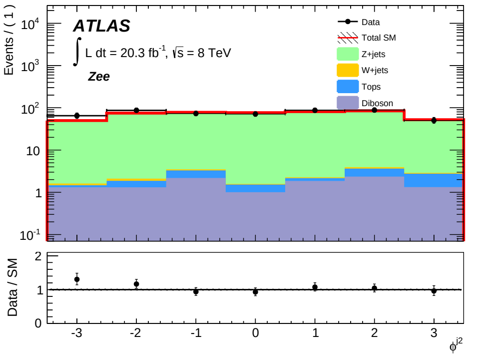


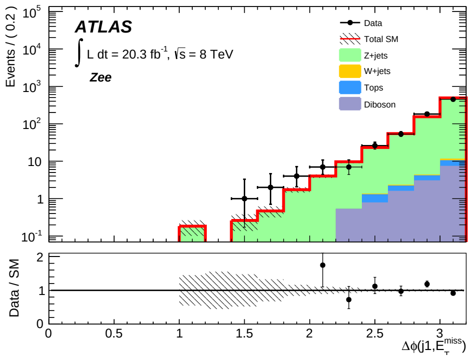


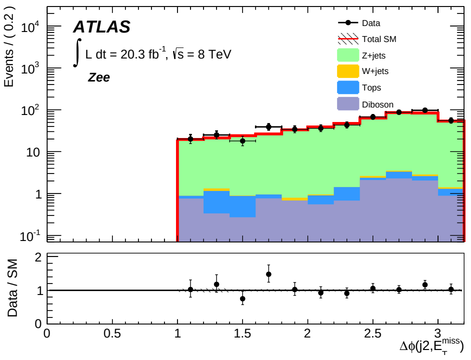


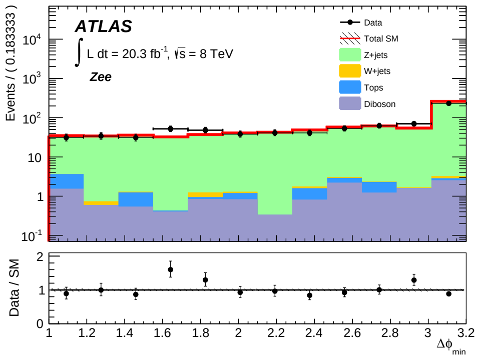


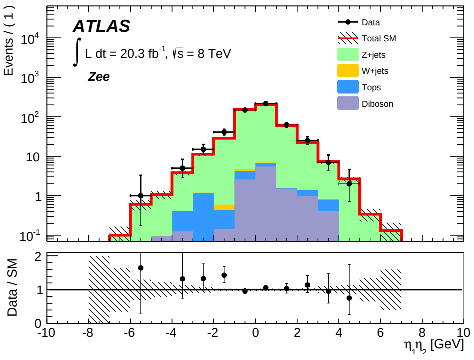


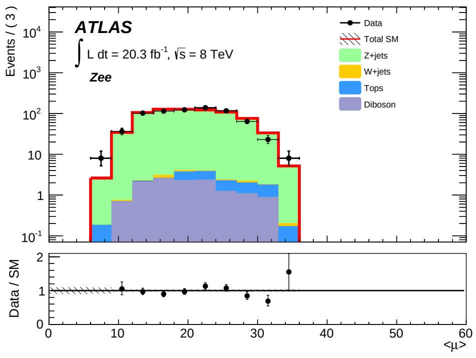


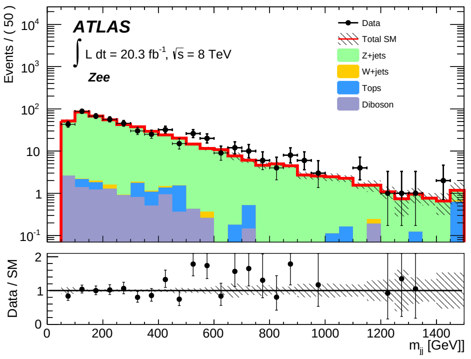

ATLAS
∫ L dt = 20.3 fb⁻¹, √s = 8 TeV
Zee
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.2)
Data / SM
Δφ(j1,E_T^miss)

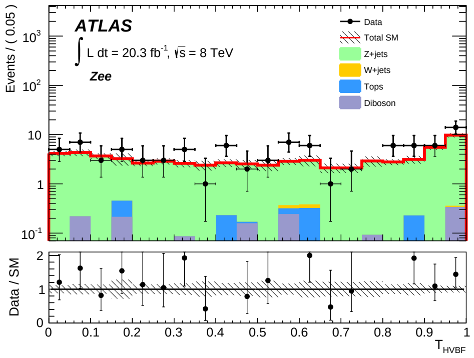

ATLAS
∫ L dt = 20.3 fb⁻¹, √s = 8 TeV
Zee
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.2)
Data / SM
Δφ(j2,E_T^miss)

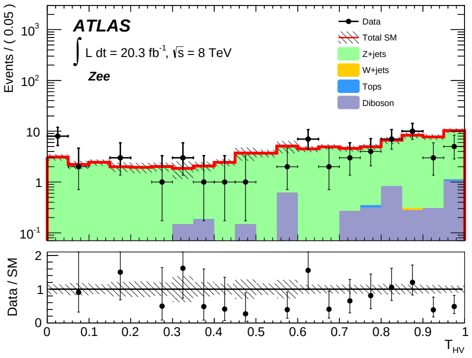

ATLAS
∫ L dt = 20.3 fb⁻¹, √s = 8 TeV
Zee
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.183333)
Data / SM
Δφ_min


ATLAS
∫ L dt = 20.3 fb⁻¹, √s = 8 TeV
Zee
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (1)
Data / SM
η_Jη_J [GeV]


ATLAS
∫ L dt = 20.3 fb⁻¹, √s = 8 TeV
Zee
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (3)
Data / SM
<μ>


ATLAS
∫ L dt = 20.3 fb⁻¹, √s = 8 TeV
Zee
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (50)
Data / SM
m_jj [GeV]]


ATLAS
∫ L dt = 20.3 fb⁻¹, √s = 8 TeV
Zee
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.05)
Data / SM
T_HVBF


ATLAS
∫ L dt = 20.3 fb⁻¹, √s = 8 TeV
Zee
Data
Total SM
Z+jets
W+jets
Tops
Diboson
Events / (0.05)
Data / SM
T_HV

Appendix D

TMVA setting for Random Forest classifier

The `ROOT` multivariate analysis package `TMVA` does not come with a predefined classifier called a Random Forest. However, by carefully choosing certain bagging parameters in the BDT method one can create a Random Forest classifier. The parameters passed to the `TMVA factory` class are given below:

```
factory->BookMethod( TMVA::Types::kBDT, "Random_Forest", \\
"!H:!V:NTrees=3000: BoostType=Bagging:MinNodeSize=2 \\
%:SeparationType=GiniIndex:nCuts=20: \\
IgnoreNegWeightsInTraining:UseRandomisedTrees=True:
\\UseNvars=4:MaxDepth=20" );
```

Bibliography

[1] P. Langacker, “The Standard Model and Beyond,” *CRC press, 2010.*

[2] The ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys.Lett. B716 (2012) 1-29*, 2012.

[3] The CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B 716 (2012) 30*, 2012.

[4] The CDS Collaboration, “Observation of Top Quark Production in $\bar{p}p$ Collisions,” *Phys.Rev.Lett.74:2626-2631*, 1995.

[5] The UA1 Collaboration, “Experimental observation of large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV,” *Phys. Lett. B 122 (1983) 103-116*, 1983.

[6] Komatsu, E.; et al., “Five-year wilkinson microwave anisotropy probe observations: Cosmological interpretation,” *The Astrophysical Journal Supplement*, vol. 180, pp. 330–76, 2009. arXiv:astro-ph/0803.0547.

[7] Clowe, D; et al., “A direct empirical proof of the existence of dark matter,” *The Astrophysical Journal*, vol. 648, pp. 109–113, 2006. arXiv:astro-ph/0608407.

[8] Abdelhak Djouadi, Oleg Lebedev, Yann Mambrini, Jeremie Quevillon, “Implications of LHC searches for Higgs–portal dark matter,” *arXiv:1112.3299 [hep-ph]*, 2012.

[9] V. Barger, R. Phillips, “Collider Physics, 1996,” *Westview Press.*

[10] H. Georgi, “Lie algebras in particle physics: from isospin to unified theories,” *Benjamin/Cummings Pub. Co.*

[11] F. Halzen and A. D. Martin, “Quarks and Leptons An Introductory Course in Modern Particle Physics,” *John Wiley and Sons*, 1984.

[12] Sir. Roger Penrose, “The Road to Reality,” *Vintage books. p. 474. ISBN 0-679-77631-1.*, 2007.

[13] Particle Data Group, “Review of Particle Physics,” *Physics Letters B Vol 667/1-5 (2008) 1-1340.*

[14] E. Noether, “Invariante variationsprobleme,” *Nachr. v. d. Ges. d. Wiss. zu Gattingen, (235- 257), 1918.*

[15] V. Barger, R. Phillips, “Collider Physics, 1996,” *p.40, Westview Press.*

[16] Zwicky, F, “On the masses of nebulae and of clusters of nebulae,” *The Astrophysical Journal*, vol. 86, p. 217, 1937. doi:10.1086/143864.

[17] Marc Kamionkowski, “Wimp and axion dark matter,” 1997. arXiv:hep-ph/9710467.

[18] Rubin, V.; Ford, W. K. Jr.; Thonnard, N, “Rotational properties of 21 sc galaxies with a large range of luminosities and radii from ngc 4605 (r = 4kpc) to ugc 2885 (r = 122kpc),” *The Astrophysical Journal*, vol. 238, p. 471, 1980. doi:10.1086/158003.

[19] Refregier, A., “Weak gravitational lensing by large-scale structure,” *Annual Review of Astronomy and Astrophysics*, vol. 41, pp. 645–668, 2003. arXiv:astro-ph/0307212.

[20] Plank Collaboration, “Planck 2013 results. xvi. cosmological parameters,” 2013. arXiv:1303.5076 [astro-ph.CO].

[21] Alcock, C.; et al., “The macho project: Microlensing results from 5.7 years of lmc observations,” *Astrophys.J.*, vol. 542, pp. 281–307, 2000.

[22] Jonathan L. Feng, Jason Kumar, “The wimpless miracle: Dark matter particles without weak-scale masses or weak interactions,” 2009. arXiv:0803.4196v3.

[23] Shinya Kanemura, Shigeki Matsumoto, Takehiro Nabeshima, Nobuchika Okada, “Can WIMP Dark Matter overcome the Nightmare Scenario?,” *Phys.Rev.D82:055026*, 2010.

[24] XENON100 Collaboration, "Dark Matter Results from 100 Live Days of XENON100 Data," *Phys. Rev. Lett. 107, 131302*, 2011.

[25] CRESST-II Collaboration, "Results on low mass WIMPs using an upgraded CRESST-II detector," *Eur. Phys. J. C (2014) 74: 3184*, 2014.

[26] CoGeNT Collaboration, "Maximum Likelihood Signal Extraction Method Applied to 3.4 years of CoGeNT Data," *arXiv:1401.6234 [astro-ph.CO]*, 2014.

[27] CERN Communication Group, "CERN faq LHC the guide," *CERN-Brochure-2008-001-Eng.*

[28] The LHC Collaboration, "LHC main website." `http://lhc.web.cern.ch/lhc/`.

[29] CERN Press Office, "LHC to restart in 2009," *CERN Press release*, 2008.

[30] The CTEQ Collabortaion *http://www.phys.psu.edu/ cteq/.*

[31] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne. `http://durpdg.dur.ac.uk/hepdata/mrs.html`.

[32] M. Glck, E. Reya, and A. Vogt. `http://zebu.uoregon.edu/~parton/partonGRV.html`.

[33] M. Glck, P. Jimenez-Delgado, and E. Reya. `http://durpdg.dur.ac.uk/hepdata/grv.html`.

[34] The NNPDF Collaboration. `http://sophia.ecm.ub.es/nnpdf/nnpdf\_pdfsets.htm`.

[35] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C.-P. Yuan, "New parton distributions for collider physics," *Phys.Rev. D82 (2010) 074024.*

[36] The ATLAS Collaboration, "Detector and physics performance technical design report," *CERN/LHCC/99-015.*

[37] The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *2008 JINST 3 S08003.*

[38] The ATLAS Collaboration, “ATLAS Website.” `http://www.atlas.ch/photos/index.html`.

[39] The ATLAS Collaboration, “Magnet system technical design report,” *CERN-LHCC-97-018.*

[40] The ATLAS Collaboration, “ATLAS Inner detector: technical design report. 1,” *CERN-LHCC-97-016.*

[41] The ATLAS Collaboration, “ATLAS Inner detector: technical design report. 2,” *CERN-LHCC-97-017.*

[42] The ATLAS Collaboration, “ATLAS pixel detector: technical design report,” *CERN-LHCC-98-013.*

[43] A. Abdesselam et al., “The barrel modules of the ATLAS semiconductor tracker,” *Nucl. Instrum. Meth. A 568 (2006) 642.*

[44] A. Abdesselam et al., “The ATLAS semiconductor tracker end-cap module,” *Nucl. Instrum. Meth. A 575 (2007) 353.*

[45] E. Abat et al., “The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance,” *2008 JINST 3 P02013.*

[46] The ATLAS Collaboration, “Liquid argon calorimeter technical design report,” *CERN-LHCC-96-041.*

[47] The ATLAS Collaboration, “Tile calorimeter technical design report,” *CERN-LHCC-96-042.*

[48] The ATLAS Collaboration, “The ATLAS forward calorimeters,” *2008 JINST 3 P02010.*

[49] The ATLAS Collaboration, “Performance of the ATLAS Trigger System in 2010,” *Eur.Phys.J.C 72 (2012) 1849.*

[50] R. Achenbach et al., “The ATLAS level-1 calorimeter trigger,” *JINST 3:P03001 (2008) 1 39.*

[51] The ATLAS Collaboration, “The ATLAS Data Acquisition and Trigger: concept, design and status,” *doi:10.1016/j.physletb.2003.10.071.*

[52] ATLAS Collaboration, "Expected photon performance in the ATLAS experiment," *ATL-PHYS-PUB-2011-007*.

[53] The ATLAS Collaboration, "Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton proton collision data," *CERN-PH-EP-2014-040*.

[54] Matteo Cacciari, Gavin P. Salam and Gregory Soyez, "The anti-kt jet clustering algorithm," *JHEP04(2008)063 doi:10.1088/1126-6708/2008/04/063*.

[55] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, "FastJet user manual," *CERN-PH-TH/2011-297*, 2011.

[56] The ATLAS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at s = 7 tev with atlas," *Eur. Phys. J. C*, 2011. arXiv:1108.5602v2 [hep-ex].

[57] R. Brun, F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl.Instrum.Meth*, vol. doi:10.1016/S0168-9002(97)00048-X, 1997.

[58] The ATLAS Collaboration, "Data quality information for 2010 and 2011 data," *ATLAS Public TWiki*, 2012.

[59] The ATLAS Collaboration, "Luminosity public results twiki," 2012.

[60] Wikipedia contributors, "Event generators," *Wikipedia, The Free Encyclopedia*, 2013.

[61] T. Gleisberg, S. Hoeche, F. Krauss, M. Schoenherr, S. Schumann, F. Siegert, J. Winter, "Event generation with sherpa 1.1," *JHEP*, 2008. arXiv:0811.4622.

[62] A. Sherstnev and R. Thorne, "Parton Distributions for LO Generators," *Eur. Phys. J. C55 (2008) 553 575, arXiv:0711.2473 [hep-ph]*.

[63] G. Corcella et al., "HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)," *JHEP 0101:010,2001*.

[64] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, Tim Stelzer, "Madgraph 5 : Going beyond," *JHEP*, 2011. arXiv:1106.0522.

[65] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, "Multiparton interactions in photoproduction at HERA," *Z. Phys. C72 (1996) 637*.

[66] Borut Paul Kersevan, Elzbieta Richter-Was, "The monte carlo event generator acermc 2.0 with interfaces to pythia 6.2 and herwig 6.5," 2012. arXiv:hep-ph/0405247.

[67] S. Frixione, B.R. Webber, "Matching nlo qcd computations and parton shower simulations," *JHEP*, 2002. arXiv:hep-ph/0204244.

[68] ATLAS Collaboration, "New atlas event generator tunes to 2010 data," *ATLAS note*, 2011. ATL-PHYS-PUB-2011-008.

[69] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables," *Phys. Rev. D78 (2008) 013004*.

[70] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG," *JHEP 0904 (2009) 002, arXiv:0812.0578 [hep-ph]*.

[71] S. Agostinelli et al., "GEANT4 - a simulation toolkit," *Nucl. Inst. and Meth A506 (2003) 250*.

[72] ATLAS Collaboration, "The ATLAS Simulation Infrastructure," *Eur. Phys. J. C 70 (2010) 823*.

[73] Paola Giovannini, on behalf of the ATLAS Liquid Argon Calorimeter Group, "Local hadron calibration with ATLAS," *Journal of Physics: Conference Series 293 012057*, 2011.

[74] ATLAS Collaboration, " ATLAS ApplyJetCalibration2012 twiki." `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ApplyJetCalibration2012`.

[75] ATLAS Collaboration, "ATLAS MissingETUtility twiki ." `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MissingETUtility`.

[76] ATLAS Collaboration, "ATLAS Muon Combined Performance Guidelines for Analyses of 2012 Data,"

[77] ATLAS Collaboration, "Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data"," *Eur. Phys. J. C72 (2012) 1909, arXiv:1110.3174 [hep-ex].*

[78] ATLAS Collaboration, "Search for new phenomena with mono-jet plus missing transverse energy signature in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *ATL-COM-PHYS-2014-1052.*

[79] K. Cranmer, A. Lazzaro, et al., "The RooStats Project," *arXiv:1009.1003*, vol. PoS ACAT2010, 2010.

[80] Kyle Cranmer, "Practical Statistics for the LHC," *arXiv:1503.07622v1 [physics.data-an]*, 2015.

[81] Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur.Phys.J.*, vol. 71, 2011.

[82] M. Baaka, G. Besjesb, et al., "HistFitter software framework for statistical data analysis," *arXiv:1410.1280 [hep-ex]*, 2014.

[83] W. Verkerke, D. P. Kirkby, "The RooFit toolkit for data modeling," *arXiv:physics/0306116*, vol. eConf C0303241, 2003.

[84] F. James, M. Roos, "MINUIT - A system for function minimization and analysis of the parameter errors and correlations," *Computer Physics Communications*, 1975.

[85] James Franklin, "The Science of Conjecture: Evidence and Probability before Pascal," 2001.

[86] Sundberg, Rolf, "Maximum likelihood theory for incomplete data from an exponential family," *Scandinavian Journal of Statistics*, 1974.

[87] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, W. Verkerke, "HistFactory: A tool for creating statistical models for use with RooFit and RooStats," *CERN-OPEN-2012-016*, 2012.

[88] I. Antcheva, M. Ballintijn, B. Bellenot, M. Biskup, R. Brun, et al., "ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization," *Comput.Phys.Commun*, 2011.

[89] Neyman, Jerzy; Pearson, Egon S, “On the Problem of the Most Efficient Tests of Statistical Hypotheses,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 231, 1933.

[90] S.S. Wilks, “The large-sample distribution of the likelihood ratio for testing composite hypotheses,” *Ann. Math. Statist*, vol. 9:60–2,, 1938.

[91] A. Wald, “Tests of statistical hypotheses concerning several parameters when the number of observations is large,” *Transactions of the American Mathematical Society*, vol. 54, no. 3, pp. 426–482, 1943.

[92] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group,” *CERN-2013-004*, 2013.

[93] kaggle.com, “Higgs Boson Machine Learning Challenge.” `https://www.kaggle.com/c/higgs-boson/forums/t/10998/winner-announcement`.

[94] Trevor Hastie, Robert Tibshirani , Jerome Friedman, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction,” *Springer; 2nd ed. 2009.*, 2011.

[95] Breiman, Leo, “Random Forests,” *Machine Learning*, vol. 45, 2001.

[96] L. Breiman, J. Friedman, R. Olshen and C. Stone, “Classification and Regression Trees,” *Wadsworth*, 1984.

[97] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, “TMVA 4 Users Guide ,” *CERN-OPEN-2007-007*, 2007.

[98] Wikipedia contributors, “CLs upper limits (particle physics).” `Wikipedia,TheFreeEncyclopedia,http://en.wikipedia.org/wiki/CLs$\_$upper$\_$limits$\_$(particle$\_$physics)`.

[99] Read, A. L., “Presentation of search results: The CL(s) technique,” *Journal of Physics G: Nuclear and Particle Physics 28 (10): 2693 2704*, 2002.

[100] Seungwon Baek, Myeonghun Park, Wan-Il Park, Chaehyun Yu, “Beyond the Dark matter effective field theory and a simplified model approach at colliders,” *arXiv:1506.06556v1 [hep-ph]*, 2015.

[101] S. Kanemura, S. Matsumoto, T. Nabeshima, N. Okada,, “Invisible Higgs and Scalar Dark Matter,” *Phys. Rev. D82*, 2010.

[102] J. Billard, L. Strigari, E. Figueroa-Feliciano, “Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments,” *arXiv:1307.5458 [hep-ph]*, 2013.

[103] ATLAS Collaboration, “Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,” *arXiv:1502.01518v1 [hep-ex]*.

[104] The ATLAS Collaboration, “The ATLAS transverse-momentum trigger performance at the LHC in 2011,” *ATLAS-CONF-2014-002*, 2014.