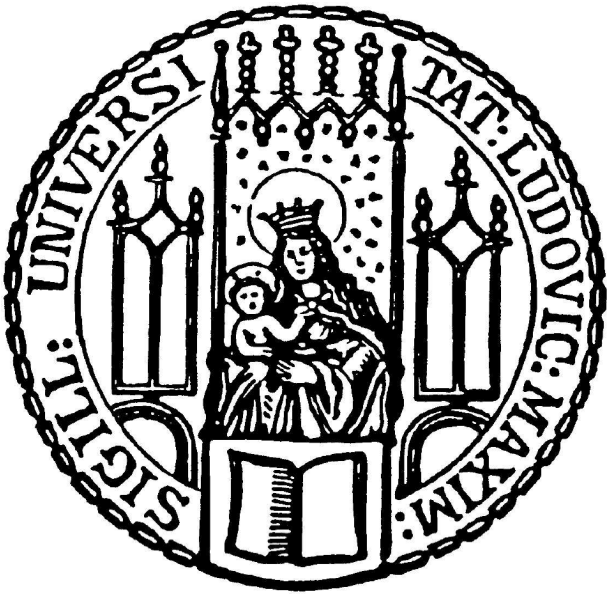
Measurement of the W^+W^- diboson production cross-section with the ATLAS Experiment at the Large Hadron Collider



Dissertation der Fakultät für Physik
der Ludwig-Maximilians-Universität München

vorgelegt von

Johannes Ebke
geboren in Tübingen

München, den 27. Juni 2012

Erstgutachterin:	Prof. Dr. Dorothee Schaile
Zweitgutachter:	Prof. Dr. Wolfgang Dünnweber

Tag der mündlichen Prüfung: 2. August 2012

Abstract

This thesis presents the first measurement of the W^+W^- production cross section in proton-proton collisions at $\sqrt{s} = 7$ TeV using the complete data set of 4.7 fb^{-1} collected by the ATLAS detector at the LHC in the year 2011. Three di-leptonic decay channels of the W^+W^- pairs are used for this analysis: decays into two electrons, two muons, and decays into one electron and one muon. For each channel, a region with a high fraction of W^+W^- events is selected and contributions from background processes are estimated using Monte Carlo simulations and data-driven techniques. In total, 1524 W^+W^- candidate events are selected, with an estimated background of 545 ± 52 events. The result of the measurement is:

$$\sigma(pp \to W^+W^-) = 52.6 \pm 2.1(\text{stat}) \pm 4.5(\text{syst}) \pm 2.1(\text{lumi}) \text{ pb}$$

This is compatible with the Standard Model cross-section prediction at next-to-leading order of $\sigma(pp \to W^+W^-) = 47.0^{+2.0}_{-1.5}$ pb.

Zusammenfassung

In dieser Arbeit wird die erste Messung des W^+W^--Produktionswirkungsquerschnitts bei Proton-Proton-Kollisionen mit $\sqrt{s} = 7$ TeV vorgestellt, die den vollständigen Datensatz von 4.7 fb^{-1} der Daten des ATLAS-Experiments am LHC aus dem Jahr 2011 benutzt. Für diese Analyse werden drei Zerfallskanäle des W^+W^--Paares verwendet: Zerfälle in zwei Elektronen, in zwei Myonen und Zerfälle in ein Elektron und ein Myon. In jedem Kanal wird eine Region im Phasenraum mit einem hohen Anteil an W^+W^--Zerfällen definiert und Beiträge von Untergrundprozessen in dieser Region werden sowohl durch Monte-Carlo-Simulationen als auch durch datenbasierte Methoden abgeschätzt. Insgesamt fallen 1524 W^+W^--Kandidaten in die Signalregion, mit einem erwarteten Untergrund von 545 ± 52 Ereignissen. Das Ergebnis der Messung ist

$$\sigma(pp \to W^+W^-) = 52.6 \pm 2.1(\text{stat}) \pm 4.5(\text{syst}) \pm 2.1(\text{lumi}) \text{ pb}$$

Dieses Ergebnis ist vergleichbar mit der Standardmodellvorhersage nächstführender Ordnung, $\sigma(pp \to W^+W^-) = 47.0^{+2.0}_{-1.5}$ pb.

Contents

Chapter 1

Introduction

In the 19th century, the scientific study of free particles began with the investigation of cathode rays by M. Faraday and J. Hittorf [1]. It took almost half a century before J. J. Thompson confirmed in 1897 that cathode rays are in fact "electrons", elementary particles that carry electric charge, and are one of the building blocks of atoms.

In the 20th century, the amount of knowledge of particle physics increased tremendously. Advances in theory were matched with expected and unexpected discoveries: The development of special relativity and the interpretation of the photoelectric effect by A. Einstein, the prediction of antimatter by P. Dirac in 1928 and the discovery of the positron by C. D. Anderson, the development of quantum field theory - and the unexpected discovery of the Muon, also by Anderson. In the 1940s and 1950s, a puzzling multitude of hadronic resonances was discovered and analysed, and eventually understood as consisting of the elementary "quarks" described by quantum chromodynamics, as proposed in 1961 by M. Gell-Mann, G. Zweig and Y. Ne'eman.

In the 1960s, quantum field theory was developed into the modern Standard Model, consisting of two coupled theories: quantum chromodynamics, describing the strong interactions between quarks, and the unified electroweak description of electromagnetism and the weak force, developed primarily by S. Glashow, S. Weinberg and A. Salam. Since then, all attempts to find deviations from the predictions of the Standard Model in particle physics experiments have failed.

One of the central theoretical constructs in the Standard Model is the electroweak gauge symmetry, which is broken spontaneously by the Higgs mechanism. The Higgs mechanism and the associated Higgs boson provide a way to add mass to the otherwise massless weak gauge bosons W and Z, but so far only indirect evidence for its existence has been found in electroweak precision measurements.

The pair production of W bosons is one of the precision probes into the electroweak sector, both because it depends on the internal structure of gauge boson interactions and because the Standard Model Higgs boson can decay into W boson pairs. W^+W^- production was first observed and measured at the Large Electron-Positron Collider at CERN in the years 1996 – 2000, and it has also been observed at the TeVatron in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV until 2011. No evidence for either the Standard Model Higgs boson or new physics beyond the Standard Model has been found at these experiments.

Since 2010 the LHC probes the Standard Model at new energy scales with a centre of mass energy of 7 TeV. Signs of the Standard Model Higgs boson are expected in many decay channels, not least in the W boson pair cross-section. In addition, the new energy regime offers a new opportunity to search for new physics, which may also result in an increase of this cross-section at higher collision energies.

In this thesis, the first measurement of this cross-section at $\sqrt{s} = 7$ TeV with the full 2011 data set of the ATLAS detector is presented, including a full estimate of statistical and systematic uncertainties. An overview of the data format, processing tools and analysis library developed for this analysis will be given in the Appendix.

No experimental result should be accepted, unless confirmed by theory

The Eddington Criterion

Chapter 2

Theoretical Foundations

2.1 The Standard Model of Particle Physics

Almost forty years ago the Standard Model of Particle Physics (SM) was formulated in its modern form in the seminal papers of Glashow [3], Salam [4] and Weinberg [5] on electroweak theory in the 1960s, and the papers of Gross, Wilzcek [6] and Politzer [7] on asymptotic freedom in quantum chromodynamics in 1973. Combined with the General Theory of Relativity [8], it predicts almost every observed effect in nature exceedingly well.

The SM describes matter using three generations of eight fermionic quantum fields. In each generation there are the up- and down-type quark fields in three colours each, one charged lepton, and one lepton-neutrino. The excitations of these fields correspond to massive particles, listed in Table 2.1.

Forces are modelled using twelve bosonic quantum fields: the massless photon γ carries the electromagnetic force, the charged W^+ and W^- and the neutral Z^0 boson are responsible for the weak force mediating nuclear decays and eight bi-coloured gluons transmit the strong force that bind the constituents of atomic nuclei. In Table 2.2 these force carriers and their essential properties are listed. The SM describes electromagnetic and weak

	1st generation		2nd generation		3rd generation	
	Flavor	Mass [MeV]	Flavor	Mass [GeV]	Flavor	Mass [GeV]
Quarks	u	$1.7 - 3.1$	c	$1.29^{+0.05}_{-0.11}$	t	172.9 ± 1.1
	d	$4.1 - 5.7$	s	$0.1^{+0.3}_{-0.2}$	b	$4.19^{+0.18}_{-0.06}$
Leptons	e	0.5110	μ	0.1057	τ	1.777
	ν_e	$\neq 0$	ν_μ	$\neq 0$	ν_τ	$\neq 0$
		< 0.000002		< 0.00019		< 0.0182

Table 2.1: Fundamental fermions in the SM with their masses [2].

Name	Interactions	Charge Carried	Mass [GeV]
Gluon g	strong	strong	0 (theory)
Photon γ	electromagnetic	none	0 (theory)
$W^{\pm}$ boson	electroweak	electroweak	80.387 ± 0.019 GeV
Z^0 boson	electroweak	none	91.1876 ± 0.0021 GeV

Table 2.2: Fundamental force carriers in the SM with their essential properties [2]. For the W boson mass, the currently most accurate measurement is given [9].

forces as effects of a *unified* electroweak force at higher energies [2].

In 1995 the final fermion predicted by the SM, the top quark, was observed for the first time at the TeVatron [10]. Since then, the last missing piece in the SM is the *Higgs boson*, a particle expected from the breaking of the electroweak symmetry, mediating the transition from the unified electroweak force to the observed separate weak and electromagnetic forces. This scalar particle must have a mass of 91^{+31}_{-24} GeV to be compatible with precision measurements in the electroweak sector [2]. Direct searches for the SM Higgs boson have now restricted its mass to 115.5 GeV $-$ 127 GeV and have observed a locally significant excess around 125 GeV, possibly a first hint of the elusive particle [11, 12].

The SM has shown to be remarkably resistant to any attempts at disproving it. In the last decade, the observation of a deficit in the number of detected neutrinos from the sun and the subsequent discovery of neutrino oscillations [13] forced an update of the theory, but no clear evidence for other physics beyond the SM has yet turned up. Observations in astronomy and cosmology, however, show that there must be more: For one, the energy content of the observed vacuum has been constrained to be about 120 orders of magnitude less than predicted by the SM, also known as "the worst theoretical prediction in the history of physics" [14]. Secondly, astronomical observations suggest that a "cold Dark Matter" that interacts gravitationally determines the large-scale structures of our universe. This Dark Matter has as yet not been detected, and it has been conclusively determined that it does not consist of SM particles. These two effects can not be explained within the SM, but none of the other candidate theories has yet been confirmed by any evidence at terrestrial experiments.

In the following, a short introduction to the Standard Model of Particle Physics is given, with a particular focus on the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism of electroweak symmetry breaking and electroweak diboson production.

2.1.1 Quantum Field Theory

Since the 19th century, we know that the classical dynamics of a physical system can be derived from a *least action principle*, that is, by minimising the integral:

$$\min_{\phi,\partial_\mu\phi} \int \mathcal{L}(\phi, \partial_\mu\phi) d^4x \tag{2.1}$$

Here $\mathcal{L}$ is the Lagrangian density of the system, which is usually constructed by writing down the difference between the kinetic and potential energy terms of the system. From this global principle a set of local Lagrange equations can be derived. These equations completely determine the evolution of ϕ, given initial values for ϕ and $\partial\phi$.

This least action principle has been generalised in the 20th century to also describe quantum fields by replacing the classical variables or fields ϕ with more abstract operators on the space of wave functions and imposing commutation relations between them. Using the requirements on relations of velocity, momentum and mass dictated by the Special Theory of Relativity, this approach yields many of the tools of Quantum Field Theory. Details on the Theory and explanation of its methods can be found e.g. in [15].

Most calculations of physical quantities in Quantum Field Theory are based on perturbation theory around small nonlinear couplings of fields[1]. Expanding Quantum Field Theory as a power series in terms of small coupling constants results in a series of integrals commonly written diagrammatically as *Feynman diagrams*. Evaluating the first few diagrams in the series, approximate results for scattering amplitudes, cross-sections and lifetimes of unstable particles can be obtained. Unfortunately, both the number and the difficulty of diagrams of higher orders grow exponentially, so for most processes only predictions to next to leading order in perturbation theory are available, and many programs to generate Monte Carlo particle scattering events only use leading order accuracy.

2.1.2 Renormalisation, Gauge Invariance and Symmetry

In the perturbative expansions of all interesting Quantum Field Theories in four dimensions there exist Feynman diagrams which contain divergences. These infinities have to cancel each other in predictions of physical values. Low-energy (infrared) divergences are often caused by emissions of soft photons in initial or final states, but can usually be re-summed to a correction factor. High-energy (ultraviolet) divergences are more problematic, but can be leveraged into a powerful tool for understanding physics at different scales by *renormalisation*.

[1] It is also possible to discretise the fields onto a lattice and use local evolution equations. This approach yields results for systems that are strongly coupled, but is extremely time-consuming, and not yet applicable for scattering.

One of these problematic ultraviolet divergences that is already known from classical electrodynamics is the self-energy of an electron. The self-energy caused by the action of the charge onto itself is infinite, unless a cutoff scale Λ at small distances and high energies is introduced, where the electron does not appear as a point-like particle anymore. The ratio of the "bare" mass and coupling parameters of the electron specified in the theory to the physical rest mass and measured coupling at low energy is infinity. To fix this, one can try to substitute the physical, measured mass and coupling in place of the "bare" values, to *renormalise* the theory. With some work, this approach also yields the *renormalisation group equations*, which predict that measured masses and couplings depend on the energy scale, and which have been tested experimentally. Intuitively this can be understood as a polarisation of the vacuum that screens the infinite, bare charge of the electron and produces the measured, finite value at large distances. At higher energies, closer to the charge, this screening effect reduces and the observed charge increases. At some very high energy scale a new theory then may take over.

However, not all theories can be renormalised. Theories where operators of a mass dimension greater than four occur (for example the original Fermi theory) are always nonrenormalisable, usually resulting in the prediction of nonphysical values for observables outside of the energy range of the theory. Even without these operators, proving renormalisability of a theory that involves interactions is extremely difficult. In contrast, theories where the interactions are restricted by *local gauge invariance* are always renormalisable if the mass dimension of the operators is at most four. Local gauge invariance is the requirement that the following replacement leaves the Lagrangian invariant:

$$\psi \to e^{-i\alpha^i(x)f^i}\psi \tag{2.2}$$

Here $\alpha^i(x)$ are arbitrary functions of space-time, and f^i are the generators of the *gauge group* operating on the fields ψ. Applying Noethers Theorem [16], one obtains the result that each gauge symmetry is accompanied by a *conserved charge*, called the charge of a field under the gauge symmetry.

Imposing this requirement on free matter fields using different gauge groups results in a rich set of interacting theories. The choice of $U(1)$ as a gauge group leads to Quantum Electrodynamics with the conserved electric charge, which yields very accurate predictions for electromagnetic effects, and was first found by Feynman and Schwinger [17]. Choosing $SU(3)$ yields Quantum Chromodynamics, a theory with three conserved charges that describes the strong interaction. However, gauge fields in these theories are always massless due to gauge symmetry - one can not introduce mass terms without sacrificing renormalisability. The weak force, however, is carried by massive bosons, which do not seem to fit into this framework.

A solution to this dilemma was developed in papers by Guralnik, Hagen, Kibble, Higgs, Brout and Englert: They showed that gauge bosons can ac-

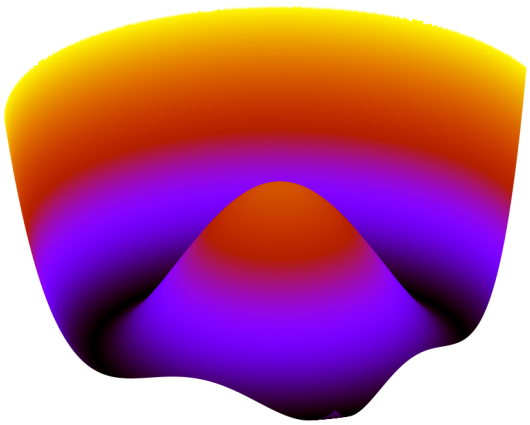


Figure 2.1: The "Mexican Hat" potential of the Higgs field.

quire mass if one introduces an additional scalar field that transforms under the symmetry, and imposes on it a quartic potential that resembles a Mexican hat (see Figure 2.1).

This leads to a dynamical breaking of the gauge symmetry, where the scalar field acquires a vacuum expectation value at the minimum of the potential. Expanding around this expectation value, one obtains three massless scalar "Goldstone" particles which can be absorbed into the original gauge fields, providing the third polarisation component needed to make them massive. Higgs also pointed out that this mechanism yields an additional massive scalar boson, now usually called the Higgs boson.

Glashow, Weinberg and Salam introduced the now accepted unified theory of electroweak interactions, based on $SU(2) \times U(1)$ combined with a four-component scalar field. There are three gauge fields of $SU(2)$, called A^1, A^2 and A^3, and one field from $U(1)$, called B. Applying the mechanism from above, one can break the symmetry of the scalar field and rewrite the fields in terms of mass eigenstates, yielding the familiar massive W^+, W^- and Z^0 bosons of the weak interaction and the massless photon γ of the electromagnetic interaction.

2.1.3 Interactions in the Standard Model

The complete Lagrangian $\mathcal{L}$ of the SM can now be written as:

$$\mathcal{L} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} \tag{2.3}$$

Quantum Chromodynamics

The first term, $\mathcal{L}_{\text{QCD}}$, consists of the fermionic quark fields q_i and the eight *gluon* fields generated by $SU(3)$ gauge symmetry G^α_μ, coupled by the strong

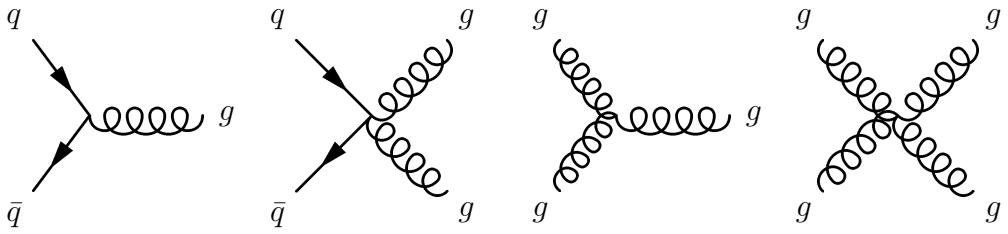


Figure 2.2: QCD Feynman diagram interaction vertices.

coupling constant g_s. There are in total 6×3 quark fields, since each of the six quark flavors comes in a *colour-triplet* of three fields with different "colour"-charges, the conserved charges under $SU(3)$. Since the $SU(3)$ symmetry is not broken, the typical "colour" assignment of "red", "green" and "blue" is arbitrary, and can be rotated at will. $\mathcal{L}_{\mathrm{QCD}}$ is given in the following equations:

$$\mathcal{L}_{\mathrm{QCD}} = \sum_j i\bar{q}_j \gamma_\mu \left(D^\mu - ig_s G^\mu\right) q_j - \frac{1}{2}\mathrm{tr} G_{\mu\nu} G^{\mu\nu} \tag{2.4}$$

$$G_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu - ig_s \left[G_\mu, G_\nu\right] \tag{2.5}$$

$$G_\mu = \sum_{\alpha=1}^{8} G_\mu^\alpha \frac{\lambda^\alpha}{2} \tag{2.6}$$

In this set of equations, D_μ already includes a set of electroweak interactions. These will be discussed in Section 2.1.3.

It has been shown with lattice computations that the binding energy given by this theory - with the rest mass of the quarks inserted manually - accounts for the observed mass spectrum of hadrons [18]. In addition, the predicted gluon radiation has been observed at colliders [19]. An interesting feature is that gluons themselves carry both colour and anticolour charge, leading to the term $-\frac{1}{2}\mathrm{tr} G_{\mu\nu} G^{\mu\nu}$, which describes gluon-gluon self-interaction. This interaction leads to self-similarity of QCD at higher energies, since quarks can also split into gluons and quarks, and gluons into quark pairs. This self-similarity leads to *anti-screening* of the QCD charge, opposite to the electromagnetic screening described in the previous section: whereas the electromagnetic "bare" charge is infinitely large, the QCD "bare" charge is infinitely small. Renormalisation therefore prescribes a reduction of QCD forces at higher energies, leading to the phenomenon of *asymptotic freedom*. Quarks and gluons behave at high energy collisions as if they were free particles. In Equation 2.4, the renormalisation is encapsulated into the strong coupling parameter g_s. Its measured numerical value falls from 0.325 at a scale of 2 GeV, to 0.1184 at the Z boson mass of 91 GeV, decreasing to 0.105 at 189 GeV.

This enables us to use perturbation theory and Feynman diagrams also for QCD processes, even if the subsequent hadronisation and fragmentation of the high-energy quarks and gluons into colourless particles still needs to be described by semi-empirical models and approximations. A list of possible Feynman diagram vertices in QCD is given in Figure 2.2.

Another important consequence of anti-screening is *confinement*: individual colour charged quarks or gluons can never be isolated on large scales but are always confined in "colourless" particles containing either an equal sum of all colours (hadrons) or an equal amount of colour and its anticolour (mesons). Otherwise, the potential energy increases linearly with the distance until enough energy is available to form a new colour-anticolour quark pair, resulting again in two or more colourless particles.

Electroweak Interactions and the Higgs boson

The remaining electroweak and Higgs parts of the SM Lagrangian are now:

$$\mathcal{L}_{\text{EW}} = \sum_k i\bar{\psi}_k \gamma_\mu D^\mu \psi_k - \frac{1}{4} F^{j\mu\nu} F^j_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \tag{2.7}$$

$$\mathcal{L}_{\text{Higgs}} = |D_\mu \phi|^2 - V(\phi) - \sum_j c_j \bar{q}_j \phi q_j - \sum_k f_k \bar{\psi}_k \phi \psi_k \tag{2.8}$$

where the covariant derivatives and field tensors are given by:

$$D_\mu = \partial_\mu - ig\frac{\boldsymbol{I}}{2}\boldsymbol{W}_\mu - ig'\frac{Y}{2}B_\mu \tag{2.9}$$

$$F^j_{\mu\nu} = \partial_\mu W^j_\nu - \partial_\nu W^j_\mu + g\epsilon^{jkl} W^k_\mu W^l_\nu \tag{2.10}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.11}$$

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \frac{\lambda}{2}(\phi^\dagger \phi)^2 \tag{2.12}$$

It has been shown experimentally that only the *left-handed* field components participate in the weak interaction introduced by $SU(2)$. We therefore require that the right-handed leptons do not transform under $SU(2)$ at all, and that left-handed leptons and quarks transform as two left-handed doublets with $|\boldsymbol{I}| = 1/2$. The charges under $SU(2)$ (the *weak isospin*) and under $U(1)$ (the hypercharge) are listed in Table 2.3.

In nature, some gauge bosons are observed to have mass. Therefore, the $SU(2) \times U(1)$ symmetry cannot be exact, but has to be "spontaneously" broken using the Higgs mechanism. This is achieved by choosing parameters μ and λ in $\mathcal{L}_{\text{Higgs}}$ so that the field ϕ obtains a vacuum expectation value. For $\lambda > 0$ and $\mu^2 > 0$, the potential V has a shape similar to Figure 2.1 with a maximum at zero, and a minimum at $\phi^\dagger\phi = \frac{\mu^2}{\lambda} =: \frac{v^2}{2}$, where v is usually defined as the vacuum expectation value. Expanding ϕ around the vacuum

	Y	I_3	$e = Y/2 + I_3$
ν_L	-1	$+1/2$	0
e_L^-	-1	$-1/2$	-1
u_L	$1/3$	$+1/2$	$2/3$
d_L	$1/3$	$-1/2$	$-1/3$

	Y	I_3	$e = Y/2 + I_3$
ν_R	0	0	0
e_R^-	-2	0	-1
u_R	$4/3$	0	$2/3$
d_R	$-2/3$	0	$-1/3$

Table 2.3: Fermion charges under the actions of $SU(2)$ (weak isospin $\boldsymbol{I}$) and $U(1)$ (hypercharge Y). Charges for left-handed fermions are given on the left, for right-handed fermions on the right side. Only the 1st-generation fermions are specified, since higher generations of fermions have the same charges.

expectation value at an arbitrary point in the minimum, three massless Goldstone modes are obtained. These can be absorbed into the W^1, W^2, W^3 and B fields, resulting in the observed mass eigenstates:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \qquad m_W = g\frac{v}{2} \tag{2.13}$$

$$Z_\mu^0 = \frac{1}{\sqrt{g^2+g'^2}}\left(gW_\mu^3 - g'B_\mu\right) \qquad m_Z = \sqrt{g^2+g'^2}\frac{v}{2} \tag{2.14}$$

$$A_\mu = \frac{1}{\sqrt{g^2+g'^2}}\left(g'W_\mu^3 + gB_\mu\right) \qquad m_A = 0 \tag{2.15}$$

$$\theta_W = \tan^{-1}(g/g') \qquad e = g\sin\theta_W \tag{2.16}$$

Substituting these into $\mathcal{L}_{\text{Higgs}}$ and $\mathcal{L}_{\text{EW}}$ yields the electroweak interactions of fermions with the electroweak bosons, shown in Figure 2.3. The photon A couples to fermions according to their electromagnetic charge. The coupling strength is $e = \sqrt{4\pi\alpha} = 0.303$ at low energies. The Z boson couples with a strength of $\frac{g}{2\cos\theta_W} \approx 0.36$ to a "V-A" current, where left-handed part of the current couples with its I_3 charge, but all parts couple with the electromagnetic charge e with the strength $-2E\sin^2\theta_W$. Finally, the W boson couples with $\frac{g}{2\sqrt{2}} \approx 0.22$ only to left-handed fermion currents where the weak isospin I_3 is raised or lowered by one unit. The relatively large numerical values of the coupling parameters are usually suppressed by the very high masses of the weak gauge bosons, which appear in the denominator of the propagator expressions.

The interactions between electroweak bosons only are shown in Figure 2.4. These triple and quadruple gauge boson vertices are particularly interesting in the study of the electroweak interaction, since they are sensitive to the mechanism of symmetry breaking.

In the SM, the coupling strengths for triple gauge boson vertices are $g_{WW\gamma} \sim e = 0.303$ and $g_{WWZ} \sim e\cos\theta_W/\sin\theta_W = 0.55$. No other triple gauge boson vertices are allowed. The quadruple gauge boson couplings are

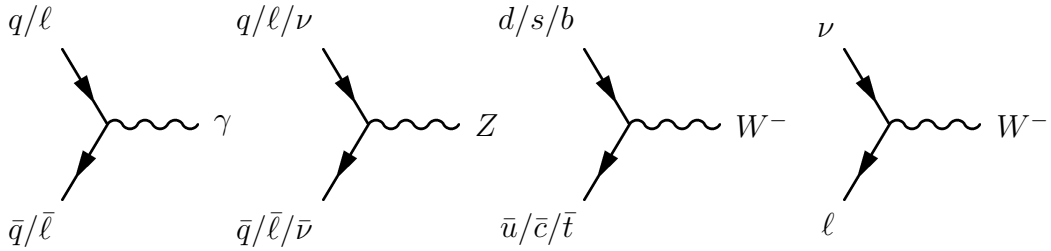


Figure 2.3: Electroweak Feynman diagram interaction vertices involving quarks and leptons. $\ell = e, \mu$, or τ.

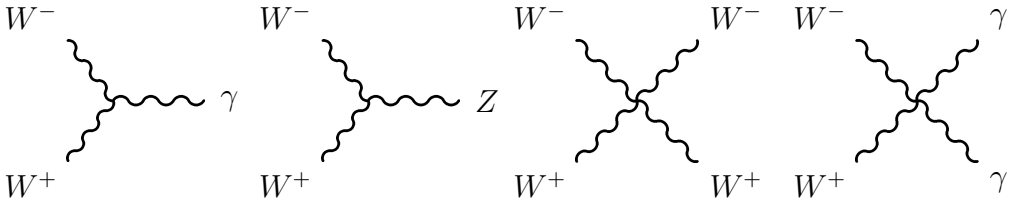


Figure 2.4: Electroweak Feynman diagram interaction vertices between electroweak bosons.

$g_{WWWW} \sim g^2 = e^2/\sin^2\theta_W = 0.40$ and $g_{WW\gamma\gamma} \sim e^2 = 0.09$, but any diagrams involving the corresponding vertices are strongly suppressed by the high masses of the gauge bosons in the propagators.

The remaining, massive degree of freedom of $|\phi|$ manifests itself as a massive scalar field - the Higgs boson, with a mass term of $m_H = \sqrt{2\lambda}v = \sqrt{2}\mu$. Since λ does not enter the Higgs couplings with massive particles, the mass of the Higgs boson cannot be predicted. However, since λ denotes the strength of the Higgs self-coupling, it is subject to the same renormalisation as was discussed in Section 2.1.2. Since $\lambda > 0$ is required for stability and $\lambda < 2\pi$ for unitarity of the interaction, given measured values for v, the mass of the Higgs boson is limited - in the case that no new physics appears before the Planck scale[2] $\lambda_{\text{Planck}} \simeq 10^{19}$ GeV - to $140 < m_H < 180$ GeV [20], with large uncertainties on the lower limit.

The simple Yukawa term $-f_k\bar{\psi}_k\phi\psi_k$ is used to add mass to fermions for every field k in Equation 2.8. Here, f_k is a free parameter that can be tuned to the rest mass measurement. This Yukawa term is both a mass term with mass $v \cdot f_k$ (since $\langle\phi\rangle = v$), as well as an interaction term with the Higgs boson with a coupling strength proportional to the mass. All interactions of the Higgs boson are given as Feynman diagrams in Figure 2.5.

[2]The Planck scale is the energy scale at which renormalisation predicts the strengths of the SM forces and gravity to be comparable.

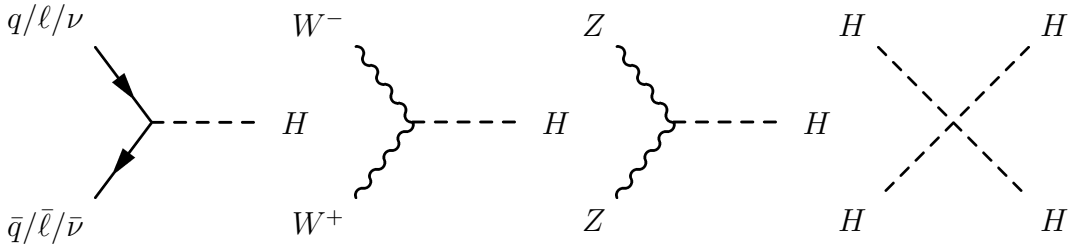


Figure 2.5: Feynman diagram interaction vertices of the Higgs boson.

Summary

In this section, a very short introduction to the Standard Model of Particle Physics has been given. The Lagrangian of the SM has been motivated by gauge symmetry and renormalisation, and all its interactions shown. Diagrammatically, they are given in Figures 2.2, 2.3, 2.4 and 2.5. From these tree-level interactions, it is now possible to build up more complex diagrams for processes that can be observed at the ATLAS detector.

Figure 2.6: Illustration of variables used in the luminosity determination.

2.2 Physics at Proton Colliders

In this section, the physics of colliding protons and their interactions are presented. Following the protons in the LHC, Section 2.2.1 starts by summarising physics of colliding charged particle beams, in particularly defining *luminosity*. Since protons are composite particles, the constituent partons have different initial momenta, summarised in *parton distribution functions* (PDFs) described in Section 2.2.2. The "hard" processes, scattering of individual partons, are presented in Section 2.2.3. Simultaneously occurring collisions between partons in the same proton and between different protons are treated in Section 2.2.4. Finally, the hadronisation of the final state into jets of hadrons is described in Section 2.2.5.

2.2.1 Particle Beams, Luminosity, and Cross-Sections

If beams of charged particles are collided, the central quantity of interest is the *luminosity* L, describing the number density of proton encounters per cross-sectional beam area and per time. The *cross-section* σ of a certain process, conversely, is the size of the (imagined) area that a proton needs to hit for this process to happen. Therefore, the rate of a process is $f = \sigma \times L$. For a long-running experiment, the quantity of interest is the *integrated luminosity* $\mathcal{L} = \int L\,dt$, which, multiplied by a cross-section, yields the number of events expected in the whole time period.

For ATLAS, the integrated luminosity is calculated using the beam parameters [21], using the equation

$$L = \frac{n_b f_r n_1 n_2}{2\pi\sigma_x\sigma_y} \tag{2.17}$$

where n_b is the number of proton bunches in the machine, n_1 and n_2 are the relevant bunch charges, f_r is the well-known revolution frequency of the LHC, and σ_x and σ_y are the Gaussian widths of the bunches in the directions perpendicular to the beam. The variables are illustrated in Figure 2.6. Since the size of the beam is difficult to measure directly, *van der Meer scans* - where the beams are shifted in known steps and the event rate is measured at the same time - are used instead. These scans are the primary method to measure $\sigma_x\sigma_y$. The total beam and the individual bunch currents $n_b n_1 n_2$ are measured using eight beam current transformers in the LHC with a sys-

tematic uncertainty of 3%, the largest contribution to the uncertainty of the luminosity measurement[3].

2.2.2 Parton Distribution Functions

Since a proton is a composite object, each parton - valence quark, sea-quark[4] or gluon - carries a different fraction of the proton forward momentum at any time, specified by the Bjorken $x = p_{\text{parton}}/p_{\text{proton}}$ variable[5]. As described in Section 2.1.3, the strong coupling is self-similar and changes with interaction energy. The momentum fractions therefore change with increasing energy as illustrated in Figure 2.7: With increasing energy, the fraction of gluons and sea-quarks increases and the two valence u quarks and one valence d quarks decrease in importance.

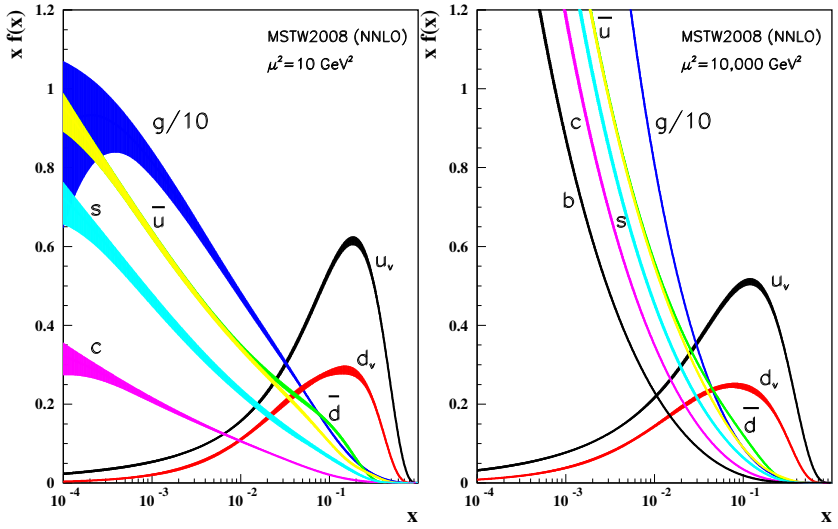


Figure 2.7: Distributions of x times the unpolarized parton distributions f(x), where f can be either the valence quarks u_v and d_v, the sea-quarks u, d, s, c and b or a gluon g. Given are the associated uncertainties using the NNLO MSTW2008 parametrisation at scales $\mu^2 = 10$ GeV^2 and $\mu^2 = 10000$ GeV^2 [2].

Given the high gluon "content" of the proton at high energies, the LHC can also be described as a gluon-gluon collider. The high sea-antiquark content also increases the cross-section for initial states with a quark and an antiquark. This property of the PDF at very high energies enables the LHC to search

[3] All other uncertainties only add up to 2.1%.

[4] A sea-quark is a quark from QCD pair-production inside the strongly interacting environment of a hadron.

[5] This is a very simplified account, details can be found e.g. in [2] and [22].

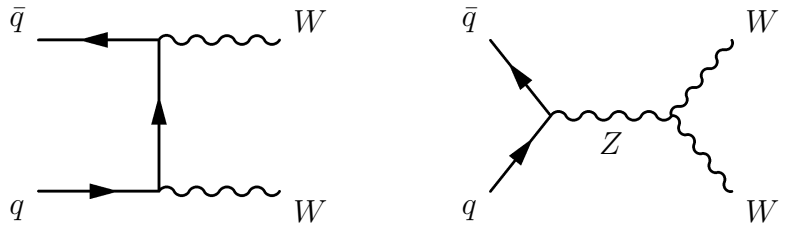


Figure 2.8: Leading order W^+W^- diboson production diagrams.

for interesting processes without colliding protons with antiprotons, as was necessary for the TeVatron collider at Fermilab at lower energy.

The PDFs used as input for MC simulation of events are listed in Section 5.1. Historically, PDFs are extracted and applied using leading-order (LO) or next-to-leading-order (NLO) calculations in perturbation theory depending on the MC generator program order. However, in the last years it has been shown that modified NLO PDFs can also improve the predictions of LO generators [23], and such PDFs are also used in this analysis.

2.2.3 Diboson Production and Background Processes

If an interesting process with high momentum transfer occurs in a collision, it is called the "hard process" in the event. For hard processes, the QCD property of asymptotic freedom (see Section 2.1.3) makes it possible to factorise the cross-section into the probability of the initial state given by the PDF and the direct calculation of the hard process cross-section given the initial state from SM Feynman diagrams. Final states may also involve gluons and quarks, the transition of these into hadrons can again be factored out since they are asymptotically free at high energies.

In this section the hard processes most important for this analysis will be presented. Cross-sections given are for illustration of the relative magnitude, and are calculated using the tree-level generator MadGraph5 1.4.2 [24]. Since the cross-section areas are very small, they are customarily given in the unit of "barn", where 1 b = 10^{-28} m^2. The cross-sections used in the actual analysis will be presented in Section 5.1. First, the signal process of W^+W^- production will be presented, followed by important background processes.

W^+W^- Production

The hard W^+W^- diboson production is the signal process of this analysis. Its calculated cross section at NLO [25] is $\sigma^{NLO} = 47.0^{+2.0}_{-1.5}$ pb. The leading order diagrams shown in Figure 2.8 contribute ≈ 30 pb, the remainder arises mostly from NLO diagrams with quark-gluon and gluon-gluon initial states as shown in Figure 2.9. Therefore, about 221,000 W^+W^- pairs are expected in

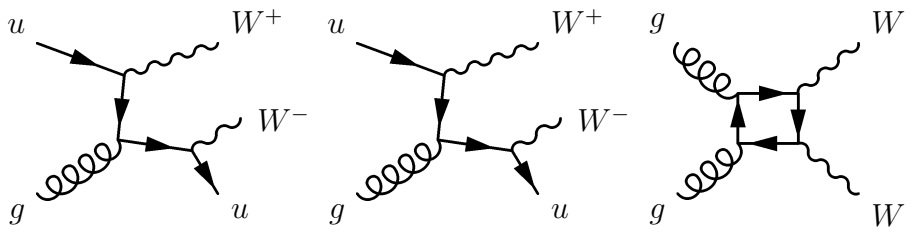


Figure 2.9: Examples of W^+W^- production diagrams of higher order. The first two show the simplest tree-level diagrams with quark-gluon initial states, which contribute $\approx 25\%$ to the cross-section at tree-level. The rightmost diagram shows W^+W^- production from a gluon-gluon initial state with fermions running in the loop. The cross-section of this process is $\approx 2\%$ of the tree-level cross-section.

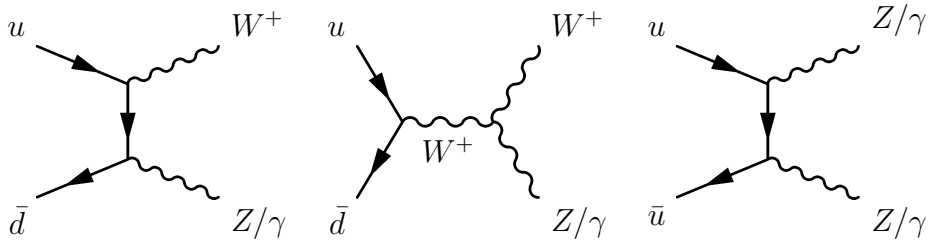


Figure 2.10: Leading order diagrams of non-W^+W^- diboson production processes.

the 4.7 fb^{-1} of data collected by the ATLAS detector in 2011. Of these, 10.5% decay into two leptons and two neutrinos, the event signature examined in this analysis.

The triple gauge boson interaction vertex in the right-hand diagram of Figure 2.8 is of particular interest, since it is one of the few vertices where the inner structure of the electroweak interaction is crucial. Anomalous couplings of the weak bosons arising in extensions of the SM can be measured [26] and the predictions from the SM formulation of electroweak theory can be tested.

Other Diboson Production

Other diboson production processes, $W^{\pm}Z$, ZZ, $Z\gamma$ and $W^{\pm}\gamma$, are also of interest, since they can be used in combination to constrain anomalous triple gauge couplings. In this analysis, all but the W^+W^- production are however considered background processes and are suppressed by the event selection. In Figure 2.10 the main production diagrams of the non-W^+W^- diboson production processes are displayed.

The cross-section of diboson production is much lower than single-boson production, since at least two weak vertices and two high-mass propagators

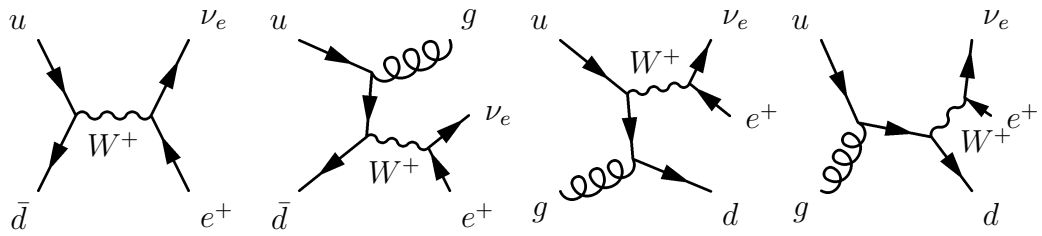


Figure 2.11: Examples of diagrams of W^+ boson production and decay to $e^+\nu_e$ in proton-proton collisions. Diagrams for W^- can be obtained by switching up- and down quarks in the diagrams.

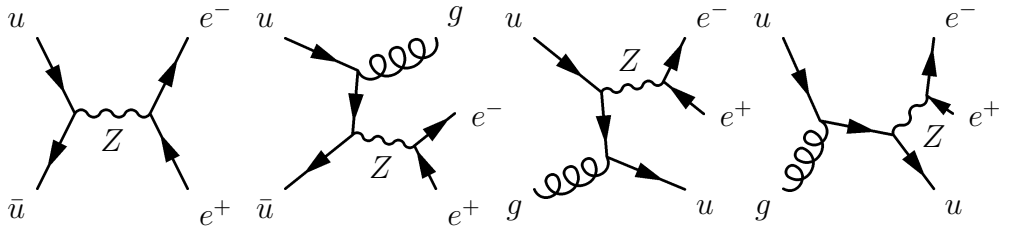


Figure 2.12: Example diagrams of Z boson production and decay to $e^+\nu_e$ in proton-proton collisions.

must be present in the Feynman diagram. The SM predictions for these processes at the LHC are calculated to NLO in [25]. For ZZ, the cross-section is 6.46 pb, for $W^{\pm}Z$ it is 11.88 pb, both still lower. $W\gamma*$ and $Z\gamma*$ are more frequent with 485 pb and 292 pb respectively, but do not have event signatures as similar to the signal as the others.

W Boson Production

W bosons are produced with a high cross section of ≈ 70 nb. Most events arise from the Feynman diagrams in Figure 2.11, and $\approx 20\%$ have associated jets from final-state quarks or gluons. These jets can possibly be mismeasured and interpreted as leptons, which causes them to be background to the $W^+W^- \to \ell\nu\ell\nu$ signal.

Z boson production

In Figure 2.12 the main production diagrams for Z boson production are depicted. Its cross-section is ≈ 20 nb, less than W production. Again, approximately 20% of the events have additional jets. This fraction is larger than pure gluon radiation would suggest, since all diagrams with gluons in the initial state are at tree level only possible if there is also at least one outgoing quark.

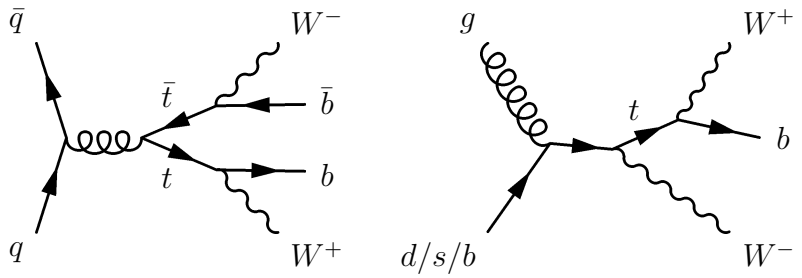


Figure 2.13: Example top pair and W-associated single-top production diagrams.

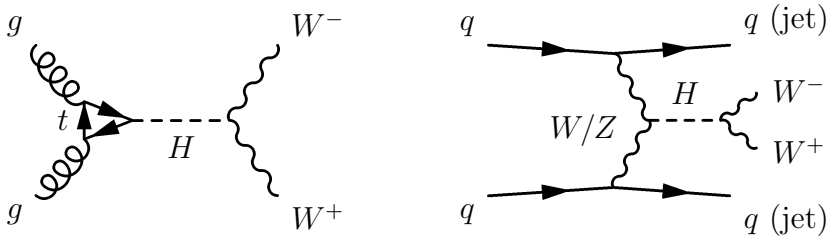


Figure 2.14: Main Higgs boson production mechanism at the LHC.

Top Pair and Single Top Production

A large fraction of background events involve top quark pairs and single tops. Most important is the top pair production, illustrated in Figure 2.13 left. With an NLO cross-section of 165 pb [27] and the production of two real W bosons, separating this background from non-top W pair production is challenging. In addition, single top quarks produced in association with a W boson can also cause events similar to the signal via the process shown in Figure 2.13 right, with a LO cross-section of 60 pb.

Higgs Production

Finally, the SM Higgs boson produced mainly in the processes shown in Figure 2.14 can also decay into a pair of W bosons. The cross section for Higgs production and the branching ratio for this decay depend strongly on the as yet unknown Higgs boson mass m_H, but is expected to be between 1.8 pb (m_H = 115 GeV) and 4.9 pb (m_H = 130 GeV). This background will be considered separately at the end of the cross-section measurement.

2.2.4 Multiparton Interaction and Pileup

In the high-energy proton-proton collisions at the LHC, asymptotic freedom allows us to use PDFs and calculate cross-sections as if the partons were not

inside a proton. However, the effects of the remaining partons of the proton also have to be taken into account. These partons are also scattered, they hadronise, and cause additional particles to appear in the final state. For this analysis, most MC simulation samples have used the Jimmy model [28], which uses a combination of perturbative and empirical models to simulate these effects.

In addition to multiparton interactions inside the proton, on average ≈ 9 other proton-proton collisions occur in each bunch crossing. Most collisions do not result in high-energy jets or leptons, and do not interfere with the measurement. However, the energy measurement scale and resolution has to be adjusted, and background from the rare occurrence of multiple simultaneous hard scatterings has to be estimated.

2.2.5 Hadronisation and Jets

Due to *confinement*, quarks and gluons in the final states of hard processes or multiparton interactions cannot continue on to be observed directly, they have to condense into colourless hadrons. During hadronisation, the potential energy built up between the moving partons is often modelled as a string. If sufficient energy is available, additional partons can be pair-produced, and the string "snaps". As this process is by its nature not accessible to perturbation theory, it must be approximated with theoretically motivated but ultimately empirical models. Hadronisation models used for MC simulations in this analysis are Herwig [28] and Pythia [29]. An example of the output of the Pythia hadronisation model for a proton-proton collision is given in Figure 2.15. No hard process has been simulated for this interaction. Green lines indicate gluons, red lines symbolise quarks and blue lines antiquarks. Black lines represent hadrons, yellow lines emitted photons. The simulation of the interactions with matter and detection of these particles by the ATLAS detector will be described in the following chapter.

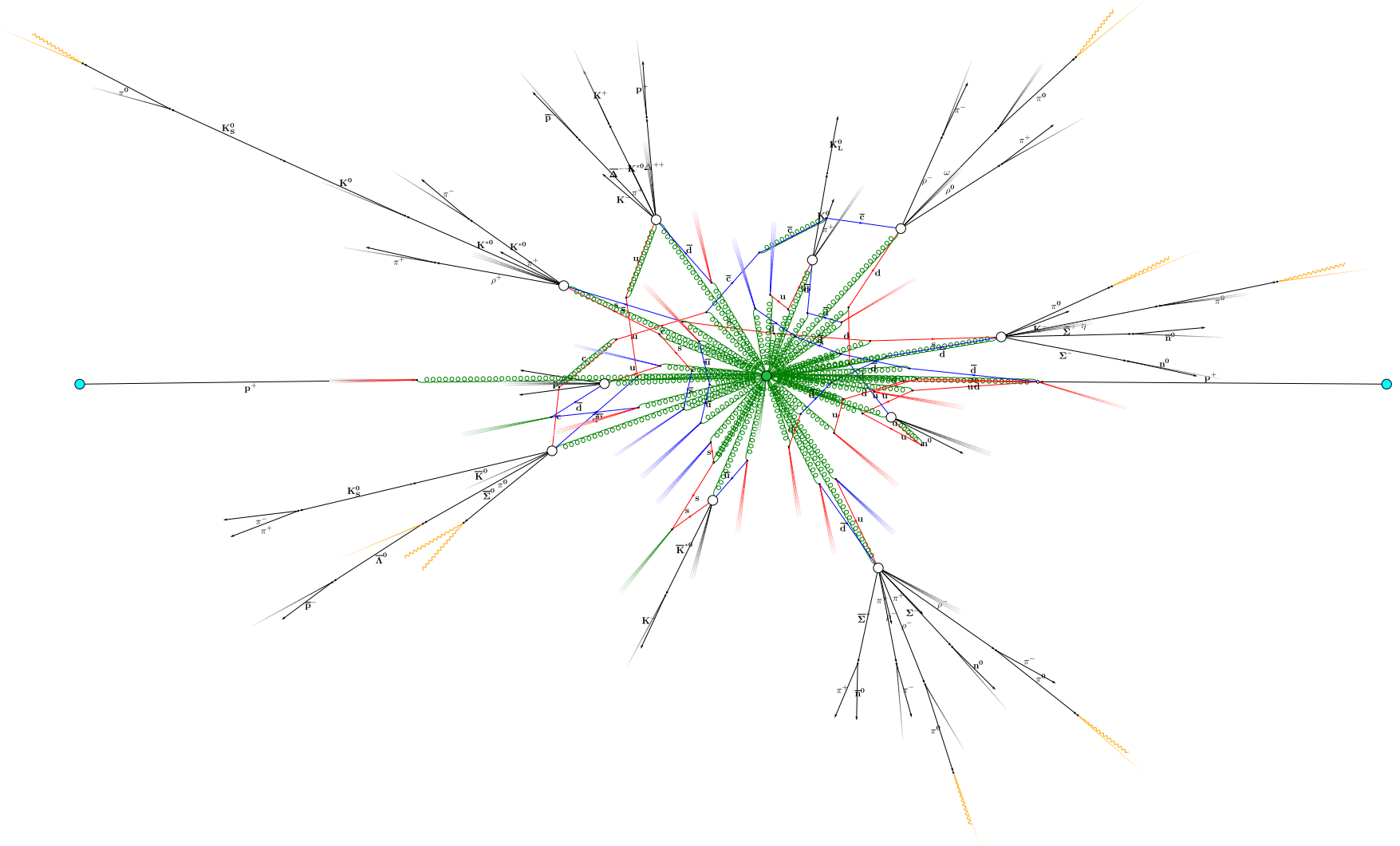


Figure 2.15: Proton-proton collision at $\sqrt{s} = 7$ TeV simulated and hadronised with the Pythia event generator. Only particles with a transverse momentum of more than 2 GeV are shown. Visualised using MCViz [30], a visualisation program written by the author in the course of this thesis.

Chapter 3

Experimental Setup

This chapter describes the origin of the data analysed for this thesis. First, in Section 3.1, the Large Hadron Collider and its performance in the year 2011 is introduced. In Section 3.2, the construction and performance of the ATLAS detector is described. The trigger and data acquisition systems are detailed in Section 3.3, and in Section 3.4 the ATLAS computing infrastructure and tools used to handle the large amount of data are presented.

3.1 The Large Hadron Collider

The Large Hadron Collider is a synchrotron-type accelerator with a diameter of 27 km situated at the European Laboratory for Particle Physics CERN on the border of Switzerland and France. It supports accelerating two counter-rotating bunched beams of protons from the injection energy of 450 GeV delivered by the CERN accelerator chain[1] (see Figure 3.1) to a design energy of 7 TeV. During 2010 and 2011 the energy was limited to 3.5 TeV, since the incident on 19 September 2008 made it clear that additional safety measures - that will have to be implemented in the 2013 shutdown - are necessary to run at the full design energy. Around the LHC ring, 1232 superconducting main dipole magnets keep the protons on the circular track, while superconducting quadrupole, hexapole, octupole and decapole magnets provide beam corrections. The two beams are brought to collision at four points around the ring. The special-purpose LHCb and ALICE detectors are situated at two of these points, LHCb being built especially to examine b-hadron decays, and ALICE observing heavy ion collisions in special runs of the LHC. The other two interaction points host the general-purpose CMS and ATLAS detectors.

The LHC run of 2011 was very successful, thanks to the excellent performance of the whole CERN accelerator chain and the LHC machine itself.

[1] The chain consists of the Linac 2 for initial acceleration, the PS Booster for acceleration to 1.4 GeV, the Proton Synchrotron (PS) up to 28 GeV, and finally the Super Proton Synchrotron (SPS) for acceleration up to the injection energy of 450 GeV.

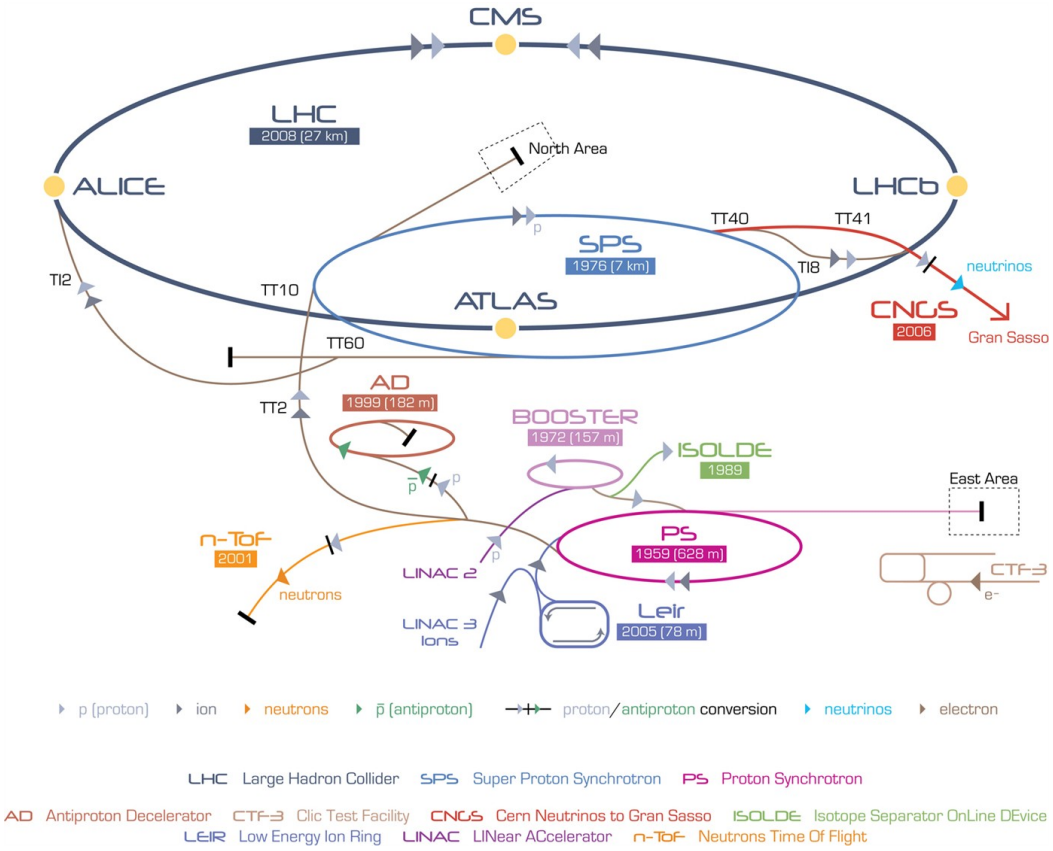


Figure 3.1: Overview over the CERN Accelerator Complex [31].

In the 2011 run period, the frequency of proton bunches was increased up to 20 MHz, half the design frequency, with a maximum of 1854 bunches in the accelerator. The maximum average bunch charge was 1.4×10^{11} protons per bunch - exceeding the LHC design specification of 1.15×10^{11} protons per bunch - while the accelerator was filled with 1331 bunches. At this configuration, the maximum instantaneous luminosity of $3.65 \times 10^{33} \mathrm{cm}^{-2}\mathrm{s}^{-1}$ was delivered to the ATLAS experiment, leading to an average of 17 interactions per bunch crossing. The total integrated luminosity delivered to the ATLAS interaction point in 2011 was 5.8 fb^{-1}, of which 5.6 fb^{-1} was delivered with stable beam conditions.

Since the goal at the beginning of 2011 had been to obtain at least 1 fb^{-1} of integrated luminosity, optimistic projections ranging around 3 fb^{-1}, the 2011 performance of the LHC can reasonably be called very successful. In 2012, the beam energy has been increased to 4 TeV, with plans to record at least 15 fb^{-1} of integrated luminosity, an amount that would be sufficient to either discover or exclude a Standard Model Higgs Boson.

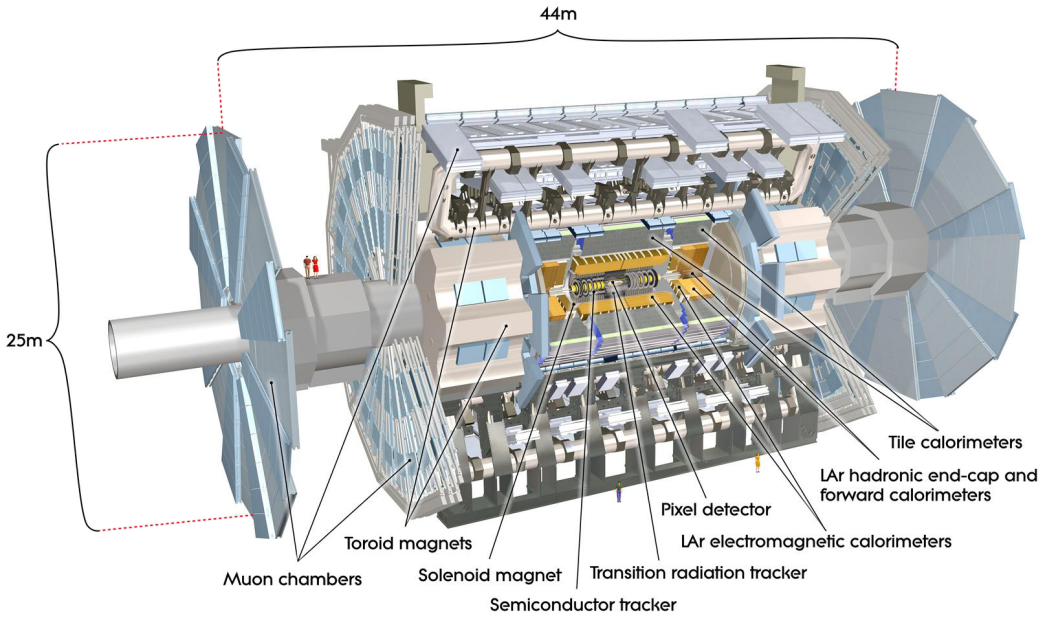


Figure 3.2: Overview of the ATLAS Detector [33].

3.2 The ATLAS detector

The ATLAS[2] detector [32] at the LHC Point 1 interaction region is the centrepiece of the ATLAS experiment. More than 3000 physicists work in the ATLAS collaboration to operate the detector, collect and store the data, and finally use the computing infrastructure of the ATLAS experiment to analyse the data and produce physics measurements.

The centre of the ATLAS detector is the nominal *interaction point*, where the proton beams of the LHC are focused and made to collide. A right-handed coordinate system is rooted at this point, with the x-axis pointing towards the centre of the LHC, the y-axis pointing upwards and the z axis pointing in the beam direction[3]. The azimuthal angle in the x-y plane transverse to the beam direction is denoted as ϕ, and the polar angle to the beam axis is denoted θ. The pseudorapidity is defined as $\eta \equiv -\log(\tan(\theta/2))$. The detector is built around a beryllium beam pipe of 58 mm diameter, containing the beam vacuum of $10^{-8} - 10^{-9}$ Pa [34]. This ultra high vacuum not only ensures long beam lifetimes of up to 100h but also minimises background from beam-gas interactions.

The subdetectors of the ATLAS detector are arranged surrounding the beam pipe, starting with the inner detector (ID, Section 3.2.2) tracking

[2]A Toroidal LHC ApparatuS

[3]Since the whole LHC tunnel is tilted, the y-axis is 0.704° off vertical.

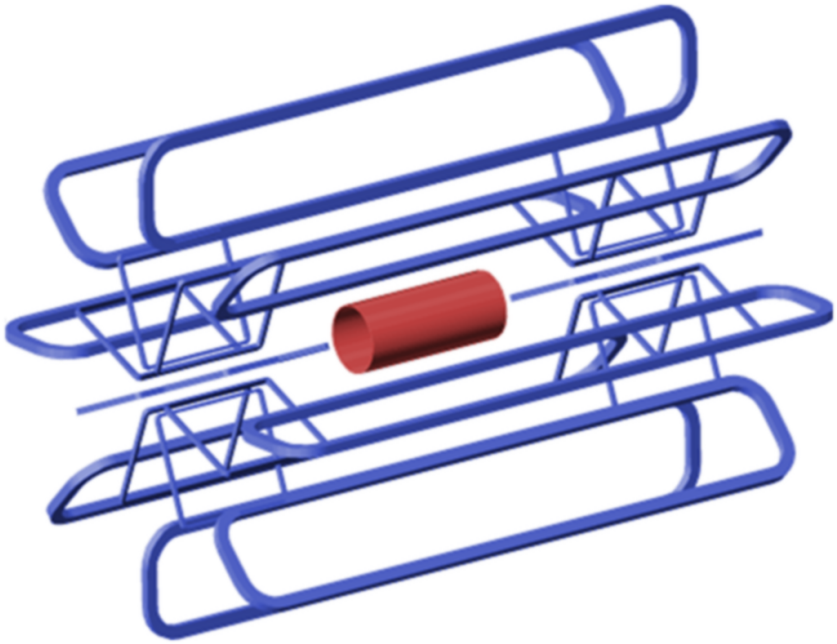


Figure 3.3: View of the bare superconducting coils of the ATLAS magnet system. The toroid barrel and endcap magnets are shown in blue, the central solenoid is shown in red [35].

charged particles, the electromagnetic calorimeter (ECal, Section 3.2.3) together with the hadronic calorimeter (HCal, Section 3.2.4) measuring the energy of charged and neutral particles, and the voluminous muon spectrometer (MS, Section 3.2.5) identifying and measuring the momenta of muons. Three superconducting magnet systems (Section 3.2.1) are used to curve the tracks of charged particles, a solenoid around the ID, and air-core barrel and end-cap toroids in the muon spectrometer. Finally, the BCM, LUCID and ALFA detectors are used in the forward regions close to the beam pipe to measure luminosity (Section 3.2.6). Figure 3.2 gives an overview over the different components.

3.2.1 Magnet Systems

The angular deflection of charged particles in a magnetic field is proportional to their momentum. Therefore, magnetic fields are commonly used in tracking detectors to measure the momenta of particles. In order to measure very high momenta where tracks are almost straight lines, it is necessary to maximise the deflection, proportional to the path integral of the magnetic field orthogonal to the particle track. At ATLAS, the magnet system is split into a solenoid around the inner detector and a toroid system consist-

ing of three separate toroid magnets (see Figure 3.3). Both systems use coils made of aluminium-stabilised superconducting Niobium-Titan alloy cooled to 4.5 K [35].

Solenoid

The ATLAS superconducting solenoid occupies the area of $122 - 132$ cm from the beam axis and is 5.3 m long, providing a nearly uniform field of 2 T inside the inner detector. The field integral along the path of charged particles is $0.5 - 2$ Tm. The additional material that particles have to traverse is minimised by using the cryostat of the liquid argon calorimeter both for support and cooling, and reusing the iron absorber of the hadronic calorimeter as the return yoke. Its 1173 windings carry a current of 7.6 kA, and its total mass is 5.7 t [35].

Toroid

The ATLAS superconducting air-core toroid barrel with an inner radius of 4.7 m, an outer radius of 10.1 m and a length of 25.3 m provides an average magnetic field of 0.5 T, and field integrals of $1 - 7.5$ Tm. It is by far the largest superconducting magnet in the world. Since the field lines are mostly contained within the magnet system, no return yoke is needed, and the total magnet assembly is very lightweight compared to its size, massing 830 t while covering a volume of 8000 m^3 (0.1 g/cm^3). This minimises scattering and energy loss of the measured particles [32, 35].

3.2.2 Inner Detector

For optimal resolution of particle tracks and identification of secondary vertices it is necessary to start tracking particles as closely to the interaction point as possible. The inner detector therefore starts tracking charged particles 5 cm from the beam axis, just outside the beam pipe, continuing on to a radius of 115 cm. Three subsystems are used in the inner detector, each optimised to the different requirements at different distances from the beam axis, and all covering a pseudorapidity[4] region of $|\eta| < 2.5$: The pixel detector, the semiconductor tracker, and the transition radiation tracker.

[4]The pseudorapidity $\eta \equiv -\log(\tan(\theta))$ is the same as the rapidity of a massless particle originating at the centre of the detector.

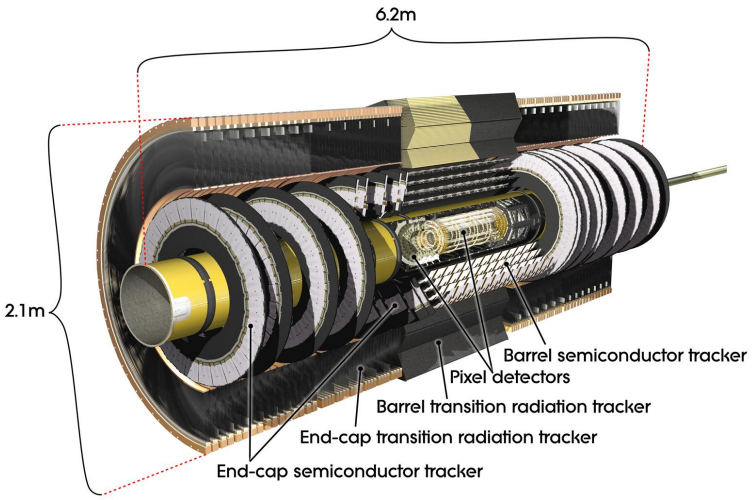


Figure 3.4: Silicon Tracker [33].

The Pixel Detector

The pixel detector consists of three layers of silicon sensors at 5, 9 and 12 cm from the beam axis, and two sets of three "endcap" discs at 50, 58 and 65 cm along the z-axis from the interaction point (see Figure 3.4). In total, the pixel detector comprises around 80 million pixels 50×400 μm in size, covering an area of 1.7 m^2. The efficiency of the pixel tracker is $\approx 99.9\%$, expected to fall to $\approx 97.8\%$ after five years of irradiation. Its spatial resolution is ≈ 12 μm, enabling precision reconstruction of primary and secondary interaction vertices, important for identifying jets originating from long-lived b-hadrons [36].

The Semiconductor Tracker

Since the number density of particles decreases with the distance from the beam axis, the semiconductor tracker uses silicon stereo microstrips instead of pixels, both to reduce the number of readout channels and to reduce cost. It consists of 4088 planar p-in-n silicon stereo micro-strip detectors with two layers rotated 50 mrad with respect to each other, covering an area of 61 m^2. Its four barrel layers are at radii from $30 - 52$ cm from the beam axis, and nine endcap discs are at each side, $85 - 273$ cm from the interaction point (see Figure 3.5). To minimise scattering, the individual sensors have a thickness of just 280 μm. Each active area is 80 μm wide and 12 cm long, providing high resolution radial tracking [37].

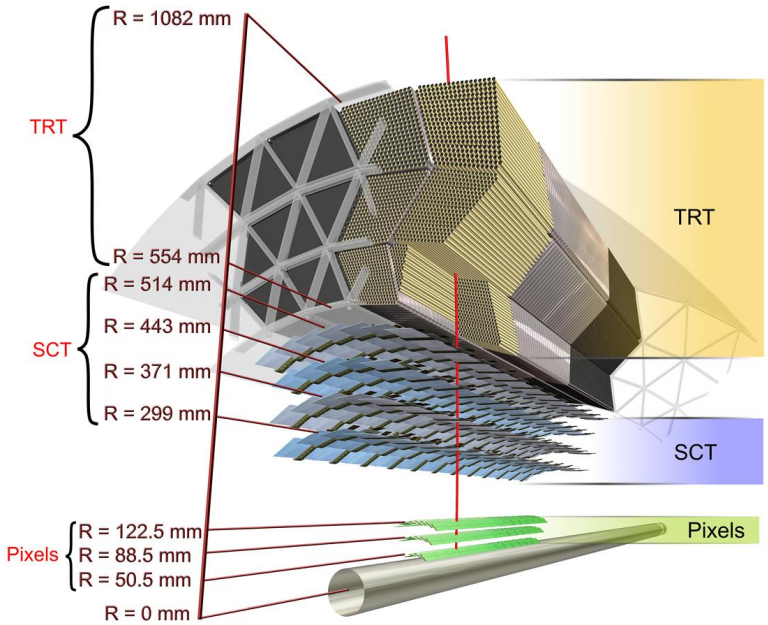


Figure 3.5: Tracking of charged particles in the inner detector [33].

The Transition Radiation Tracker

The outermost subdetector of the inner detector is the transition radiation tracker. In this region 175424 thin-walled proportional drift tubes with a diameter of 4 mm are used to obtain continuous tracking - each particle in the barrel region crosses around $35 - 40$ tubes. The tubes provide a spatial resolution of ≈ 130 μm, and also measure the transition radiation emitted by charged particles crossing the tube boundaries. This transition radiation is sensitive to e/m and therefore helps in the discrimination of electrons and pions [38].

Figure 3.5 summarises the subsystems of the inner detector and illustrates the relative sizes and positions in the barrel region. At 115 cm from the beam axis, the inner detector is bounded by the cryostat of the liquid argon calorimeter.

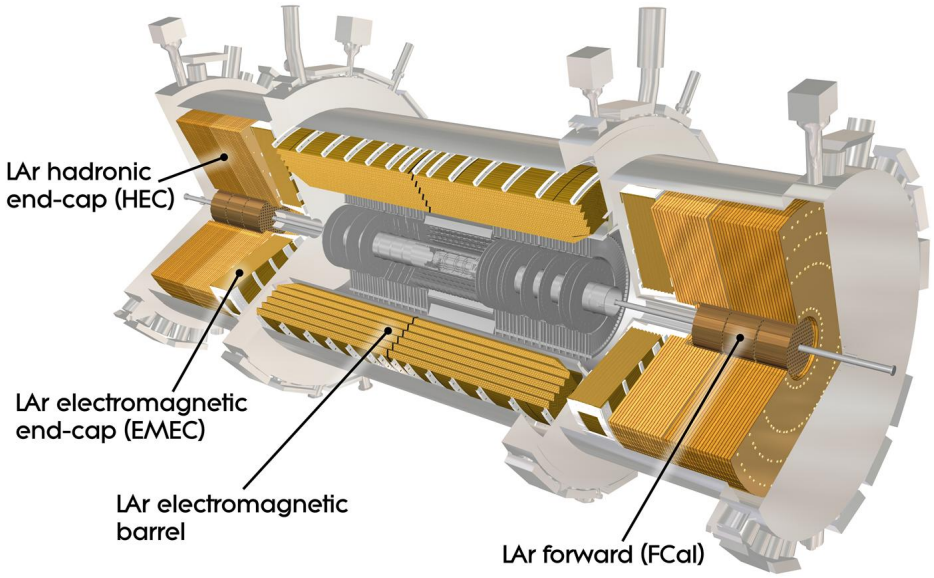


Figure 3.6: Electromagnetic calorimeter systems [33].

3.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (Figure 3.6) is a finely segmented, non-compensating sampling calorimeter, using an accordion lead structure as absorber and liquid argon as an active material. It covers the pseudorapidity region $|\eta| < 3.2$, the barrel section extending to $|\eta| < 1.475$ and the endcaps covering the region of $1.375 < |\eta| < 3.2$. In the transition region of barrel and endcap the electron identification is currently problematic, therefore, electrons falling in this region are currently rejected in analyses. It is segmented in radial and azimuthal regions as well as in depth (see Figure 3.7). The innermost thin layer with a granularity of $\Delta\eta \times \Delta\phi = 0.0031 \times 0.1$ makes it possible to distinguish $\pi^0 \to \gamma\gamma$ decays and single photons. Together with the second, main layer with $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, this enables a rough reconstruction of the particle direction even without tracking. The third layer with a granularity of $\Delta\eta \times \Delta\phi = 0.05 \times 0.025$ is used to check the shower shape and protects against misidentification of hadronic showers as electrons or photons. The design resolution of the electromagnetic calorimeter, verified by initial measurements, is $\frac{\Delta E}{E} = 10\%/\sqrt{E[\text{ GeV}]} \oplus 0.7\%$ [40].

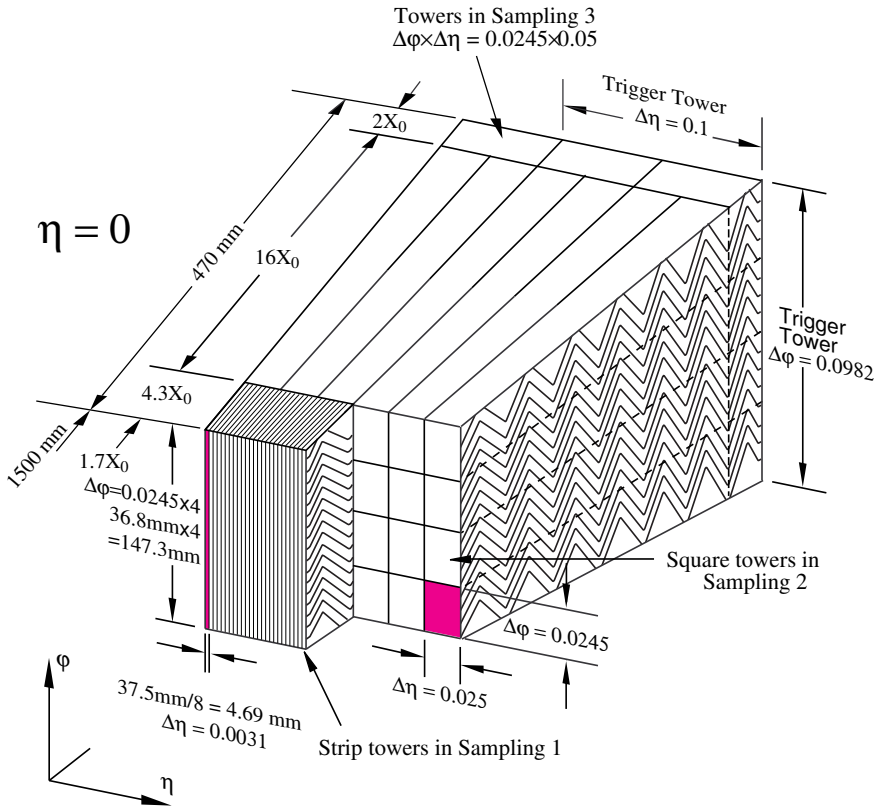


Figure 3.7: Electromagnetic calorimeter segmentation [39].

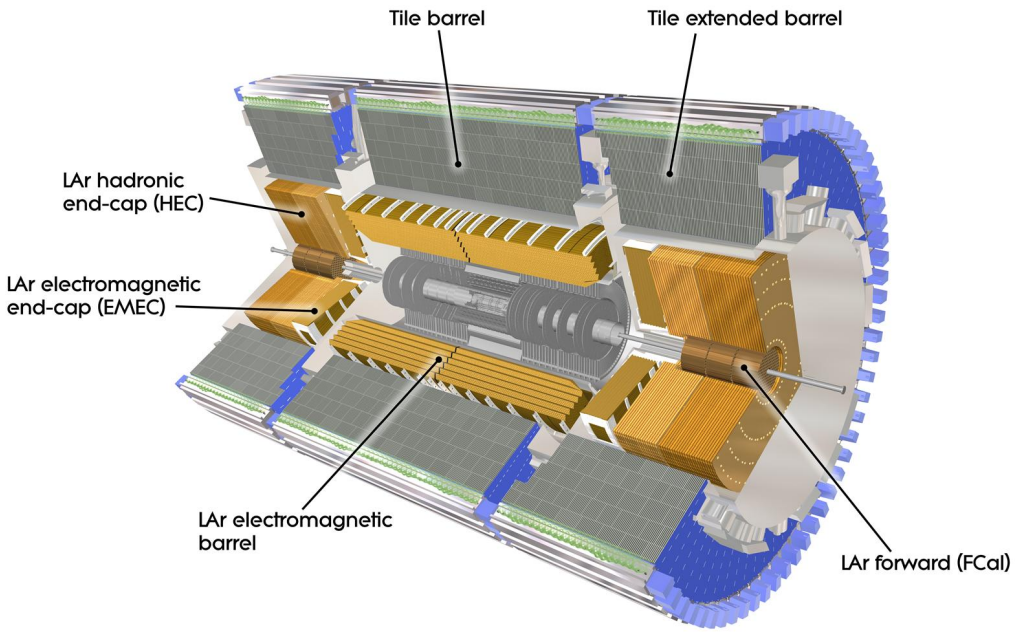


Figure 3.8: Hadronic calorimeter systems [33].

3.2.4 Hadronic Calorimeter

Around the electromagnetic calorimeter barrel wraps the hadronic tile calorimeter, visualised in Figure 3.8. It consists of interleaved iron and plastic scintillator plates, segmented into wedges of $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.1$. The tile calorimeter consists of the central barrel with $|\eta| < 1.0$ and two extended barrels with $0.8 < |\eta| < 1.7$. The readout, using wavelength-shifting fibres to transport the signal to photomultipliers (see Figure 3.9), is separated into three layers, making it possible to follow the development of the hadronic shower. It occupies the radial region between 2.3 and 4.3 m. At $\eta = 0$, a particle crossing the calorimeter passes material corresponding to 9.7 nuclear interaction lengths for proton[5]. The design resolution of the tile calorimeter is $\frac{\Delta E}{E} = 50\%/\sqrt{E[\text{ GeV}]} \oplus 3\%$, confirmed by initial measurements [41]. Built into the tile calorimeter are the minimum bias trigger scintillators (MBTS), which consist of 16 large scintillator plates on each side of the central barrel. These can be used to time and trigger on collision events in the first years of LHC running.

To cover the region of $|\eta| > 1.7$, a hadronic liquid argon calorimeter is built into the endcaps, sharing cryostats with the electromagnetic liquid

[5]A nuclear interaction length is defined as the distance after which only $1/e$ of the original relativistic particles are present.

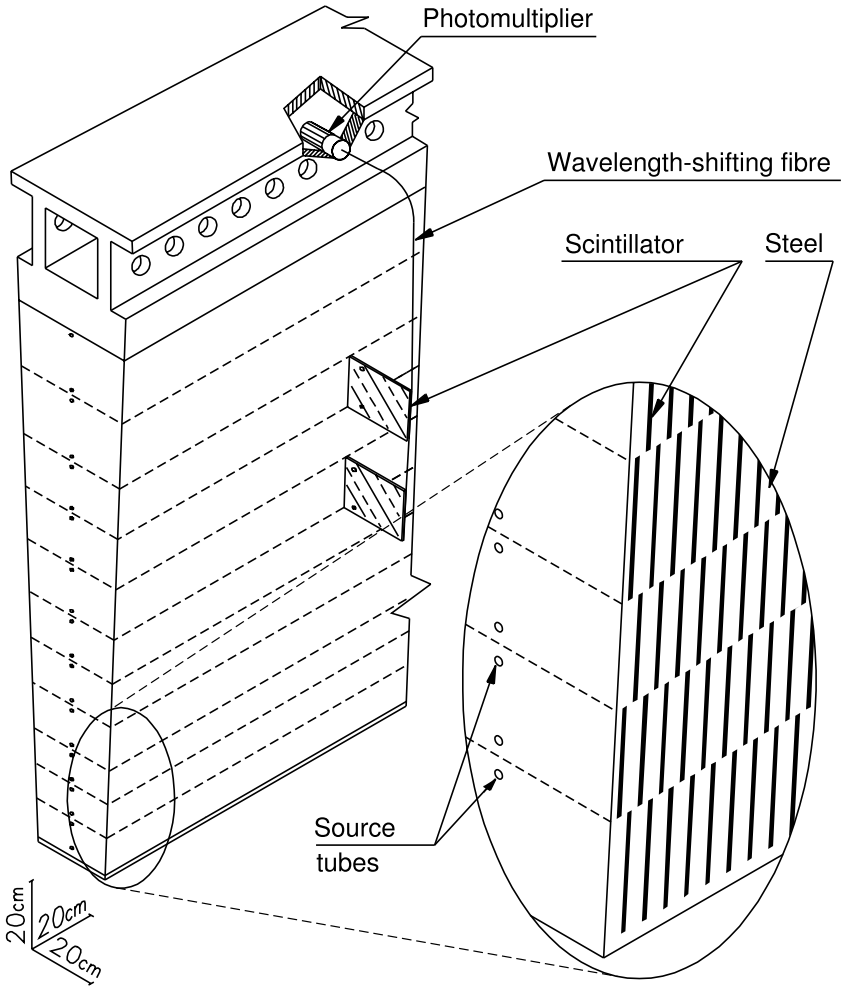


Figure 3.9: Hadronic tile calorimeter segmentation [41].

argon calorimeters, covering $1.5 < |\eta| < 3.2$. Its design is similar to the electromagnetic calorimeter, with copper being used instead of lead as absorber material. In the same endcap, a forward liquid argon calorimeter with absorbers of copper and tungsten measures both electromagnetic and hadronic showers, reaching up to $3.1 < |\eta| < 4.9$. This ensures high coverage and detection of all particles leaving the interaction region [39].

3.2.5 Muon Spectrometer

The only charged particles that usually escape the calorimeter systems are muons. Due to their high mass compared to electrons, bremsstrahlung losses are suppressed, and in contrast to hadrons they do not participate in strong interactions. The Muon Spectrometer (MS) detects and triggers on the escaping eponymous muons, and is able to measure their momentum at high

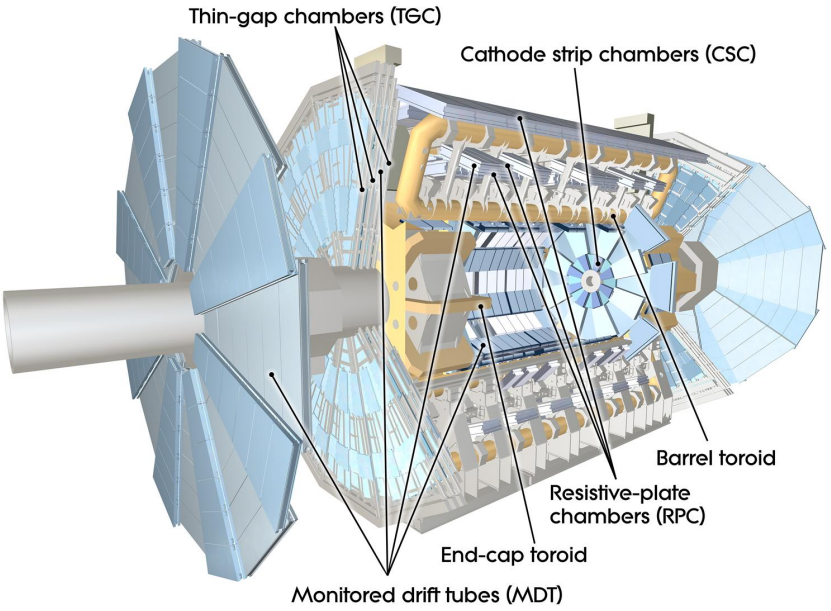


Figure 3.10: Subdetectors of the Muon Spectrometer [33].

accuracy. From the four subsystems, the resistive plate chambers (RPC) and thin gap chambers (TGC) are able to deliver quick trigger signals to the trigger system, whereas the monitored drift tubes (MDT) and cathode strip chambers (CSC) perform high-precision tracking and momentum measurement [42].

In Figure 3.10 the layout and the components of the muon system are visualised. In the barrel region, the MDT and RPC subsystems are consolidated into three stations, which are positioned 5, 7.5 and 10 m from the beam axis. The MDT system is used together with the TGC trigger system in the large endcap wheels at distances of 7.4, 10.8 and 14 m from the interaction point. The CSC system is only used close to the beam, since it copes better with high rates of charged particles expected there.

The large distances of the muon stations provide a good lever arm and large integration length in the magnetic field for the measurement of very high momenta. This results in a good energy resolution for muons up to TeV scale, design values ranging from 3% at 100 GeV to 10% at 1 TeV [32]. The measured resolution using collision data is currently $\approx$ 3.7% at 100 GeV to 17% at 1 TeV [43, 44] due to preliminary alignment and calibration constants.

Monitored Drift Tubes

The MDT subdetector of the muon spectrometer consists of aluminium drift tubes with a diameter of 30 mm. Central tungsten-rhenium wires at 3080 V

provide a sufficient electric field to cause measurable charge avalanches if a charged muon traverses the tube. The drift gas mixture of Argon-CO_2 has been chosen for its good ageing properties, depositing little material on the wire. For each muon track in the system, an autocalibration procedure is used to infer the radius-time relation. The resolution of a single tube is $60 - 250$ mm, decreasing with increasing radius [32].

Cathode Strip Chambers

Since the rate of charged particles in the forward region of the detector is too high for MDTs - greater than 150 Hz/cm^2 - multiwire proportional chambers with segmented cathodes at high granularity are used, capable to record with up to 1000 Hz/cm^2. Positions of charged particle tracks are measured by comparing charges deposited in adjacent strips, resulting in a resolution of up to 65 μm [32].

Resistive Plate Chambers

To provide fast trigger signals, the three layers of RPCs in the barrel are used. Charged particles are triggered by coincidence in successive stations, with low-p_T and high-p_T signals derived from the deflection of the particle in the toroidal field. RPCs do not use wires, but two charged parallel plates separated by a 2 mm gas-filled gap. The readout is segmented into strips. The time resolution of the RPCs is less than 2 ns, enabling both exact triggers and also timing-based rejections of muons from cosmic rays [32, 45].

Thin Gap Chambers

TGCs, used as trigger system in the endcaps, are again multiwire proportional chambers, with anode wires between two parallel cathode layers separated by 2.8 mm. The small distance and the wire spacing of 1.8 mm allows a good time resolution and fast response, which is synchronised with the LHC clock and does not allow time-of-flight measurements [32].

3.2.6 Forward Detectors

In addition to the main ATLAS detector, several special-purpose detectors also share the interaction point:

Beam Condition Monitor

The diamond beam condition monitor BCM [46] provides fast luminosity per bunch at a distance of 184 cm from the interaction point. It also provides a beam abort signal in the case that unusually high radiation is measured.

LUCID

The LUCID[6] detector is 17 m from the interaction point. It measures the particles arising in proton-proton collisions with low momentum transfer, where the deflection angle is very small, using these to provide a further measurement of the luminosity. In contrast to the BCM, LUCID provides only an integrated measurement, and does not provide per-bunch luminosity.

ALFA

In the very forward region, 240 m from the interaction point, the ALFA[7] detector provides a measurement of absolute luminosity, only possible in special runs where the number of interactions is low [47].

Zero Degree Calorimeter

The ATLAS zero degree calorimeter is a special-purpose detector, located at zero degrees with respect to the beams, 140 m from the interaction point. It detects neutral particles produced in collisions with $|\eta| > 8.3$, and is used in both heavy-ion and proton collisions to examine this very special region of phase space [48].

3.3 Trigger and Data Acquisition

At the LHC in 2011 the frequency of proton-proton collisions occur with a frequency of 20 MHz with nine simultaneous interactions on average. Since the size of the data recorded for an event is ≈ 1.4 MB, recording all events is not feasible. Instead a system of hardware and software triggers is applied to reduce the rate of events stored to ≈ 600 Hz. An overview of the ATLAS trigger and data acquisition system is given in Figure 3.11.

3.3.1 Hardware Triggers and Readout

The first level (Level 1) of the trigger system is implemented in hardware. Custom hardware receives inputs from the calorimeter and from muon detectors, and decides in ≈ 1 μs to accept or reject an event. If a decision is made to accept an event, the data is collected from the pipeline memories of the subdetectors and delivered to the read out drivers (RODs), situated in the counting room in a cavern next to the ATLAS detector. The Level 1 reduces the rate of events to ≈ 100 kHz.

[6] LUminosity Cerenkov Integrating Detector
[7] Absolute Luminosity For ATLAS

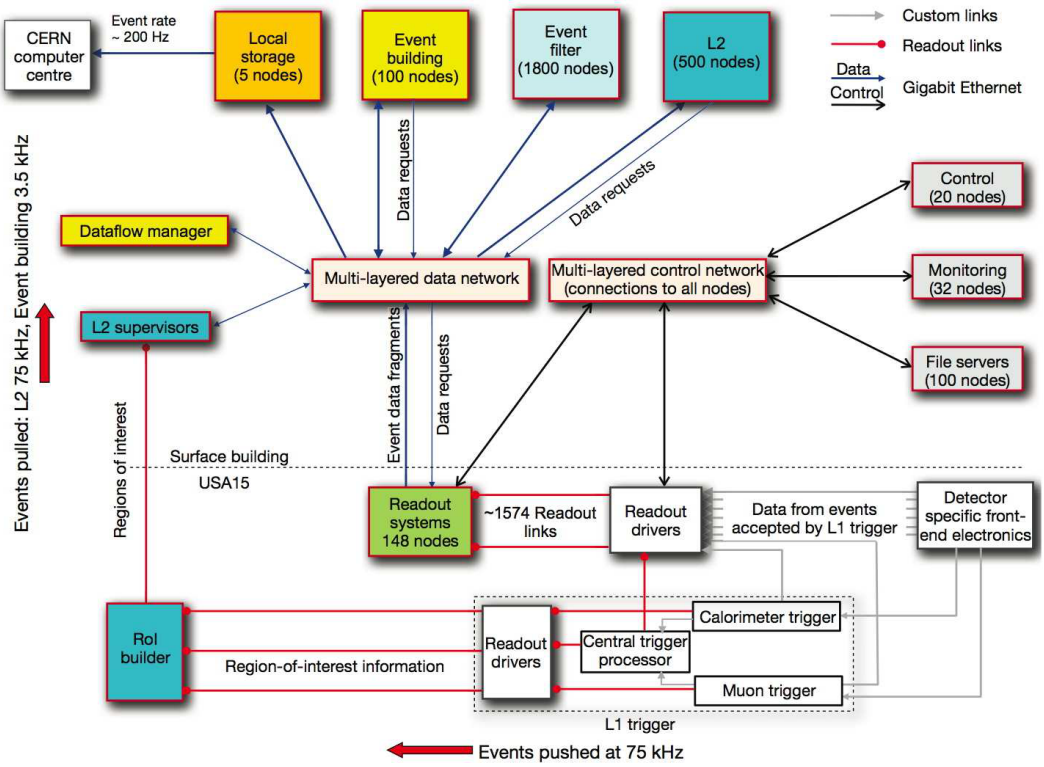


Figure 3.11: Diagram of the ATLAS trigger and data acquisition system [32].

3.3.2 Software Triggers

At this point, the RODs process and pass the event information to the read out buffers, which can be accessed by the Level 2 trigger system on the surface. This system is now purely software based. To cope with the still very high rate from Level 1, the Level 2 trigger software is seeded with the regions of interest identified by the Level 1 hardware, and inspects data in this region only to confirm or reject the Level 1 decision. This reduces the rate to $\approx$ 4 kHz. Events passing the Level 2 trigger system are then passed to the high-level trigger (HLT) computing cluster, which proceeds to do a fast reconstruction of the events using the regular ATLAS reconstruction software. This system reduces the rate to the final $400 - 600$ Hz, which are then stored on tape and at the same time distributed to the international system of ATLAS data centres, where the events are then re-reconstructed and further analysed. The final data rate from the ATLAS detector data acquisition system is therefore ≈ 1 GB/s.

3.4 Computing Tools and Infrastructure

Storing, distributing and analysing the large amount of data produced at the ATLAS experiment is a formidable challenge. In addition, a comparable amount of data simulated using Monte Carlo methods has to be produced - taking several minutes per event - and analysed as well, in a manner as similar to real data as possible. The ATLAS software framework Athena (Section 3.4.1) provides a flexible system for these problems, whereas the large computing power necessary to simulate and analyse data is harnessed using the tools of grid computing (Section 3.4.2).

3.4.1 Event Processing

The main toolset for the ATLAS experiment is the Athena software framework, with over six million lines of code. It is based on the Gaudi analysis framework [49]. Gaudi provides the tools to write modular algorithms, tools and services in the C++ programming language, and to combine and configure them using the Python scripting language. Athena provides core functions for ATLAS data analysis:

- Reconstruction of simulated and recorded data
- ATLAS geometry and magnetic field modelling
- Data storage and retrieval in a flexible format, using a transient/persistent interface
- Interface to Monte Carlo simulation tools
- A version of the Geant4 [50] simulation toolkit to simulate the passage of particles through the ATLAS detector
- Digitisation of Geant4 results producing simulated ATLAS data

The transient/persistent interface and the data storage system is very flexible and can dynamically store data generated by algorithms or user programs. However, the overhead per event does make the fast, repeated analysis of all events quite slow for simple analyses. For physics analysis, it is therefore common to produce derived data sets with reduced event size and process them using the ROOT analysis toolkit [51]. For this thesis, however, a new data storage format based on the Google protocol buffer library was designed. In addition, a powerful set of processing tools has been developed, including an analysis and histogramming library. This toolset, called A4, is used to generate and process derived data sets. Details of A4 can be found in Appendix B.

3.4.2 Grid Computing

Since the large amount of data produced by ATLAS and the other LHC experiments makes it unfeasible to process all data at CERN in a single data centre, the LHC computing Grid has been designed and deployed. It consists of a multitude of computing centres, connected by a common organisation and LCG middleware. The ATLAS computing Grid is organised in a hierarchy of four tiers: CERN as Tier0, the origin of experimental data, eleven special Tier1 data centres with dedicated connections to CERN, receiving, storing and re-processing the data, and also taking a central role in Monte Carlo simulation. Each Tier1 site has an associated "cloud" of Tier2 and Tier3 centres, which provide processing power for simulation and analysis and contain derived data, but do not generally have a copy of the full ATLAS data in the original format.

Using the Grid infrastructure is achieved using "Grid User Interfaces" like Ganga [52, 53], programs which assist in preparing analysis jobs and submit them to the grid. One of the policies is to send these analysis jobs to the centres that hold the data they are analysing, rather than transferring large amounts of data on demand. For this thesis, the GangaTasks utility described in Appendix A for automating the submission and distribution of jobs and the collection of the analysis results has been developed and used.

The grid infrastructure has so far worked remarkably well, and enabled the distributed analysis of the approximately 1 billion (10^9) recorded and simulated events used in this thesis.

Chapter 4

Measurement Overview

The precise measurement of the W^+W^- production cross-section is an important test of the Standard Model (SM): comparison with the SM prediction can constrain additional terms in the Lagrangian that could arise from new physics. Understanding the W^+W^- production process in detail is also vital in searches where W^+W^- production is a background, not least the search for the SM Higgs Boson in the $H \to W^+W^-$ decay channel.

4.1 Signature of W^+W^- decays

The main contribution to W^+W^- production at the LHC is from a quark-antiquark initial state as shown in Figure 4.1. The SM prediction at next-to-leading order (NLO) for this process is $47.0^{+2.0}_{-1.5}$ pb [25]. Due to the high energy of the protons, contributions from a gluon-gluon initial state as shown in Figure 4.2 are also not negligible, contributing $\approx 3\%$ to the event rate.

The W boson decay fractions are 10.5% each for decays into $e\nu_e$, $\mu\nu_\mu$ and $\tau\nu_\tau$, and 68.5% for decays in quark-antiquark pairs [2]. A pair of W bosons can therefore decay either into four quarks (47%), two quarks and an $\ell\nu$ pair (43%) or into two $\ell\nu$ pairs (10%). Since the four jet QCD rate and the direct production of Ws with associated jets are both large with respect to the W^+W^- cross-section, only the decays of W bosons into leptons are used for

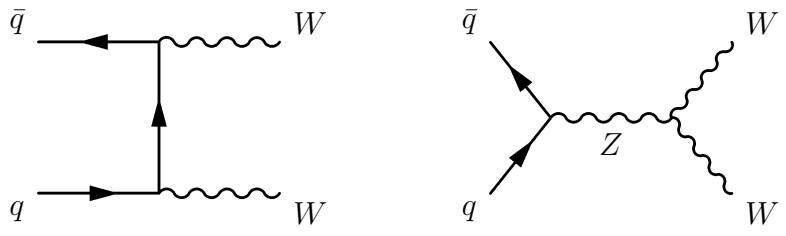


Figure 4.1: Leading order W^+W^- diboson production diagrams.

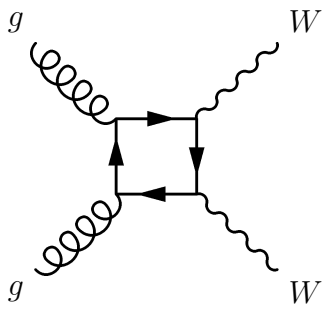


Figure 4.2: W^+W^- production diagram from gluon-gluon initial state.

the measurement. Since electrons and muons have different reconstruction and trigger efficiencies, and some background processes are specific to one flavor combination, the analysis is split into three "channels": two observed electrons ee, two observed muons $\mu\mu$, and one observed electron and one observed muon $e\mu$. Since it is difficult to reliably reconstruct tau leptons directly, only the tau decays into electrons or muons are considered directly and subsumed into the respective channel. In all channels a significant amount of missing transverse energy $\not\!\!E_T$ from the two neutrinos is expected.

The main background processes that can produce similar signatures are:

- Z boson decays with associated jets, where mismeasurements or tau lepton decays cause significant missing transverse energy $\not\!\!E_T$.
- Production of $t\bar{t}$ or a single top associated with a W boson. To remove these events, it is required that no additional jet is present in the event.
- W boson decays with associated jets or photons, if one jet or photon is mistakenly identified as a second lepton, or a non-prompt lepton is produced in the decay of heavy-flavor jets.
- WZ production if a lepton is not reconstructed in $WZ \to \ell\nu\ell\ell$ decays.
- $ZZ \to \ell\ell\nu\nu$ decays where the leptons are not close to the Z peak.

4.2 Summary of Event Selection Criteria

To suppress the background processes while at the same time selecting real W^+W^- decays with good efficiency, the following selection criteria for the signal region $\ell\ell + \not\!\!E_T$ are applied:

- Exactly two isolated, oppositely charged leptons with a transverse momentum p_T of at least 15 GeV. In the ee and $\mu\mu$ channels, the leading lepton is required to have a $p_T > 25$ GeV, in the $e\mu$ channel the electron

must satisfy this criterion. This difference in selection is justified by the high rate of jets misidentified as electrons in the region of p_T between 15 and 25 GeV, which would lead to large uncertainties.

- Two same-flavor leptons (ee and $\mu\mu$ channel) must have an invariant mass $m_{\ell\ell}$ greater than 15 GeV to suppress contributions from Υ decays. The invariant mass must also lie outside a window of 15 GeV around the Z mass to reduce the background from Z decays.

- In the $e\mu$ channel an invariant mass $m_{\ell\ell} > 10$ GeV is required to exclude the low-mass spectrum not modelled well in MC simulation.

- At least $55, 50, 25$ GeV of $\not{E}_{T,rel}$ are required in the ee, $\mu\mu$ and $e\mu$ channel respectively. This suppresses the contribution from Z decays. The variable $\not{E}_{T,rel}$ is defined as the perpendicular component of the missing transverse energy $\not{E}_T$ to the closest identified lepton or jet if that is closer than $\pi/2$ in the ϕ plane, otherwise it is defined equal to $\not{E}_T$.

- Veto the event if a jet with $E_T > 25$ GeV and $|\eta| < 4.5$ or a jet that has a b-tag with $E_T > 20$ GeV is present. This effectively suppresses events from $t\bar{t}$ decays, at the cost of introducing a dependence on the jet energy scale.

- The subleading lepton must have a transverse momentum $p_T > 20$ GeV. This requirement further reduces the background due to W+jets.

A summary of these cut requirements is given in Table 4.1. The details of the analysis objects - electrons, muons, jets and missing transverse energy - and further details on the event selection are explained in Section 6.

4.3 Cross-Section Measurement

To measure the total cross-section of the process $pp \to W^+W^-$ the following formula is used:

$$\sigma_{W^+W^-} = \frac{N - B}{\mathcal{L} \cdot \mathcal{A} \cdot \epsilon \cdot \mathrm{BR}} \tag{4.1}$$

All the parameters in the formula need to be measured to calculate the cross-section. The total observed number of events in a signal region N is obtained by counting selected events, and the contribution from background processes in this region B is estimated using data-driven methods, where possible, and MC simulations otherwise. The total luminosity $\mathcal{L}$ delivered to the experiment is 4.70 ± 0.17 fb^{-1} (see Section 5.2). The acceptance of the signal region $\mathcal{A}$ is usually evaluated together with ϵ as one number, $\mathcal{A}\epsilon$, and treated as a total efficiency. In MC simulations, $\mathcal{A}\epsilon$ corresponds to the fraction of initially generated events passing the object selection. The efficiencies ϵ

$ee + \not{E}_{T,rel}$	$\mu\mu + \not{E}_{T,rel}$	$e\mu + \not{E}_{T,rel}$
Exactly two oppositely charged leptons with $p_T > 15$ GeV		
Leading Lepton $p_T > 25$ GeV		Leading Electron $p_T > 25$ GeV
Invariant mass $m_{\ell\ell} > 15$ GeV		$m_{\ell\ell} > 10$ GeV
$\lvert m_{\ell\ell} - m_Z \rvert > 15$ GeV		-
$\not{E}_{T,rel} > 55$ GeV	$\not{E}_{T,rel} > 50$ GeV	$\not{E}_{T,rel} > 25$ GeV
No jets with $E_T > 25$ GeV and $\lvert\eta\rvert < 4.5$		
No b-jets with $E_T > 20$ GeV		
Subleading Lepton $p_T > 20$ GeV		

Table 4.1: Summary of signal event selection criteria for the three channels.

are mainly estimated using MC simulation, but are corrected to data using factors provided by the ATLAS performance groups. Finally, BR denotes the branching fraction of W^+W^- decays into fully leptonic modes, given in [2].

Chapter 5

Data Sets

5.1 Monte Carlo Simulation

For the simulation of signal and background processes, several different Monte Carlo (MC) generators are used. For all simulated events, the response of the detector is calculated using a detailed computer model of the ATLAS detector [54] using the Geant4 program [50].

The main contribution to W^+W^- production, from initial states with two incoming quarks, is simulated by the MC@NLO 4.01 [55] generator interfaced with the Herwig 6.520/Jimmy 4.31 [28] programs. This combination allows incorporation of the NLO QCD matrix elements into the parton shower. With this generator, especially important for $H \to W^+W^-$ background studies, the spin correlation between the two W bosons is treated correctly. For the gluon contribution to W^+W^- production the special gg2WW generator [56] is used. A full list of signal samples is provided in Table 5.1.

The samples for the background processes of W and Z production with associated jets, heavy flavor jets or photons are simulated using the Alpgen 2.13 generator [57], also interfaced with Herwig/Jimmy. The Alpgen generator can also produce events with additional partons in the final state. Therefore, these samples are labelled by "NpX", where X is the number of additional partons. The overlap between the heavy flavor samples and the regular samples is removed using the heavy flavor overlap removal procedure described in [58]. Top pair production is simulated with MC@NLO and Herwig/Jimmy, single top as well as Wt production by AcerMC [59]. Diboson production of WZ and ZZ is modelled directly by Herwig, and background from di-jets with heavy flavor is modelled using the PythiaB [60] fork of the Pythia generator [29].

Additional background to W^+W^- production can arise by direct decay of an off-shell photon associated with a W boson into a di-lepton pair [61], which is not simulated in the Alpgen samples of W+photon. This background is simulated using a modified MadGraph that supports massive photons coupled

process	cross-section [fb]	filter eff.	N_{MC}
$q\bar{q} \to W^+W^- \to e^+\nu e^-\bar{\nu}$	501	1.0	199949
$q\bar{q} \to W^+W^- \to e^+\nu \mu^-\bar{\nu}$	501	1.0	200000
$q\bar{q} \to W^+W^- \to e^+\nu \tau^-\bar{\nu}$	501	1.0	200000
$q\bar{q} \to W^+W^- \to \mu^+\nu \mu^-\bar{\nu}$	501	1.0	199000
$q\bar{q} \to W^+W^- \to \mu^+\nu e^-\bar{\nu}$	501	1.0	199949
$q\bar{q} \to W^+W^- \to \mu^+\nu \tau^-\bar{\nu}$	501	1.0	200000
$q\bar{q} \to W^+W^- \to \tau^+\nu \tau^-\bar{\nu}$	501	1.0	499676
$q\bar{q} \to W^+W^- \to \tau^+\nu e^-\bar{\nu}$	501	1.0	199950
$q\bar{q} \to W^+W^- \to \tau^+\nu \mu^-\bar{\nu}$	501	1.0	200000
$gg \to W^+W^- \to e^+\nu e^-\bar{\nu}$	14.5	0.9895	10000
$gg \to W^+W^- \to e^+\nu \mu^-\bar{\nu}$	14.5	0.9899	10000
$gg \to W^+W^- \to e^+\nu \tau^-\bar{\nu}$	14.5	0.9232	10000
$gg \to W^+W^- \to \mu^+\nu \mu^-\bar{\nu}$	14.5	0.9890	9999
$gg \to W^+W^- \to \mu^+\nu e^-\bar{\nu}$	14.5	0.9869	10000
$gg \to W^+W^- \to \mu^+\nu \tau^-\bar{\nu}$	14.5	0.9288	10000
$gg \to W^+W^- \to \tau^+\nu \tau^-\bar{\nu}$	14.5	0.9289	10000
$gg \to W^+W^- \to \tau^+\nu e^-\bar{\nu}$	14.5	0.9219	10000
$gg \to W^+W^- \to \tau^+\nu \mu^-\bar{\nu}$	14.5	0.3269	10000

Table 5.1: Monte Carlo samples of W^+W^- production processes. The filter efficiency given is the efficiency of the generator level requirements on leptons imposed on events. N_{MC} is the number of simulated MC events.

with Pythia 6.425 [29] as the parton shower generator.

All MC samples have been produced in the 2011 ATLAS MC production campaign "MC11c" [62]. Even though different generators are used, several parameters are kept constant: The masses and widths of particles are taken from the PDG 2010 [2], and the related ATLAS Underlying Event Tunes 1 (for MC@NLO) [63], 2 (for Alpgen, Herwig and Jimmy) [63] and 2B (for Pythia) [64] are used. The NLO parton distribution function provided to MC@NLO is CT10 [65], the leading-order (LO) PDF for Alpgen is CTEQ6L1. Herwig/Jimmy as well as Pythia use the LO** PDF MRSTMCal.LHgrid [23]. Finally, samples generated by leading order generators are scaled by k-factors to NLO or NNLO (if available) calculations [27].

A list of all MC background samples, their cross-sections and k-factors is given in the following Table 5.2.

generator	process	cross-section [pb]	k-factor	ϵ_{filter}	N_{MC}
Herwig	ZZ	4.6	1.41	0.21152	249999
Herwig	WZ	3.4	1.58	1.0	999896
Alpgen	$Z \to ee$, Np0	668.3	1.25	1.0	6618284
Alpgen	$Z \to ee$, Np1	134.4	1.25	1.0	1334897
Alpgen	$Z \to ee$, Np2	40.5	1.25	1.0	2004195
Alpgen	$Z \to ee$, Np3	11.2	1.25	1.0	439949
Alpgen	$Z \to ee$, Np4	2.9	1.25	1.0	149948
Alpgen	$Z \to ee$, Np5	0.8	1.25	1.0	50000
Alpgen	$Z \to \mu\mu$, Np0	668.7	1.25	1.0	6615230
Alpgen	$Z \to \mu\mu$, Np1	134.1	1.25	1.0	1334296
Alpgen	$Z \to \mu\mu$, Np2	40.3	1.25	1.0	1999941
Alpgen	$Z \to \mu\mu$, Np3	11.2	1.25	1.0	549896
Alpgen	$Z \to \mu\mu$, Np4	2.8	1.25	1.0	150000
Alpgen	$Z \to \mu\mu$, Np5	0.8	1.25	1.0	20000
Alpgen	$Z \to \tau\tau$, Np0	668.4	1.25	1.0	10613179
Alpgen	$Z \to \tau\tau$, Np1	134.8	1.25	1.0	3334137
Alpgen	$Z \to \tau\tau$, Np2	40.4	1.25	1.0	1004847
Alpgen	$Z \to \tau\tau$, Np3	11.2	1.25	1.0	509847
Alpgen	$Z \to \tau\tau$, Np4	2.8	1.25	1.0	144999
Alpgen	$Z \to \tau\tau$, Np5	0.8	1.25	1.0	45000
Alpgen	$Z \to \nu\nu$, Np0	3533.8	1.25	0.01	54949
Alpgen	$Z \to \nu\nu$, Np1	732.3	1.25	0.61	909848
Alpgen	$Z \to \nu\nu$, Np2	222.3	1.25	0.88	169899
Alpgen	$Z \to \nu\nu$, Np3	62.0	1.25	0.97	144999
Alpgen	$Z \to \nu\nu$, Np4	15.8	1.25	0.99	309899
Alpgen	$Z \to \nu\nu$, Np5	4.4	1.25	1.0	84999
Alpgen	$Z \to eebb$, Np0	6.6	1.25	1.0	409999
Alpgen	$Z \to eebb$, Np1	2.5	1.25	1.0	160000
Alpgen	$Z \to eebb$, Np2	0.9	1.25	1.0	60000
Alpgen	$Z \to eebb$, Np3	0.4	1.25	1.0	30000
Alpgen	$Z \to \mu\mu bb$, Np0	6.6	1.25	1.0	409949
Alpgen	$Z \to \mu\mu bb$, Np1	2.5	1.25	1.0	100000
Alpgen	$Z \to \mu\mu bb$, Np2	0.9	1.25	1.0	60000
Alpgen	$Z \to \mu\mu bb$, Np3	0.4	1.25	1.0	29999
Alpgen	$Z \to ee$, Np0 ($m_{ll} < 40$)	3051.6	1.22	0.0351	1939447
Alpgen	$Z \to ee$, Np1 ($m_{ll} < 40$)	87.9	1.22	0.48	999998
Alpgen	$Z \to ee$, Np2 ($m_{ll} < 40$)	41.4	1.22	1.0	999946
Alpgen	$Z \to ee$, Np3 ($m_{ll} < 40$)	8.4	1.22	1.0	149998
Alpgen	$Z \to ee$, Np4 ($m_{ll} < 40$)	1.9	1.22	1.0	40000
Alpgen	$Z \to ee$, Np5 ($m_{ll} < 40$)	0.5	1.22	1.0	10000
Alpgen	$Z \to \mu\mu$, Np0 ($m_{ll} < 40$)	3051.6	1.22	0.04	1934946
Alpgen	$Z \to \mu\mu$, Np1 ($m_{ll} < 40$)	87.9	1.22	0.49	999898
Alpgen	$Z \to \mu\mu$, Np2 ($m_{ll} < 40$)	41.5	1.22	1.0	999995
Alpgen	$Z \to \mu\mu$, Np3 ($m_{ll} < 40$)	8.4	1.22	1.0	150000
Alpgen	$Z \to \mu\mu$, Np4 ($m_{ll} < 40$)	1.9	1.22	1.0	39999
Alpgen	$Z \to \mu\mu$, Np5 ($m_{ll} < 40$)	0.5	1.22	1.0	10000
Alpgen	$Z \to \tau\tau$, Np0 ($m_{ll} < 40$)	3055.1	1.22	1.0	999649
Alpgen	$Z \to \tau\tau$, Np1 ($m_{ll} < 40$)	84.9	1.22	1.0	299999
Alpgen	$Z \to \tau\tau$, Np2 ($m_{ll} < 40$)	41.5	1.22	1.0	498899
Alpgen	$Z \to \tau\tau$, Np3 ($m_{ll} < 40$)	8.4	1.22	1.0	150000
Alpgen	$Z \to \tau\tau$, Np4 ($m_{ll} < 40$)	1.9	1.22	1.0	39999
Alpgen	$Z \to \tau\tau$, Np5 ($m_{ll} < 40$)	0.5	1.22	1.0	10000
PythiaB	$\mu 10 \mu 10$X	2830.0	1.0	1.0	296599
PythiaB	$\mu 10 e 10$X	4017.0	1.0	1.0	795695
PythiaB	$e10e10$X	1693.0	1.0	1.0	2920985
McAtNlo	$t\bar{t}$	164.6	1.0	0.55551	11584773
AcerMC	Wt	15.7	1.0	1.0	994897
AcerMC	t (t-channel) $e\nu$	6.9	1.0	1.0	758871
AcerMC	t (t-channel) $\mu\nu$	6.8	1.0	1.0	759784
AcerMC	t (t-channel) $\tau\nu$	7.3	1.0	1.0	845149
AcerMC	t (s-channel) $e\nu$	0.5	1.0	1.0	199899
AcerMC	t (s-channel) $\mu\nu$	0.5	1.0	1.0	199850
AcerMC	t (s-channel) $\tau\nu$	0.5	1.0	1.0	499500
Alpgen	$W \to e\nu$, Np0	6921.6	1.2	1.0	3458883
Alpgen	$W \to e\nu$, Np1	1304.3	1.2	1.0	2499645

generator	process	cross-section [pb]	k-factor	ϵ_{filter}	N_{MC}
Alpgen	$W \to e\nu$, Np2	378.3	1.2	1.0	3768632
Alpgen	$W \to e\nu$, Np3	101.4	1.2	1.0	1008947
Alpgen	$W \to e\nu$, Np4	25.9	1.2	1.0	250000
Alpgen	$W \to e\nu$, Np5	7.0	1.2	1.0	69999
Alpgen	$W \to \mu\nu$, Np0	6919.6	1.2	1.0	3462942
Alpgen	$W \to \mu\nu$, Np1	1304.2	1.2	1.0	2498593
Alpgen	$W \to \mu\nu$, Np2	377.8	1.2	1.0	3768737
Alpgen	$W \to \mu\nu$, Np3	101.9	1.2	1.0	1008446
Alpgen	$W \to \mu\nu$, Np4	25.8	1.2	1.0	254950
Alpgen	$W \to \mu\nu$, Np5	6.9	1.2	1.0	70000
Alpgen	$W \to \tau\nu$, Np0	6918.6	1.2	1.0	3418296
Alpgen	$W \to \tau\nu$, Np1	1303.2	1.2	1.0	2499194
Alpgen	$W \to \tau\nu$, Np2	378.2	1.2	1.0	3750986
Alpgen	$W \to \tau\nu$, Np3	101.5	1.2	1.0	1009946
Alpgen	$W \to \tau\nu$, Np4	25.6	1.2	1.0	249998
Alpgen	$W \to \tau\nu$, Np5	7.0	1.2	1.0	65000
Alpgen	Wbb, Np0	47.3	1.2	1.0	474997
Alpgen	Wbb, Np1	35.8	1.2	1.0	205000
Alpgen	Wbb, Np2	17.3	1.2	1.0	174499
Alpgen	Wbb, Np3	6.6	1.2	1.0	69999
Alpgen	Wcc, Np0	127.5	1.2	1.0	1274846
Alpgen	Wcc, Np1	104.7	1.2	1.0	1049847
Alpgen	Wcc, Np2	52.1	1.2	1.0	524947
Alpgen	Wcc, Np3	17.0	1.2	1.0	170000
Alpgen	Wc, Np0	650.0	1.2	1.0	6498837
Alpgen	Wc, Np1	205.0	1.2	1.0	2069646
Alpgen	Wc, Np2	50.8	1.2	1.0	519998
Alpgen	Wc, Np3	11.4	1.2	1.0	115000
Alpgen	Wc, Np4	2.8	1.2	1.0	30000
Alpgen/Herwig	$W\gamma$, Np0 pt20	213.1	1.432	1.0	1249998
Alpgen/Herwig	$W\gamma$, Np1 pt20	52.2	1.432	1.0	499998
Alpgen/Herwig	$W\gamma$, Np2 pt20	17.3	1.432	1.0	175000
Alpgen/Herwig	$W\gamma$, Np3 pt20	5.3	1.432	1.0	264999
Alpgen/Herwig	$W\gamma$, Np4 pt20	1.4	1.432	1.0	64999
Alpgen/Herwig	$W\gamma$, Np5 pt20	0.3	1.432	1.0	20000
MadGraph/Pythia	$W\gamma* \to l\nu ee$	4.8	1.0	1.0	294999
MadGraph/Pythia	$W\gamma* \to l\nu\mu\mu$	1.5	1.0	1.0	149900
MadGraph/Pythia	$W\gamma* \to l\nu\tau\tau$	0.2	1.0	1.0	50000

Table 5.2: Monte Carlo samples of background processes. The filter efficiency ϵ_{filter} is the efficiency of the generator level requirements imposed on events. N_{MC} is the number of simulated Monte Carlo events. NpX for X=0 − 5 specifies the number of additional partons in the event.

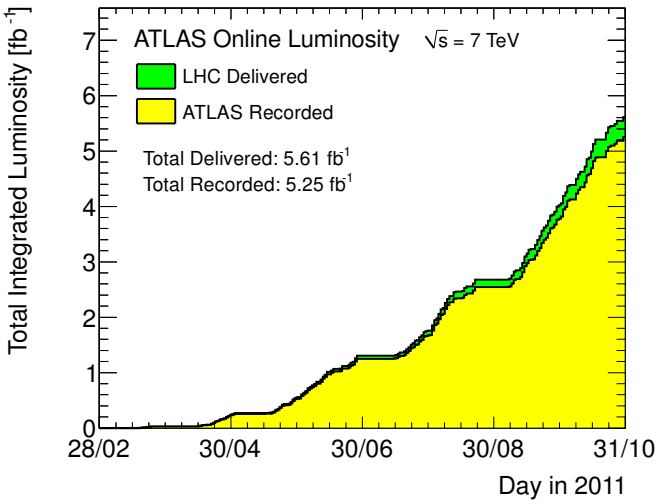


Figure 5.1: Integrated luminosity recorded by the ATLAS detector in the course of 2011 [67, 68].

5.2 Data from the ATLAS Experiment

Between the 15th of April and the 30th of October 2011 the ATLAS experiment recorded proton-proton collisions at $\sqrt{s} = 7$ TeV corresponding to a total integrated luminosity $\mathcal{L} = 5.2$ fb^{-1} (see Figure 5.1). During this period, comprising the ATLAS internal data periods D to M, the spacing between bunches in the LHC was 50 ns[1]. The runs in the time before the 15th of April - internally called periods A-C - had a larger bunch spacing and correspond to very little usable luminosity, and are therefore not used in this analysis. In total, 4701 pb^{-1} of integrated luminosity fulfil the data quality requirements of the ATLAS Standard Model group. The luminosity is determined from a list of good runs using the standard ATLAS luminosity tool [66]. The current uncertainty on the luminosity measurement from van der Meer scans is 3.7%, dominated by the uncertainty on the bunch charge [21]. This yields the integrated luminosity $\mathcal{L} = 4.70 \pm 0.17$ fb^{-1} for the data used in this analysis.

5.2.1 Event Cleaning

In addition to applying the data quality requirements of the Standard Model group in the form of a "good runs list", the following problems have to be addressed:

- In the periods E-H four front end boards (FEBs) of the liquid argon calorimeter (LAr) did not operate, leading to a blind spot in the region

[1] The design value of the LHC is 25 ns.

of $(0.0 < \eta < 1.45) \times (-0.78847 < \phi < -0.59213)$. Events in which a jet of more than 25 GeV[2] of transverse momentum is measured to intersect this region are rejected. In MC simulations, this procedure is applied to a percentage of events corresponding to the integrated luminosity affected by this problem. This procedure is called "LAr Hole Cleaning" and is recommended by the ATLAS Jet-Etmiss performance group [69]. In total, 0.6% of the events in data and 0.3% of the events in simulation are rejected by this procedure.

- Reconstructed hadronic jets can be "bad jets" which are not associated to real energy deposits in the calorimeters. Sources of bad jets can be hardware problems, beam conditions, and cosmic-ray showers [69]. If a bad jet with a transverse momentum p_T of more than 20 GeV is identified in the event and it does not overlap with an identified muon or electron[3], the whole event is rejected, since any $\not{E}_T$ measurement would be unreliable. In both data and MC simulation, 0.14% of events are affected.

- In irregular intervals, the liquid argon calorimeter experiences noise bursts [70]. These are automatically detected and a time period of one second after the first excessive noise occurrence is flagged as unsuitable for analysis [71]. The loss of luminosity is with 0.3% very small, but is accounted for in the luminosity calculation [66].

5.2.2 Triggers

The data from the ATLAS experiment is exported from the detector in several streams, according to the triggers they fired. For this analysis both the "egamma" and "muons" streams are used, containing all events that triggered any electromagnetic or muon trigger. In total, this amounts to 744 million events in all runs marked as "good". The nominal transverse momentum thresholds for the lepton triggers are 20 GeV (or 22 GeV, beginning with period K) for electrons and 18 GeV for muons. In Table 5.3 the triggers for each time period are given.

For each time period, these triggers are the *lowest unprescaled single-lepton triggers* listed in Table 5.3, having the lowest transverse momentum (p_T) threshold while still being recorded in full[4]. The numbers in the trigger names indicate the p_T threshold in GeV. All but the first muon trigger require a "medium" object quality. Since this object quality criterion is also

[2]This includes electrons, since they are always also initially reconstructed as a jet.

[3]"Jets" that are reconstructed from deposits caused by muons or electrons are often classified as "bad" hadronic jets.

[4]Prescaling with a value of n refers to the practice of only accepting one in n firings of the trigger, reducing the effective recorded luminosity for that trigger by $1/n$.

period	muon trigger	electron trigger
D-I	mu18_MG	e20_medium
J	mu18_MG_medium	e20_medium
K	mu18_MG_medium	e22_medium
L,M	mu18_MG_medium	e22vh_medium1

Table 5.3: Triggers used for selecting events in different time periods.

required during the object selection, there is no impact on the analysis. To reduce the high rate of electron triggers caused by high-energy hadronic jets an additional hadronic veto[5] requirement indicated by "vh" is added to the electron trigger in the two last periods L and M.

The trigger efficiencies of the muon triggers is cross-checked between simulation and data using a tag-and-probe approach by the muon trigger performance group [72]. The electron trigger and reconstruction efficiencies are determined by the electron performance group in [73]. Both studies yielded differences between data and simulation, and scale factors depending on p_T and η are calculated to correct the simulation. Scale factors are applied on the simulation event weight. In the same procedure, systematic uncertainties on the scale factors are obtained, which can be used to estimate the uncertainty on the trigger and reconstruction efficiency. The combined trigger scale factors on the signal sample are 0.999 for the ee, 0.998 for $e\mu$, and 0.995 for the $\mu\mu$ channel.

Trigger efficiencies for reconstructed muons with $p_T > 20$ GeV are about 80% in the endcap and 90% in the barrel region. The single electron trigger efficiency is close to 99% for reconstructed electrons with $p_T > 25$ GeV, reflecting the tight reconstruction requirements.

Since both leptons could have potentially triggered the event, the trigger scale factor for the event has to be calculated as $SF = \frac{1-(1-\epsilon_{1,data})\times(1-\epsilon_{2,data})}{1-(1-\epsilon_{1,MC})\times(1-\epsilon_{2,MC})}$, where the ϵ are the data/MC trigger efficiencies for the first and second lepton, respectively.

Finally, to protect against spurious trigger firings and to obtain a consistent picture, it is required that the trigger is actually fired by one of the reconstructed leptons, matching trigger objects inside $\Delta\phi < 0.15$ for electrons and $\Delta\phi < 0.1$ for muons. This requirement is not met in 0.55% of simulated and 0.57% of data events.

[5]That is, the item triggers only if the energy deposit in the hadronic calorimeter in the same region is not too high with respect to the deposit in the electromagnetic calorimeter.

Chapter 6

Physics Object Reconstruction and Identification

To obtain a good measurement at a collider experiment, a set of reconstructed "physics objects" must be defined and well understood. In Sections 6.1 to 6.5 the physics objects are defined: electrons, muons, (hadronic) jets, b-jets, and missing transverse energy $\not{E}_T$. Although photons can be reconstructed in the ATLAS detector, they are not used explicitly in this analysis but subsumed into jets. Also, no tau lepton identification is used, since it currently has low efficiency and is not well suited for rejecting backgrounds involving tau decays, for example decays of Z bosons into tau lepton pairs.

Section 6.6 describes how primary interaction vertices are reconstructed and counted - with on average 9 and up to 17 average interactions per bunch crossing, this becomes an important part of understanding the data, and several quantities related to physics objects have to be corrected to the amount of "pileup" activity in the event.

Finally, some measurements can be interpreted as two different physics objects. In Section 6.7 it is therefore specified which physics objects to prefer in case of overlap.

6.1 Electrons

Electron reconstruction in the ATLAS detector is seeded by clusters of energy deposited in the liquid argon (LAr) calorimeter. Such clusters are then matched with tracks reconstructed in the inner detector. On these objects "loose", "medium" and "tight" selection quality criteria are defined, using hadronic leakage, shower shape, track quality and track-cluster matching as discriminants [74]. For data taken during 2011, the criteria have been re-optimised leading to the new "loose++", "medium++" and "tight++" selections, since the high number of interactions per bunch crossing lead to a decrease in electron identification efficiency of up to 20% [75] with the ori-

ginal definitions. In this analysis, a high purity of reconstructed electrons is desired, therefore the "tight++" quality selection is used.

To reject non-prompt electrons produced inside hadronic jets, it is also required that the electron physics object is *isolated*, that is, the transverse sum of energy deposited in the calorimeter inside a cone of 0.3 in the $\eta - \phi$ plane around the electron $E_{T,cone30}$ must be less than 14% of the electron transverse energy, and the sum of transverse momentum of the tracks in this area $p_{T,cone30}$ must be less than 13% of the electron transverse energy. These values are optimised for high efficiency and good rejection, and are used as a common baseline for all diboson analyses and Higgs boson searches at the ATLAS experiment [76].

In summary, criteria for electrons in this analysis are:

- Reconstructed by the ATLAS electron algorithm [74]
- At least 15 GeV of transverse energy in the calorimeter
- Not in regions with LAr problems (in affected runs, see Section 5.2.1)
- Fully inside the calorimeter acceptance and outside the "crack" region between barrel and endcap ($|\eta| < 2.47$ and not $1.37 < |\eta| < 1.52$)
- Passes "tight++" selection
- The perigee of the track must lie within[1] $|\Delta z| < 1$ mm and[2] $\frac{\Delta d_0}{\sigma(d_0)} < 10$ of the reconstructed primary interaction vertex
- Isolated: $E_{T,cone30} < 0.13 \times p_{T,e}$ and $p_{T,cone30} < 0.14 \times p_{T,e}$

The reconstruction efficiencies of the electron physics objects defined in this way are determined by the e/γ performance group [73] using the tag-and-probe method on $Z \to e^+e^-$ samples. In total, the electron selection efficiency on the signal sample is $\approx 78\%$ for the central region with $|\eta| < 0.8$ and 64% in the forward region with $|\eta| > 2.0$, with uncertainties of $2 - 5\%$ dominated by the background uncertainty on the tag-and-probe samples.

Since the tag-and-probe method indicates that the simulated reconstruction and identification efficiency does not match the one observed in data, scale factors depending on η and E_T are calculated and applied, using the official tools provided by the e/γ performance group [73]. The numerical values of these scale factors lie between 0.98 and 1.13, the effect on the normalisation being only 0.02%. Since the scale factors do not include the isolation criteria used in this analysis, they are separately studied, but no significant deviation from unity is found [77]. The uncertainty on the isolation efficiency is on the order of $1 - 2\%$.

[1] Δz is the distance in the direction parallel to the beam

[2] Δd_0 is the distance in the plane perpendicular to the beam

6.2 Muons

For the reconstruction of muons in the ATLAS detector several approaches and algorithms are available [78]. In this analysis, the STACO group of algorithms is used: First, tracks are extrapolated from the outer layer of the muon spectrometer towards the interaction point and reconstructed as standalone tracks. These standalone tracks are subsequently extrapolated to the interaction vertex and there combined with tracks reconstructed in the inner detector using a chi-squared fit of track vectors and covariance matrices.

The combined muon track is then required to have a minimal number of hits in each subdetector [79], to ensure that there is little contamination from spurious hits and cosmic muons not intersecting the beam pipe.

In summary:

- Combined muon found by the STACO algorithm chain
- Transverse momentum $p_T > 15$ GeV
- Minimum number of hits in each subdetector
- The perigee of the track must lie within $|\Delta z| < 1$ mm and $\frac{\Delta d_0}{\sigma(d_0)} < 3$ of the reconstructed primary interaction vertex
- Isolated: $E_{T,cone30} < 0.14 \times p_{T,\mu}$ and $p_{T,cone30} < 0.15 \times p_{T,\mu}$

Similar to electrons, the isolation requirements are chosen to reject muons originating in heavy flavor decays inside jets, and are requirements common to ATLAS electroweak analyses. They have been optimised separately to the electron requirements, yielding small differences in the cut values.

The MC simulation of muon reconstruction and identification matches the data very well. The scale factors calculated by the Combined Muon Performance Group [79] are found to be close to unity, with an uncertainty of $1-2\%$. In addition, the isolation scale factors are estimated to be compatible with unity with an uncertainty of 0.5% [77]. The total effect of both scale factors on the normalisation is 0.2%.

6.3 Jets

Hadronic jets are reconstructed using the anti-k_T algorithm [80], the default jet reconstruction algorithm used in both ATLAS and CMS. This algorithm has the desirable property of producing jets that are very close to cone-like and - unlike the related k_T-clustering - not being overly sensitive to underlying activity and pileup, while still being infrared- and collinear safe. The resolution parameter is set to 0.4, approximately corresponding to the cone size in the $\eta - \phi$ plane. As input for the algorithm, topological calorimeter

clusters calibrated to the electromagnetic scale are used. The reconstructed jets are then corrected to the hadronic *Jet Energy Scale* (JES), derived from detailed MC simulation. The uncertainty on this calibration, which currently is one of the major systematic uncertainties, is estimated from di-jet balance in Z+jets events [81].

As discussed in Section 5.2.1, cuts that reject "bad jets" are imposed [69]. These cuts assure that the majority of the deposit is in the hadronic calorimeter, that both calorimeters report good quality data recording, and no coherent noise is registered in the electromagnetic calorimeter. Furthermore, they require that the deposit is not purely in the hadronic calorimeter as would be the case for some cosmic ray showers, and no single noisy cell has caused more than 99% of the deposit. These requirements are also applied to the jet selection, however, the presence of a "bad jet" also triggers the rejection of the whole event at a later stage.

Jets are required to have a minimum of 25 GeV of transverse energy E_T (at the hadronic energy scale), and to lie within $|\eta| < 4.5$. In addition, at least 75% of the scalar sum of the transverse momenta p_T of the tracks inside a 0.4 cone around the jet must originate from the primary interaction vertex, to suppress jets originating from other, simultaneous proton-proton interactions.

6.4 *b*-Jets

To identify jets originating from long-lived b hadrons, several identification algorithms are combined with a neural network into the "MV1" b-tagging algorithm [82]. The algorithms fed into the neural network try to determine if some of the tracks originate from a secondary vertex close to the primary, using the impact parameter ("IP3D" algorithm), try constructing a secondary vertex ("SV1"), or use a neural network themselves ("JetFitterCombNN"). In addition, the measured jet momentum is provided to the network.

Since tracking is necessary for all b-tagging algorithms, b-jets are only available in the inner detector acceptance region of $|\eta| < 2.5$. Scale factors for b-jets that calibrate mis-tagging rate and tagging efficiency to data are available from the performance group for jet $E_T > 20$ GeV, which is used as a lower limit instead of 25 GeV for non-b-jets. The effects of the scale factors on the MC prediction are $\approx 0.15\%$. For this analysis, b-jets are identified in addition to regular jets, and are only used to suppress the background from $t\bar{t}$ decays.

6.5 Missing Transverse Energy

Since the ATLAS detector has a coverage of almost the full 4π solid angle, it is possible to use momentum balance to measure the missing energy in the

event. Since the observed collisions have proton constituents with unknown Bjorken x (fraction of the proton momentum) in the initial state, only the missing energy in the transverse plane is expected to be zero if no neutrinos are present. This missing transverse energy $\not\!\!E_T$ is an important variable for distinguishing W^+W^- decays with neutrinos from other processes like Z decays into muon or electron pairs.

At the ATLAS experiment, the missing transverse energy is calculated using calorimeter clusters in the calorimeter acceptance region of $|\eta| < 4.9$ and the reconstructed muon momenta. Calorimeter clusters are calibrated according to the associated physics objects. Tracks from the inner detector are used to recover muons that lie in inefficient regions of the muon spectrometer. Low-momentum tracks not reaching the calorimeter are also considered separately. A detailed performance analysis is given by [83].

In this analysis, Z+jets production and decay to e or μ pairs is the most important background without neutrinos in the final state. It can be rejected by requiring a minimum $\not\!\!E_T$ in an event. Since in this process most $\not\!\!E_T$ arises due to limited resolution, a modified variable $\not\!\!E_{T,rel}$ that is less sensitive to lepton or jet mismeasurements can be defined:

$$\not\!\!E_{T,rel} = \begin{cases} \not\!\!E_T \times \sin(\Delta\phi_{\not\!\!E_T,l}) & \text{if} \Delta\phi_{\not\!\!E_T,l} < \pi/2 \\ \not\!\!E_T & \text{if} \Delta\phi_{\not\!\!E_T,l} \geq \pi/2 \end{cases} \tag{6.1}$$

Here $\Delta\phi_{\not\!\!E_T,l}$ denotes the difference of the azimuthal angle ϕ between $\not\!\!E_T$ and the closest selected lepton or jet projected on the x-y plane transverse to the beam. Any unexpectedly high energy loss or underestimate of a lepton or a jet momentum does not cause a high $\not\!\!E_{T,rel}$, leading to a reduced tail of high-$\not\!\!E_T$ events in Z+jets decays.

6.6 Vertices

Tracks from the inner detector are used to reconstruct a set of interaction vertices. The *primary vertex* is then selected as the one with the highest sum of p_T^2 from the associated tracks. This vertex is used for all the vertex cuts on z and d_0 as described in Sections 6.2-6.3. To remove cosmic ray and beam background, only reconstructed vertices with at least three tracks are considered.

The number of reconstructed vertices is also used to estimate the expected additional energy in the isolation cones for muons and electrons, and to reduce the dependency of the selection efficiency on the instantaneous luminosity.

6.7 Overlap Removal

There are several cases where real tracks or energy deposits can be reconstructed as multiple physics objects:

- Two tracks can point to the same calorimeter cluster, both reconstructed as electrons.
- A muon can leave a bremsstrahlung deposit in the calorimeter, which is reconstructed as an electron.
- An electron is also always reconstructed as a jet, since jets are also reconstructed from all calorimeter deposits.

Most important is the removal of the jets that overlap electrons: any jet that is closer than 0.3 in $\eta - \phi$ space to a selected electron is removed from consideration (e/j overlap removal). Since selected electrons are already required to be well isolated, no energy deposit is ignored, and the clustering of additional jets is not affected.

To remove "fake" electrons due to muon bremsstrahlung, any electron closer than 0.1 to the muon is removed (μ/e overlap removal). As in the e/j case, the muon isolation already ensures that this effect is rare, and if it is present, the deposited energy of the bremsstrahlung is small.

Finally, from any two electrons closer than 0.1 in $\eta - \phi$ space the one with lower transverse momentum is removed. Due to the tight quality criteria used for electrons, this effect is exceedingly rare.

Chapter 7

Event Selection

For the W^+W^- production cross-section measurement, three exclusive final states are considered: $e^+e^-\not{E}_{T,rel}$, $\mu^+\mu^-\not{E}_{T,rel}$ and $e^\pm\mu^\mp\not{E}_{T,rel}$. For all channels, a common preselection is applied, described in Section 7.1. The selection criteria for the final signal region described in Section 7.2 are very similar for all channels, differing only in cut values and the application of the Z mass veto only in the ee and $\mu\mu$ channels. The cut values are optimised for maximum $S/\sqrt{S+B}$, given the lowest unprescaled single-lepton triggers available. In Section 7.3 the signal acceptance and its data-driven corrections are described, and in Section 7.4 results from data and MC simulation are shown and compared.

7.1 Preselection of Events

To select events suitable for further analysis, the following *preselection cuts* are applied to the data sets described in Chapter 5. Most have already been presented in their context in previous chapters and are summarised here with references to their detailed description.

- *Stream Overlap Removal*: Since events from both the "egamma" and "muons" trigger streams are used, events from the "egamma" stream are removed, if they are also in the "muons" stream (see also Section 5.2.2).
- *Data Quality*: Events from runs or luminosity blocks that are not in the "good runs list" are rejected (see Section 5.2).
- *Heavy Flavor Overlap Removal*: Simulated events that are labelled as redundant by the heavy flavor overlap removal tool (see Section 5.1) are removed. This affects $\approx 5.7\%$ of the W+jets MC sample, and less than 1% of the Drell-Yan sample.
- *Event Cleaning*: The event cleaning requirements as described in Section 5.2.1 are applied.

- *Primary Vertex*: The primary vertex is required to be associated with at least three tracks (see Section 6.6).

- *Trigger*: One of the relevant triggers for the channel must have fired (see Section 5.2.2). For the $e\mu$ channel, either the single-electron or the single-muon trigger is accepted.

- *Two Leptons*: Exactly two isolated, oppositely charged leptons with a transverse momentum $p_T > 15$ GeV are required, as described in Chapter 6. In addition, the leading lepton (the leading electron for the $e\mu$ channel) must have a transverse momentum $p_T > 25$ GeV, to ensure that the plateau of the trigger efficiency is reached and scale factors provided by the performance groups can be used.

- *Trigger Match*: At least one of the selected leptons must have fired the trigger (see Section 5.2.2).

In Table 7.1 the observed and predicted events at each step of this preselection are shown. In addition, the scale factors presented in Chapters 5 and 6 are applied to the MC prediction.

In all channels, the observed number of events exceeds the prediction by $\approx 3 - 4\%$. This difference is for example compatible with the systematic uncertainty on the luminosity, but could also arise from correlated systematic errors on electron and muon efficiencies. In addition, the current total theoretical uncertainty on the cross-section of the dominant Drell-Yan $Z/\gamma^* \to \ell\ell$ process at the LHC is also on the order of 4.1% [84]. In this thesis, the Drell-Yan contribution to the signal region will be estimated from a control region, so the uncertainties on the cross-section are expected to cancel.

At this point, it is important to check for differences between observations and predictions, not only in the event count but also in the differential distributions. One of the key distributions is the distribution of the invariant mass $m_{\ell\ell}$ of the selected leptons, shown in Figure 7.1. In the ee and $\mu\mu$ channels, di-lepton events from Z boson decays dominate with 99.4% of events according to MC simulation. In the $e\mu$ channel, 54% of events are top quark decays, 30% arise from Z boson decays and already 5% from W^+W^- decays. No major differences are visible, except an excess around $m_{\ell\ell} \approx 10$ GeV in the ee and $\mu\mu$ channels that can be identified as the Υ resonance, which is not contained in the MC simulations. A closer look also indicates that the peak from Z boson decays is slightly narrower in simulation than in data.

Cut/SF	WW Signal MC	Sum of MC	Observed
initial events	23136 ± 23	238544000 ± 41000	825746369
stream overlap	23136 ± 23	238544000 ± 41000	820290478
data quality	23136 ± 23	238544000 ± 41000	743749597
heavy flavor	23136 ± 23	229592000 ± 41000	743749597
b-jet SF	23264 ± 23	229266000 ± 41000	743749597
event cleaning	23114 ± 23	228323000 ± 41000	735911263
pileup weight	23095 ± 23	228207000 ± 41000	735911263
muon SF	22915 ± 23	227657000 ± 41000	735911263
electron SF	22924 ± 23	227611000 ± 41000	735911263
primary vertex	22810 ± 23	226471000 ± 41000	732763097
electron trigger	8448 ± 14	27265000 ± 11000	170988644
ee channel selection	735.9 ± 4.2	986580 ± 740	1025524
trigger match	735.5 ± 4.2	986260 ± 740	1025073
trigger SF	734.5 ± 4.2	985240 ± 740	1025073
electron or muon trigger	14882 ± 19	62635000 ± 19000	356029424
$e\mu$ channel selection	1748.5 ± 6.5	15017 ± 58	15618
trigger match	1747.0 ± 6.5	14991 ± 58	15579
trigger SF	1743.7 ± 6.5	14956 ± 58	15579
muon trigger	8442 ± 14	35469000 ± 16000	185186996
$\mu\mu$ channel selection	1279.4 ± 5.5	1728560 ± 960	1773911
trigger match	1263.3 ± 5.5	1719020 ± 960	1763866
trigger SF	1256.8 ± 5.5	1711010 ± 960	1763866

Table 7.1: Cut flow table of the observed and predicted events data for the preselection. Scale factors from Chapters 5 and 6 are also applied. The weight of MC events is adjusted so the distribution of the expected number of collisions matches the number in data. Proper comparison of data and MC before the channel selection is not possible, since many MC samples necessary for a full single-lepton analysis are not included. Only the statistical uncertainty on the MC prediction is shown.

In Figures 7.2 and 7.3 the transverse momenta of the leading and subleading leptons are shown. In these plots, a small difference in shape of the lepton spectrum can be observed as well. However, the magnitude of all of the differences are below 10%, the eventual uncertainty on the Drell-Yan contribution. In Figure 7.4 the relative missing transverse energy $\not{E}_{T,rel}$ discussed in Section 6.5 is shown. The shapes are in reasonably good agreement, but - as expected - there are some differences at high values.

7.2 Selection of the Signal Region

To reduce the large fraction of Z+jets events in the same-flavor channels, the invariant mass of the di-lepton system $m_{\ell\ell}$ is required to be outside a 15 GeV window around the Z mass (Z *veto*): $|m_{\ell\ell} - m_Z| > 15$ GeV, $m_Z = 91.1876$ GeV [2]. In addition, $m_{\ell\ell}$ is also required to be larger than 15 GeV to exclude the Υ resonance seen in Figure 7.1, which is broadened by limited detector resolution. For the $e\mu$ channel, the only requirement is that $m_{\ell\ell} > 10$ GeV, to exclude the low invariant mass region that is not very well described by MC simulation and also susceptible to mismeasurement, due to the small opening angles involved.

Further reduction of the fraction of events from Z+jets decays is achieved by requiring a minimum $\not{E}_{T,rel}$ of $55, 50$ and 20 GeV for the $\mu\mu$, ee and $e\mu$ channels, respectively. The $\not{E}_{T,rel}$ distributions for the three channels after the Z mass cut are shown in Figure 7.5. The slightly broader $\not{E}_{T,rel}$ distribution for $\mu\mu$ is the reason for the difference in the ee and $\mu\mu$ channel cuts. The $e\mu$ channel is much less affected by the Drell-Yan background events with low $\not{E}_{T,rel}$, reflected in the low cut value.

The large number of background events from $t\bar{t}$ decays at this stage are more problematic. Since two real W bosons are produced in the decay, both lepton and neutrino ($\not{E}_T$) properties match the signal well. The best possibility to suppress the decays is, therefore, identifying and vetoing additional jets originating from the decays of the two produced b-quarks.

For this reason, all events where a selected jet ($E_T > 25$ GeV, $|\eta| < 4.5$) or a selected b-jet[1] ($E_T > 20$ GeV, $|\eta| < 2.5$) is present are rejected. In Figure 7.6 the distribution of the number of jets after the previous cuts, and in Figure 7.7 the number of b-jets after the additional jet veto is given.

Finally, the required transverse momentum of the subleading lepton $p_{T,2}$ is raised to 20 GeV, to further reduce background where the second lepton is a "fake" lepton (a misidentified jet) or originates from heavy flavor decays[2].

[1] See Section 6.4 for b-Jet selection.

[2] The positioning of this cut at the end of the cutflow allowed for easier comparison of earlier steps with other ATLAS analysis groups.

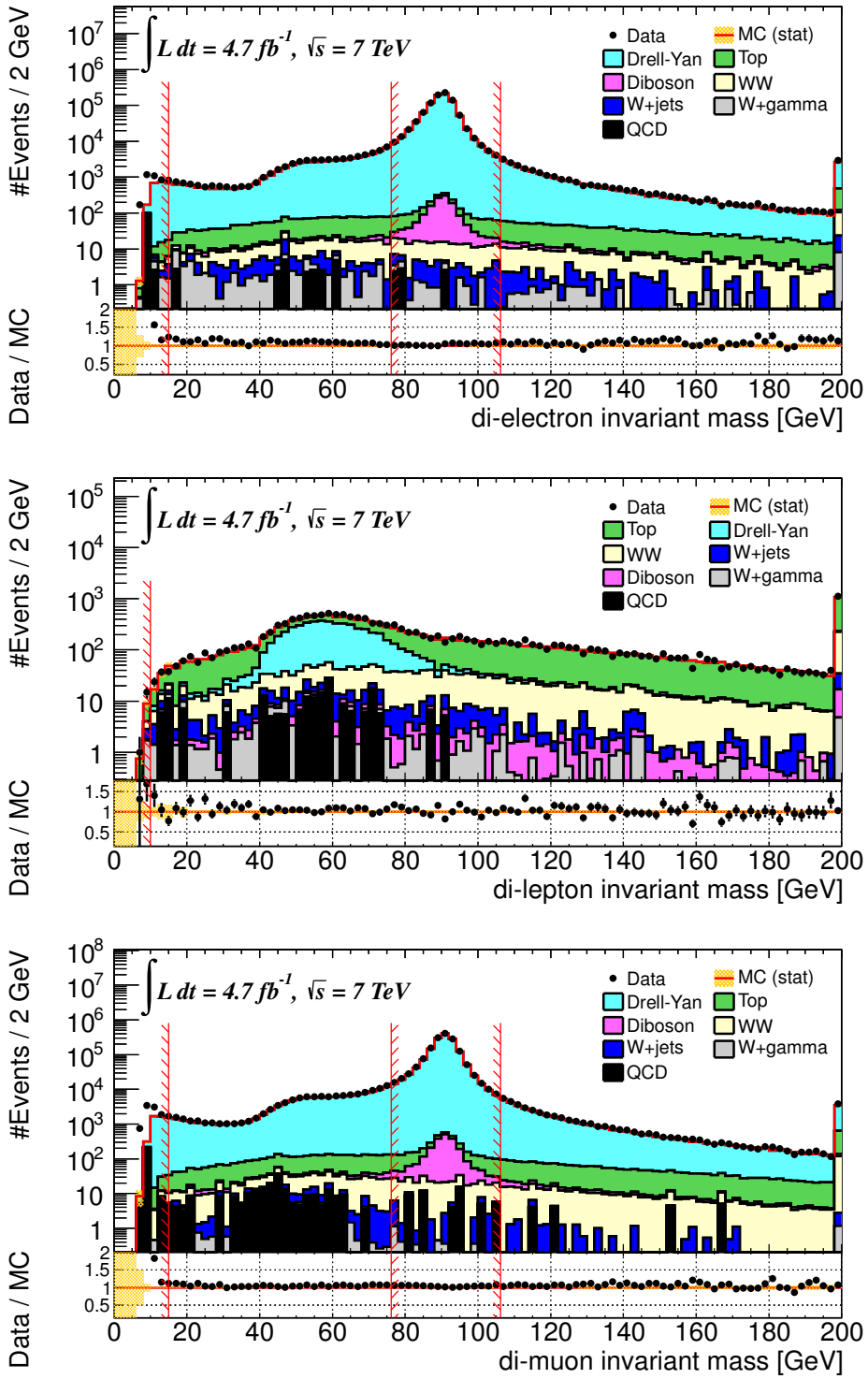


Figure 7.1: Invariant di-lepton mass $m_{\ell\ell}$ in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels (logarithmic scale) after preselection. Cut values are indicated by red vertical lines.

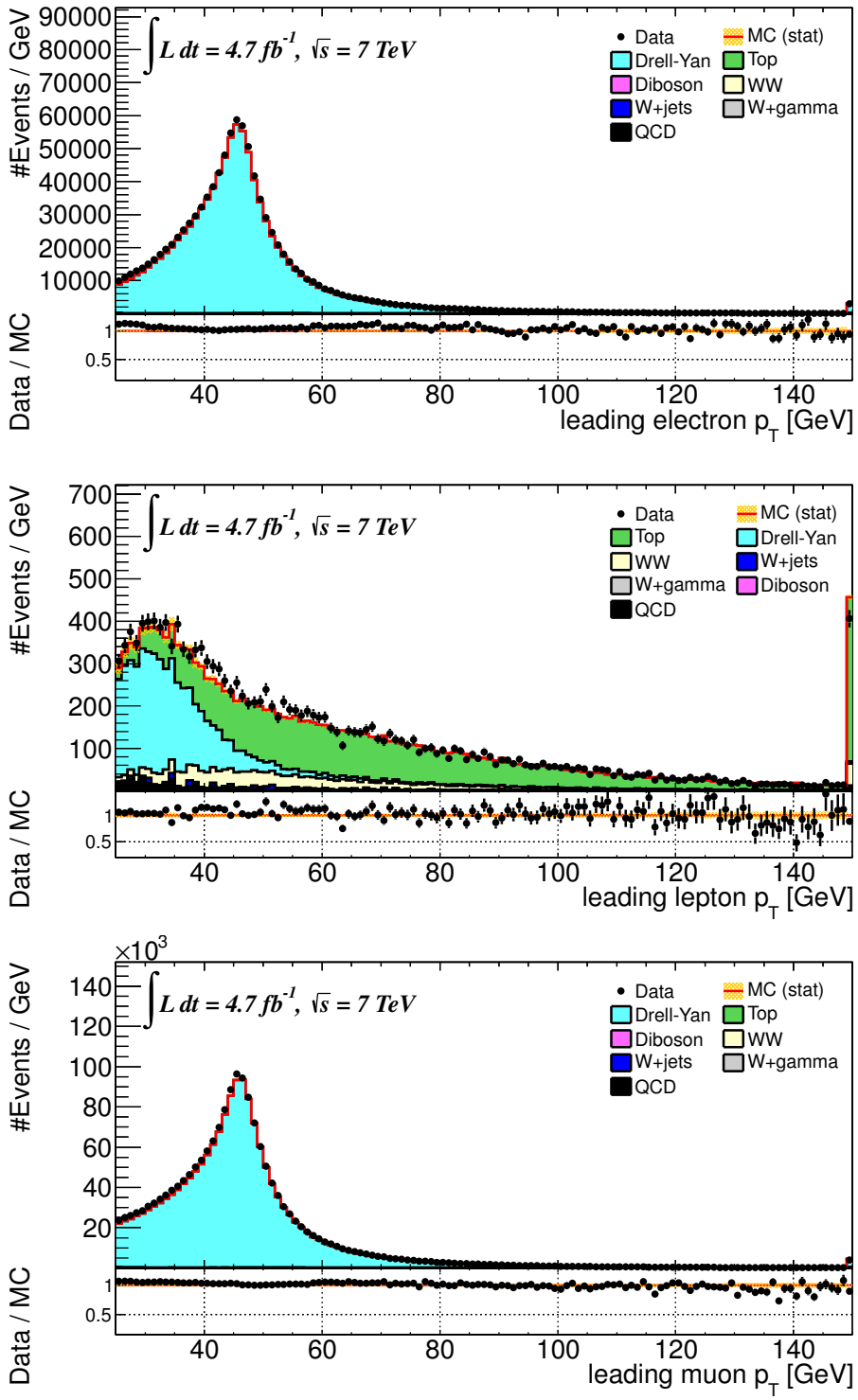


Figure 7.2: Leading lepton transverse momentum p_T in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels after preselection.

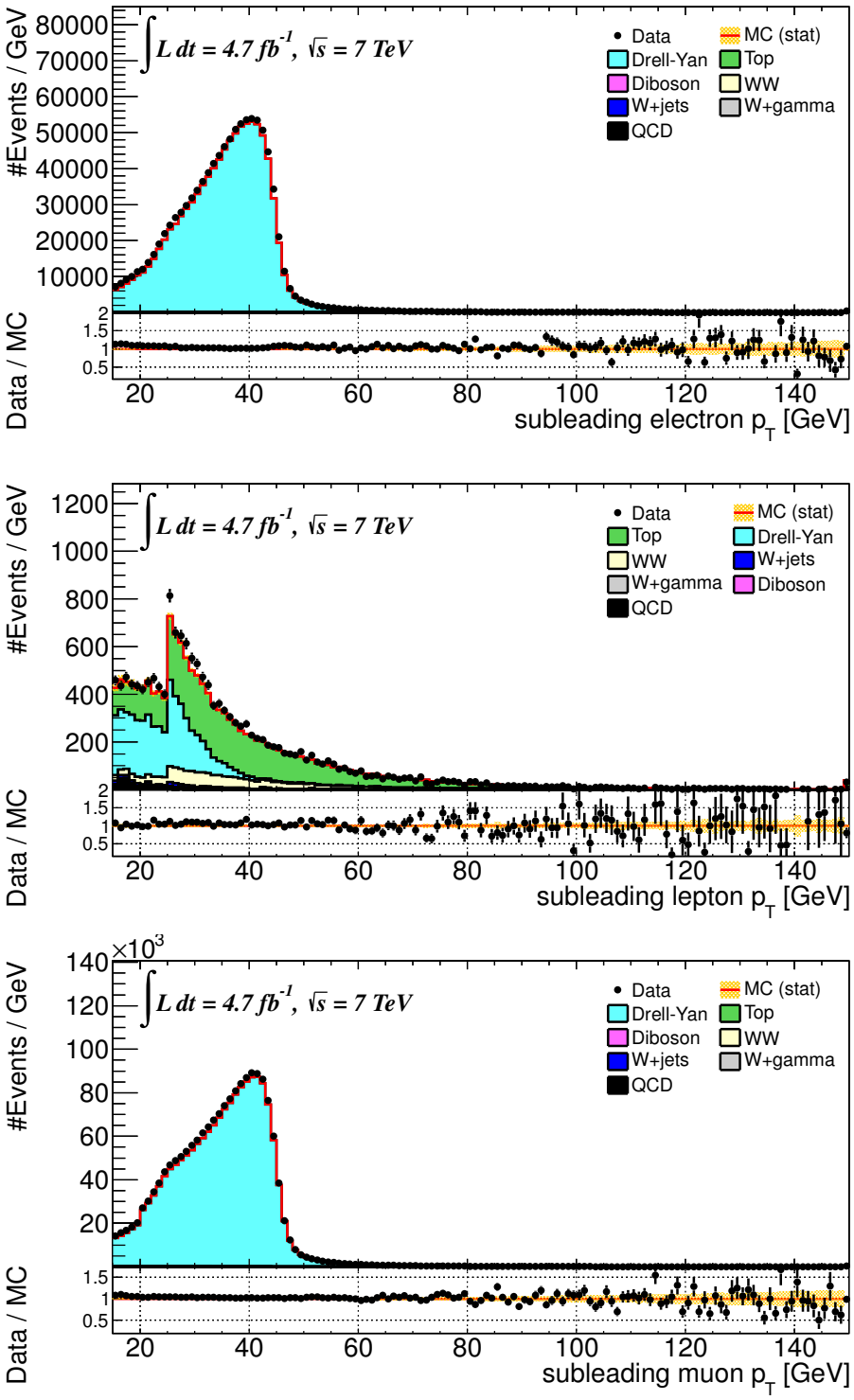


Figure 7.3: Subleading lepton transverse momentum p_T in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels after preselection. In the $e\mu$ distribution the electron peak is clearly visible, since electrons are required to have $p_T > 25$ GeV.

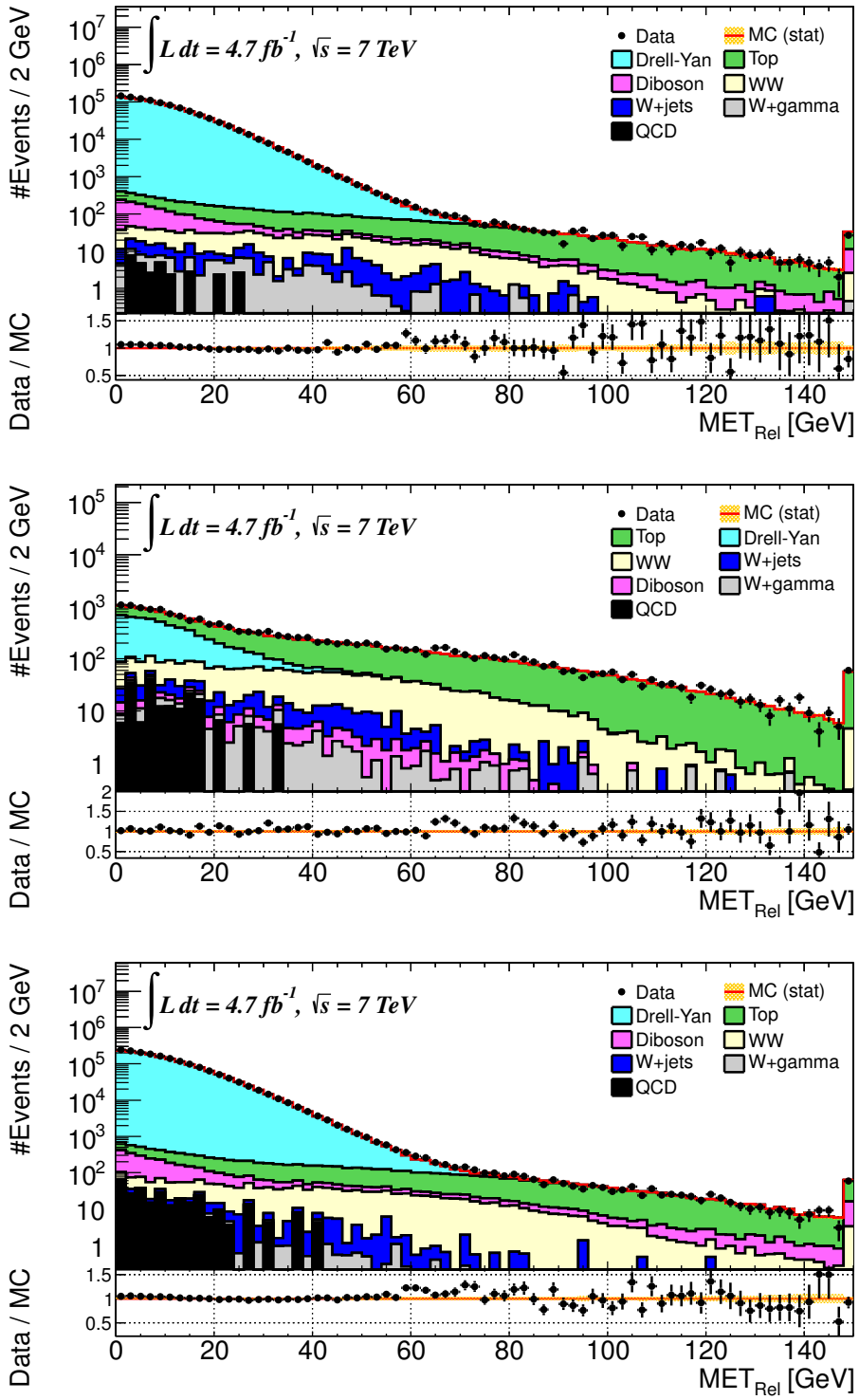


Figure 7.4: Relative missing transverse energy $\not\!\!E_{T,rel}$ in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels (logarithmic scale) after the rejection of low mass resonances.

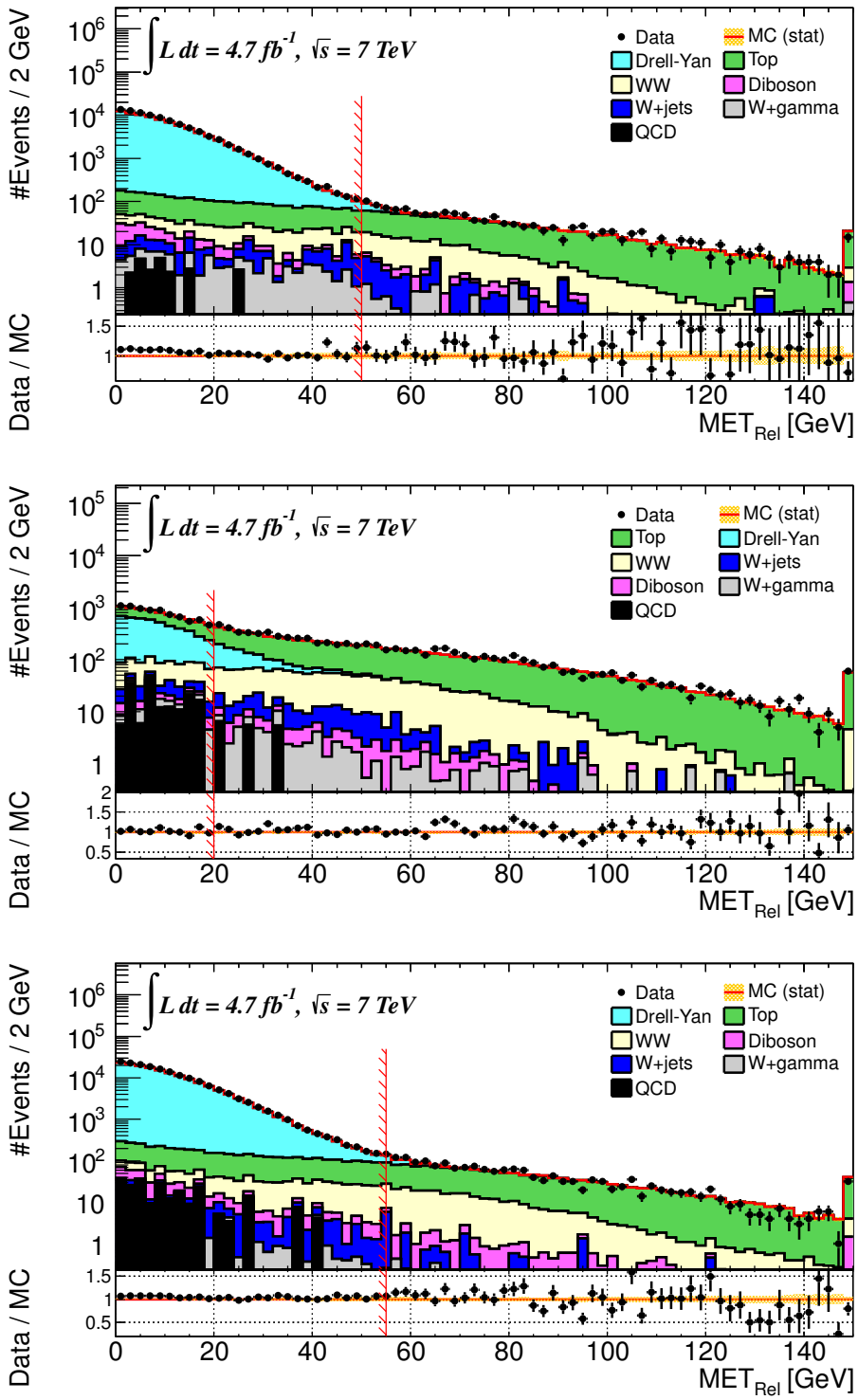


Figure 7.5: Relative missing transverse energy $\not\!\!E_{T,rel}$ in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels (logarithmic scale) after the Z veto.

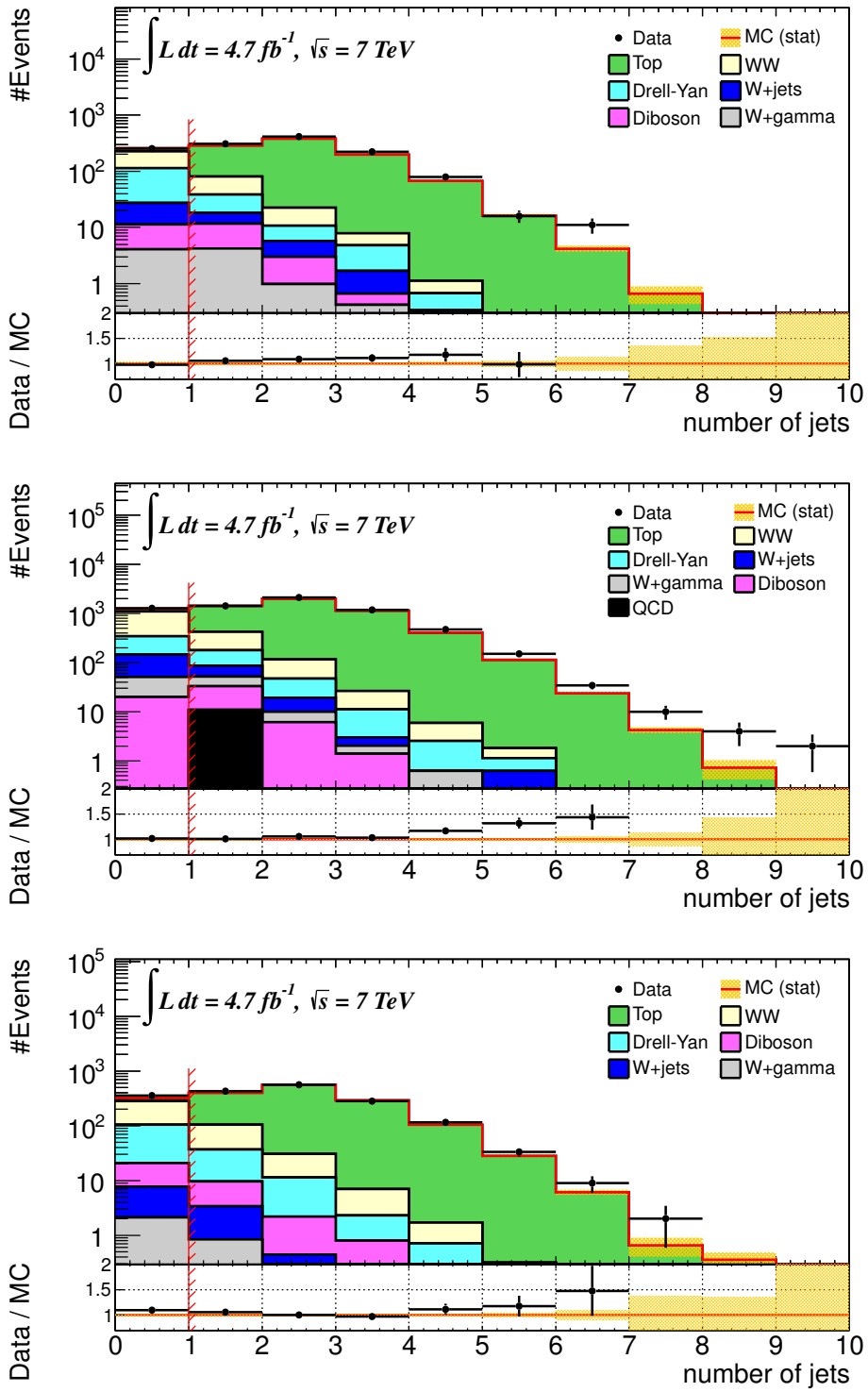


Figure 7.6: Number of jets with $E_T > 25$ GeV in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels after the $\not\!\!E_{T,rel}$ cut.

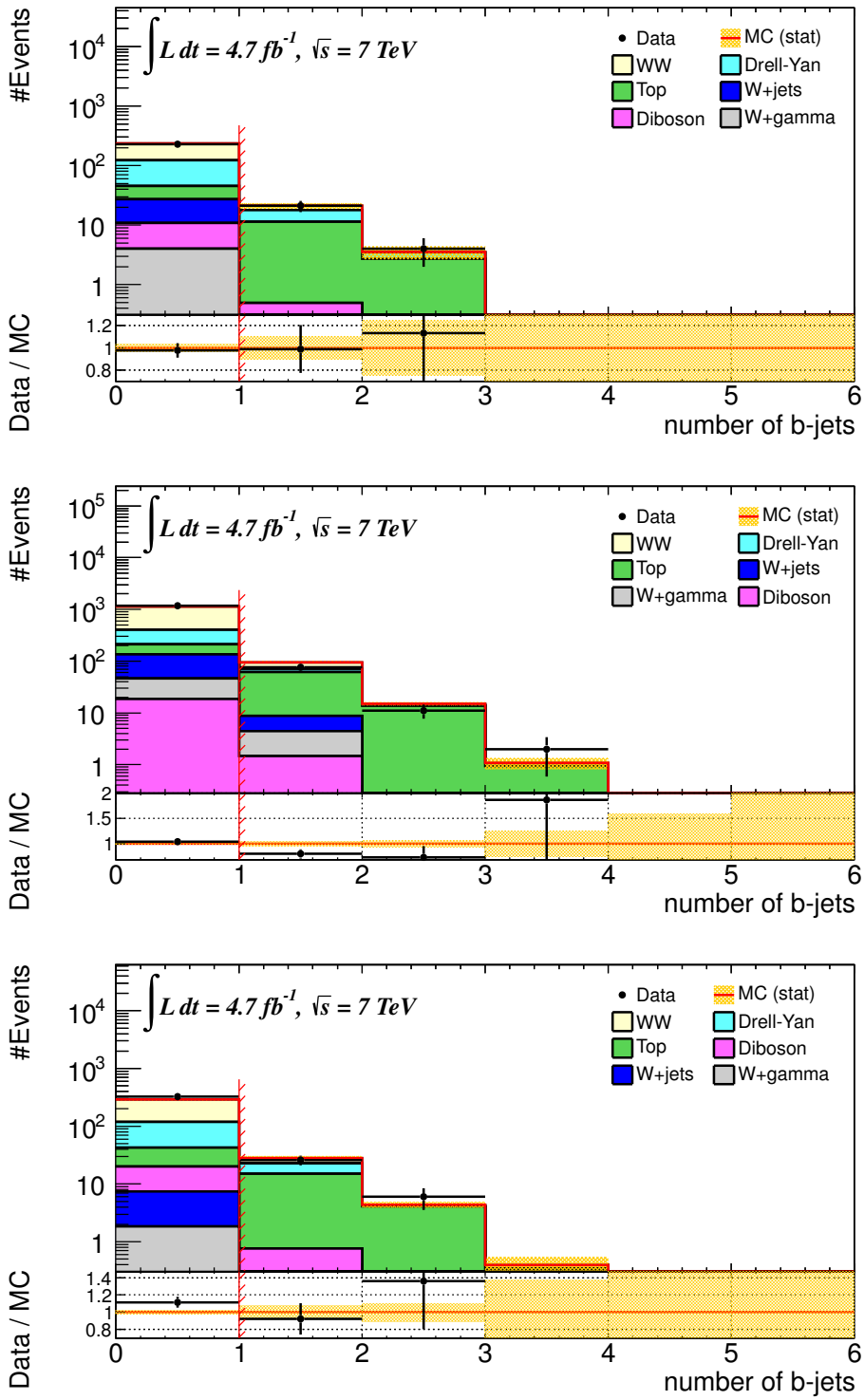


Figure 7.7: Number of b-jets with $E_T > 20$ GeV in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels after the jet veto.

Cut	$ee + \not{E}_{T,rel}$		$e\mu + \not{E}_{T,rel}$		$\mu\mu + \not{E}_{T,rel}$	
	$e\nu e\nu$	$\tau\nu\ell\nu$	$e\nu\mu\nu$	$\tau\nu\ell\nu$	$\mu\nu\mu\nu$	$\tau\nu\ell\nu$
total events	2578.0	12823.6	5156.0	12823.6	2578.0	12823.6
preselection	2538.1	12770.6	5020.2	12770.6	2481.0	12770.6
trigger	2063.0	3435.9	4119.3	6667.0	2031.9	3485.9
2 leptons	659.3	96.2	1773.4	247.0	1167.4	163.6
opposite sign	653.9	95.6	1765.2	245.7	1167.4	163.6
$p_{T,\ell_1} > 25$ GeV	643.7	92.0	1560.8	187.6	1127.9	151.5
trig. match & SF	642.6	91.7	1556.7	186.9	1109.9	146.9
$m_{\ell\ell} > X$ GeV	638.6	91.3	1555.5	186.7	1100.4	145.9
Z veto	499.7	69.4	1555.5	186.7	859.8	112.6
$\not{E}_{T,rel} > X$ GeV	156.7	14.6	980.8	105.0	248.5	22.1
jet veto	100.8	7.9	650.5	66.8	156.3	13.0
b-jet veto	97.9	7.7	629.2	64.4	151.3	12.4
$p_{T,\ell_2} > 20$ GeV	86.9	5.4	586.5	53.1	133.6	9.2
Acceptance	3.37%	0.04%	11.38%	0.41%	5.18%	0.07%

Table 7.2: Signal yields and total acceptance of W^+W^- decays split into the different analysis channels listed at different stages in the analysis. Here, $\ell = e, \mu, \tau$. The cross-contribution from ee, $e\mu$ and $\mu\mu$ decays into different channels is negligible.

7.3 Signal Acceptance and Corrections

7.3.1 Jet Energy Scale Acceptance Correction

Since the efficiency of the jet veto is expected to be not very well modelled in MC simulation, an additional study [85] has been performed, using inclusive Z decays selected with the same preselection cuts. The jet veto efficiency in these events is measured and compared to similar measurements using both MC@NLO and Alpgen generators, resulting in a scale factor on the jet veto efficiency where parts of the jet energy scale uncertainty cancel. In total, a correction factor of $0.953 \pm 0.001(\text{stat}) \pm 0.048(\text{MC}) \pm 0.007(\text{JER} \oplus \text{JES})$ to the jet veto efficiency is determined and is applied in all following tables and results. In addition, the systematic uncertainty on the total selection efficiency due to jet resolution and scale effects is evaluated to be 3.22% (ee), 1.89% ($e\mu$) and 3.17% ($\mu\mu$).

7.3.2 Signal Acceptance

In Table 7.2 the MC prediction - corrected by the JES scale factor - for signal yields and the total signal acceptance are listed. The yields are normalised to 4701 pb^{-1} using the NLO cross-section. Events involving τ decays are

Cut	$ee + \not{E}_{T,rel}$		$e\mu + \not{E}_{T,rel}$		$\mu\mu + \not{E}_{T,rel}$	
	$e\nu e\nu$	$\tau\nu\ell\nu$	$e\nu\mu\nu$	$\tau\nu\ell\nu$	$\mu\nu\mu\nu$	$\tau\nu\ell\nu$
preselection	98.45	99.59	97.37	99.59	96.24	99.59
trigger	81.28	26.90	82.05	52.21	81.90	27.30
2 leptons	31.96	2.80	43.05	3.70	57.45	4.69
opposite sign	99.18	99.38	99.54	99.47	100.00	100.00
$p_{T,\ell_1} > 25$ GeV	98.44	96.23	88.42	76.35	96.62	92.60
trig. match & SF	99.83	99.67	99.74	99.63	98.40	96.96
$m_{\ell\ell} > X$ GeV	99.38	99.56	99.92	99.89	99.14	99.32
Z veto	78.25	76.01	100.00	100.00	78.14	77.18
$\not{E}_{T,rel} > X$ GeV	31.36	21.04	63.05	56.24	28.90	19.63
jet veto	64.33	54.11	66.32	63.62	62.90	58.82
b-jet veto	97.12	97.47	96.73	96.41	96.80	95.38
$p_{T,\ell_2} > 20$ GeV	88.76	70.13	93.21	82.45	88.30	74.19

Table 7.3: Individual cut acceptances of W^+W^- signal samples in % split into the different analysis channels.

displayed in separate columns. In Table 7.3 the efficiencies of the individual cuts are shown. In all channels the trigger efficiency is $\approx$ 82%, but the probability of observing two well-reconstructed, isolated leptons with $p_T >$ 15 GeV is only 32%, 43% and 57% in the ee, $e\mu$ and $\mu\mu$ channels. The total selection efficiencies are 25% for ee, 30% for $e\mu$ and 43% for $\mu\mu$.

In the analysis step, the strongest cut is the one on $\not{E}_{T,rel}$, followed by the jet veto. In total, the analysis cuts have an efficiency of 13.5% for ee, 37.7% for $e\mu$ and 12.0% for $\mu\mu$.

The final selection efficiencies with all corrections relative to all di-leptonic decays of the W^+W^- signal are $0.3994\% \pm 0.0065\%$ (ee), $2.7645\% \pm 0.0156\%$ ($e\mu$), $0.6177\% \pm 0.0078\%$ ($\mu\mu$) and $3.7816\% \pm 0.0195\%$ (combined). The uncertainties are only due to MC statistics, systematic uncertainties on the efficiencies will be calculated in Chapter 9.

7.4 Cutflow and Results

In Table 7.4 the detailed yields of the MC simulations of background and signal processes at all stages of the selection are presented and compared with the observed events. The uncertainties are only MC statistical uncertainties. In the right-hand column, the percentage difference between observed events and data is shown. The fact that no systematic uncertainties have yet been applied has to be taken into account when comparing data and MC simulations.

In the last rows of the respective channel tables, the MC predictions and uncertainties for top quark, Drell-Yan and W+jet decays are replaced by their data-driven estimates including the total statistical and systematic uncertainties. The estimates are detailed in Section 8.3 for the top quark decays, Section 8.2 for Drell-Yan and Section 8.1 for W+jet decays. The W^+W^- scale factor described in Section 7.3 is applied already at the jet veto step.

The differential event distributions in the signal regions are shown in Figures 7.8 - 7.16. Of particular interest is the transverse momentum p_T of the di-lepton system shown in Figure 7.15, which is a prime candidate for further selection cuts for future analyses. The good separation of events from Drell-Yan and W^+W^- decays is clearly visible, and there seems to be good agreement in shape between simulation and data.

ee **channel**	W+gamma	QCD	Top	Diboson	Drell-Yan	W+jets	WW	Total SM	Observed	D/MC
preselection	108.9 ± 7.2	125 ± 18	3511 ± 13	1320.1 ± 5.9	979260 ± 740	179 ± 15	734.5 ± 4.2	985240 ± 740	1025073	+4.0%
$m_{\ell\ell} > 15$ GeV	98.4 ± 6.8	30.0 ± 8.7	3488 ± 13	1319.7 ± 5.9	977460 ± 740	175 ± 15	730.1 ± 4.2	983300 ± 740	1021353	+3.9%
Z veto	84.9 ± 6.3	22.2 ± 7.4	2739 ± 12	156.3 ± 2.0	84890 ± 210	137 ± 13	569.2 ± 3.7	88600 ± 210	95949	+8.3%
$\not{E}_{T,rel} > 50$ GeV	10.0 ± 2.0	0.0 ± 0.0	878.3 ± 6.7	16.86 ± 0.66	113.7 ± 7.8	26.7 ± 5.2	171.3 ± 2.0	1217 ± 12	1303	+7.1%
jet veto	4.1 ± 1.1	0.0 ± 0.0	32.4 ± 1.4	7.19 ± 0.43	85.0 ± 6.9	16.2 ± 4.2	108.8 ± 1.6	253.7 ± 8.4	254	+0.1%
b-jet veto	4.0 ± 1.1	0.0 ± 0.0	18.6 ± 1.0	6.78 ± 0.42	77.7 ± 6.5	16.2 ± 4.2	105.6 ± 1.6	229.0 ± 8.1	229	+0.0%
$p_{T,\ell_2} > 20$ GeV	2.8 ± 1.1	0.0 ± 0.0	15.93 ± 0.97	5.69 ± 0.38	65.2 ± 6.0	9.8 ± 3.2	92.4 ± 1.5	191.9 ± 7.1	196	+2.1%
(data-driven)	2.8 ± 1.1	0.0 ± 0.0	14.0 ± 3.6	5.69 ± 0.38	70.6 ± 6.5	19.8 ± 10.5	92.4 ± 1.5	205.4 ± 13.1	196	−4.6%

$e\mu$ **channel**	W+gamma	QCD	Top	Diboson	Drell-Yan	W+jets	WW	Total SM	Observed	D/MC
preselection	103.8 ± 6.8	176 ± 33	8119 ± 20	102.8 ± 1.6	4434 ± 38	276 ± 17	1743.7 ± 6.5	14956 ± 58	15579	+4.2%
$m_{\ell\ell} > 10$ GeV	102.0 ± 6.8	176 ± 33	8114 ± 20	102.7 ± 1.6	4434 ± 38	275 ± 17	1742.4 ± 6.5	14946 ± 58	15563	+4.1%
Z veto	102.0 ± 6.8	176 ± 33	8114 ± 20	102.7 ± 1.6	4434 ± 38	275 ± 17	1742.4 ± 6.5	14946 ± 58	15563	+4.1%
$\not{E}_{T,rel} > 20$ GeV	54.1 ± 5.0	10.9 ± 7.7	4673 ± 15	50.2 ± 1.1	331 ± 10	139 ± 12	1085.9 ± 5.1	6344 ± 24	6653	+4.9%
jet veto	30.7 ± 3.6	0.0 ± 0.0	144.4 ± 2.9	20.01 ± 0.72	201.2 ± 8.4	94 ± 11	717.3 ± 4.1	1208 ± 15	1265	+4.7%
b-jet veto	27.7 ± 3.4	0.0 ± 0.0	77.2 ± 2.1	18.41 ± 0.68	191.1 ± 8.2	90 ± 11	693.7 ± 4.0	1098 ± 15	1176	+7.1%
$p_{T,\ell_2} > 20$ GeV	17.6 ± 2.6	0.0 ± 0.0	70.8 ± 2.0	16.66 ± 0.65	142.2 ± 7.1	69.9 ± 9.3	639.6 ± 3.9	957 ± 13	1041	+8.8%
(data-driven)	17.6 ± 2.6	0.0 ± 0.0	70.8 ± 15.3	16.66 ± 0.65	142.2 ± 14.4	68.3 ± 30.2	639.6 ± 3.9	955 ± 37	1041	+9.0%

$\mu\mu$ **channel**	W+gamma	QCD	Top	Diboson	Drell-Yan	W+jets	WW	Total SM	Observed	D/MC
preselection	30.4 ± 1.5	492 ± 47	5543 ± 17	2128.4 ± 7.3	1701450 ± 950	115.0 ± 9.2	1256.8 ± 5.5	1711010 ± 960	1763866	+3.1%
$m_{\ell\ell} > 15$ GeV	27.0 ± 1.4	270 ± 35	5497 ± 17	2127.8 ± 7.3	1697310 ± 950	114.5 ± 9.2	1246.3 ± 5.5	1706590 ± 950	1753907	+2.8%
Z veto	22.1 ± 1.2	223 ± 31	4298 ± 15	234.4 ± 2.4	163060 ± 290	91.6 ± 7.8	972.3 ± 4.8	168900 ± 290	178777	+5.8%
$\not{E}_{T,rel} > 55$ GeV	3.07 ± 0.37	0.0 ± 0.0	1288.4 ± 8.0	21.78 ± 0.74	123.0 ± 7.7	8.6 ± 2.4	270.6 ± 2.5	1715 ± 12	1784	+4.0%
jet veto	2.10 ± 0.31	0.0 ± 0.0	41.6 ± 1.5	13.09 ± 0.57	84.4 ± 6.6	5.6 ± 2.1	169.3 ± 2.0	316.1 ± 7.4	357	+12.9%
b-jet veto	1.83 ± 0.28	0.0 ± 0.0	22.9 ± 1.1	12.56 ± 0.56	76.7 ± 6.3	5.6 ± 2.1	163.8 ± 1.9	283.4 ± 7.0	325	+14.7%
$p_{T,\ell_2} > 20$ GeV	1.60 ± 0.27	0.0 ± 0.0	20.1 ± 1.1	10.87 ± 0.52	63.1 ± 5.7	3.7 ± 1.6	142.9 ± 1.8	242.2 ± 6.3	287	+18.5%
(data-driven)	1.60 ± 0.27	0.0 ± 0.0	25.2 ± 5.9	10.87 ± 0.52	71.8 ± 6.5	5.1 ± 2.2	142.9 ± 1.8	257.5 ± 9.2	287	+11.5%

Table 7.4: Cut flow table of the ee (top), $e\mu$ (middle) and $\mu\mu$ channel (bottom) signal selection. Only statistical errors are given for MC. The last rows contain the results using the data-driven values and uncertainties for top quark, Drell-Yan and W+jets decays.

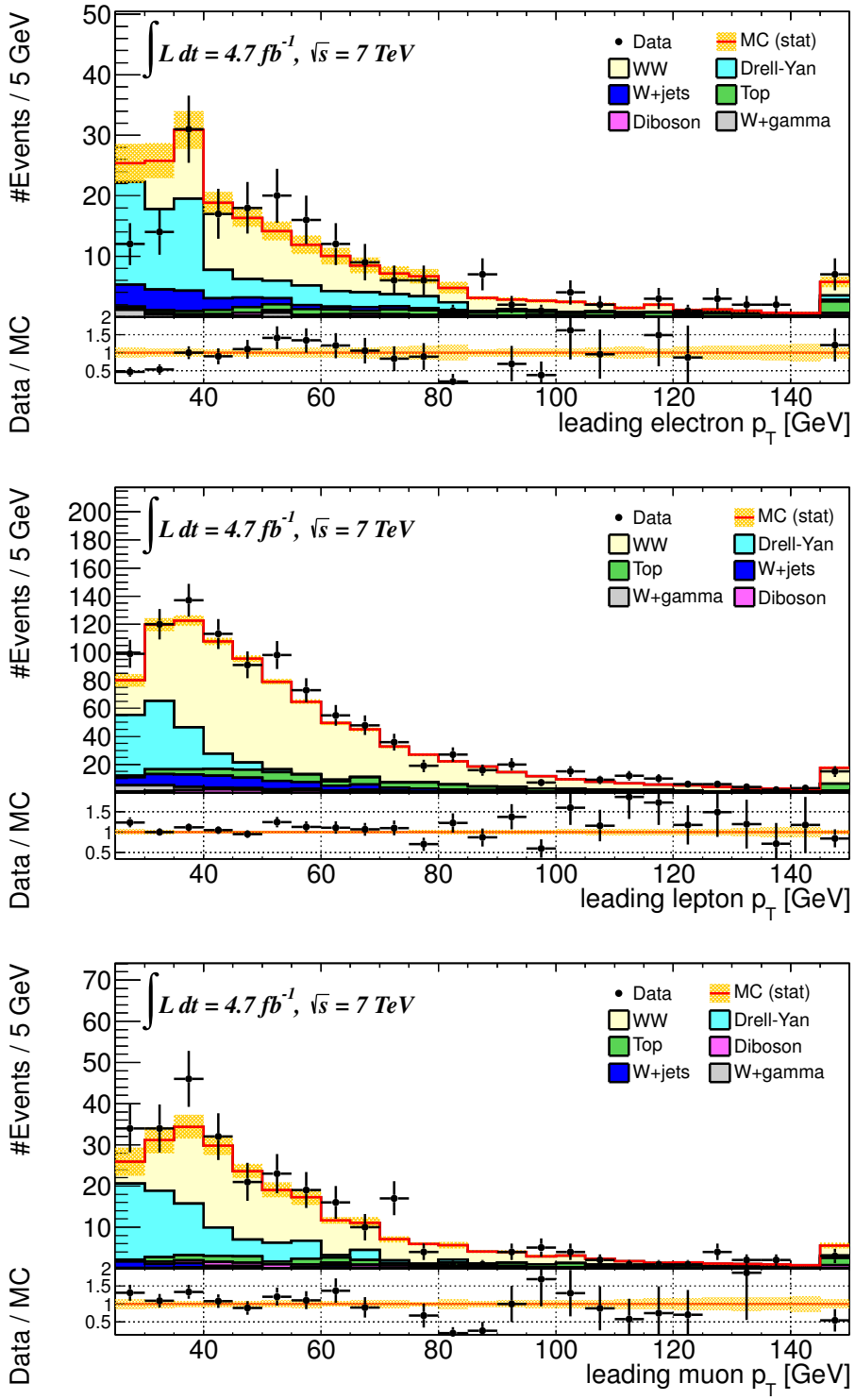


Figure 7.8: Leading lepton p_T in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region.

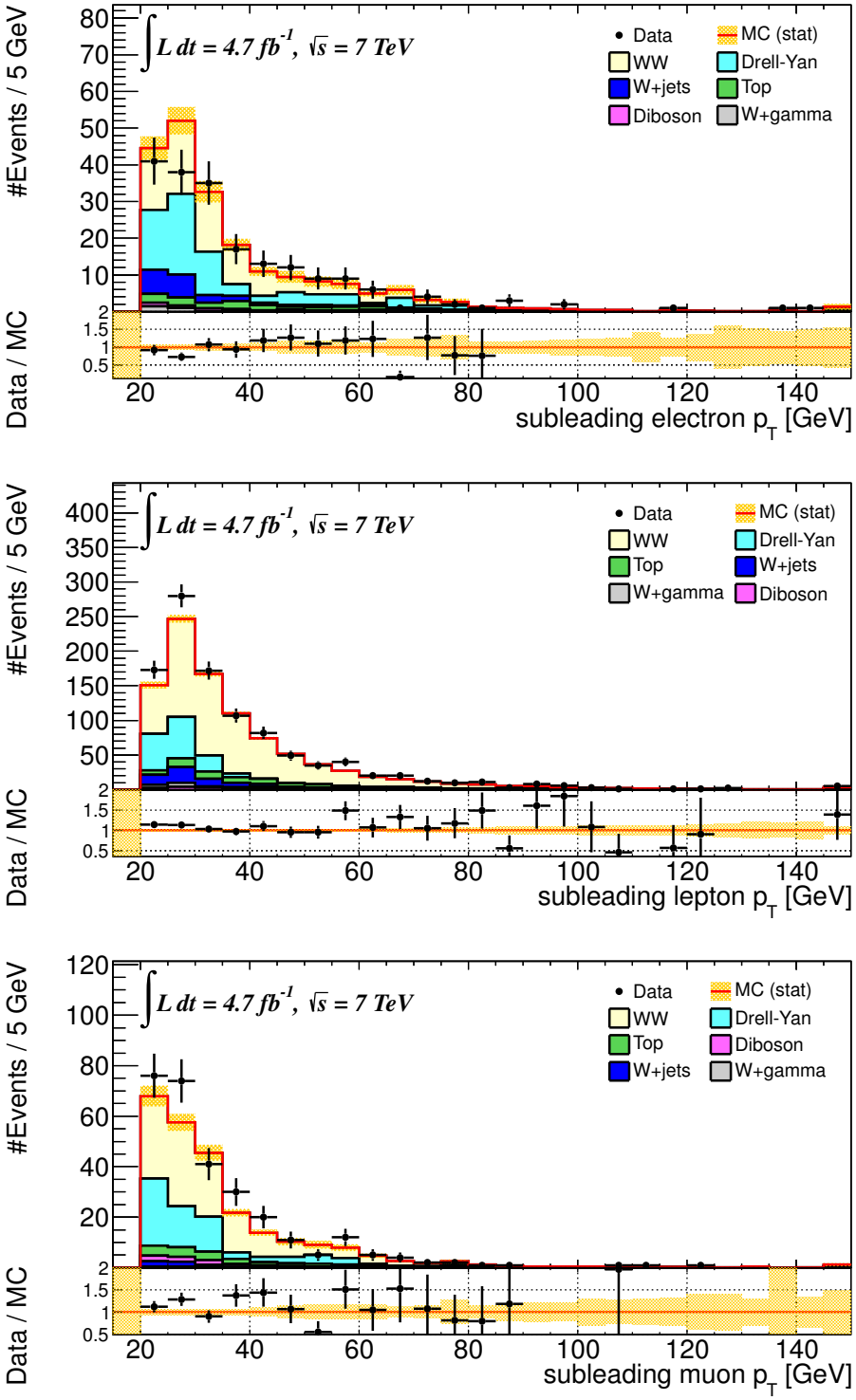


Figure 7.9: Subleading lepton p_T in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region.

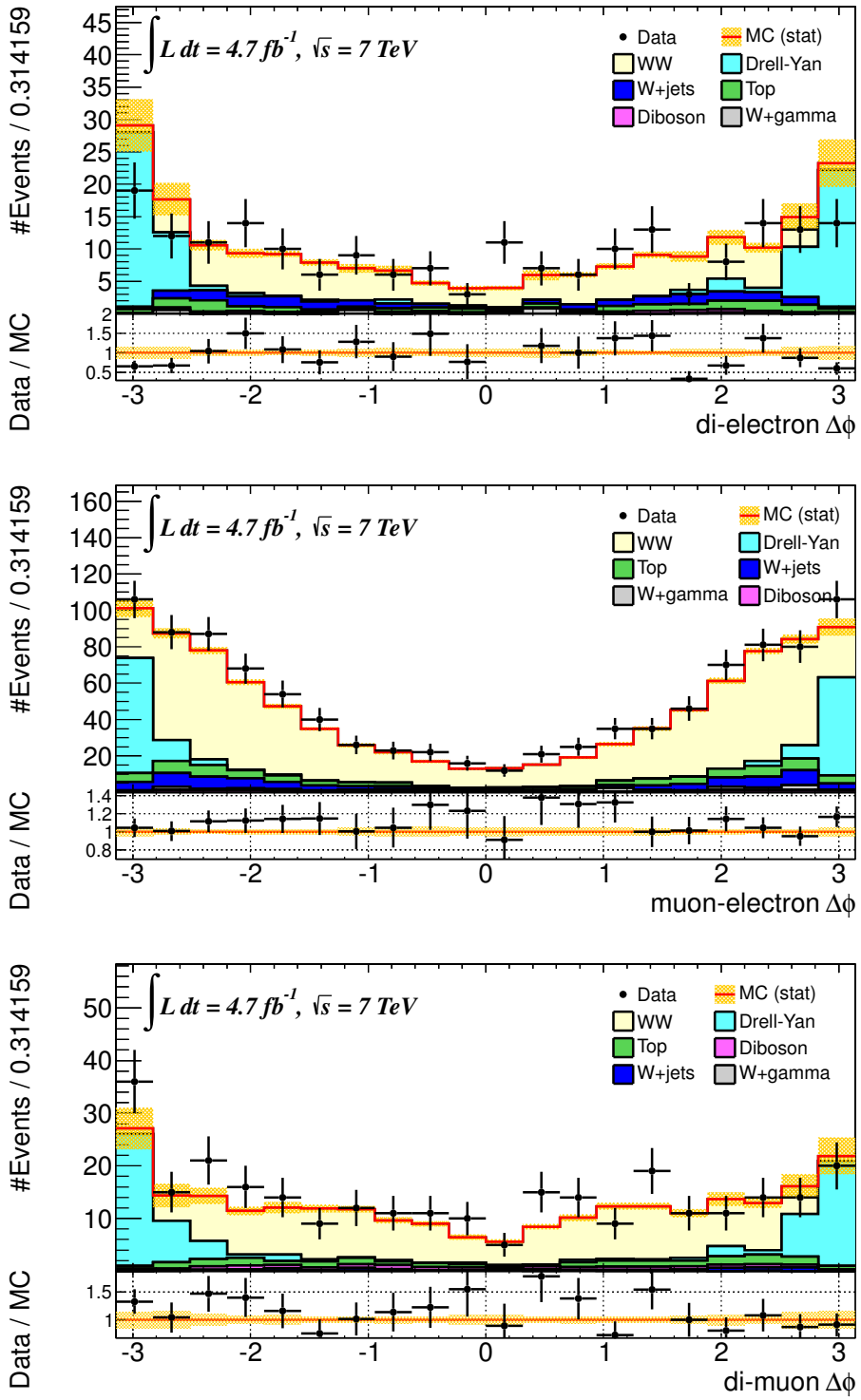


Figure 7.10: Di-lepton azimuthal opening angle in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region. The angle is measured clockwise from the leading lepton.

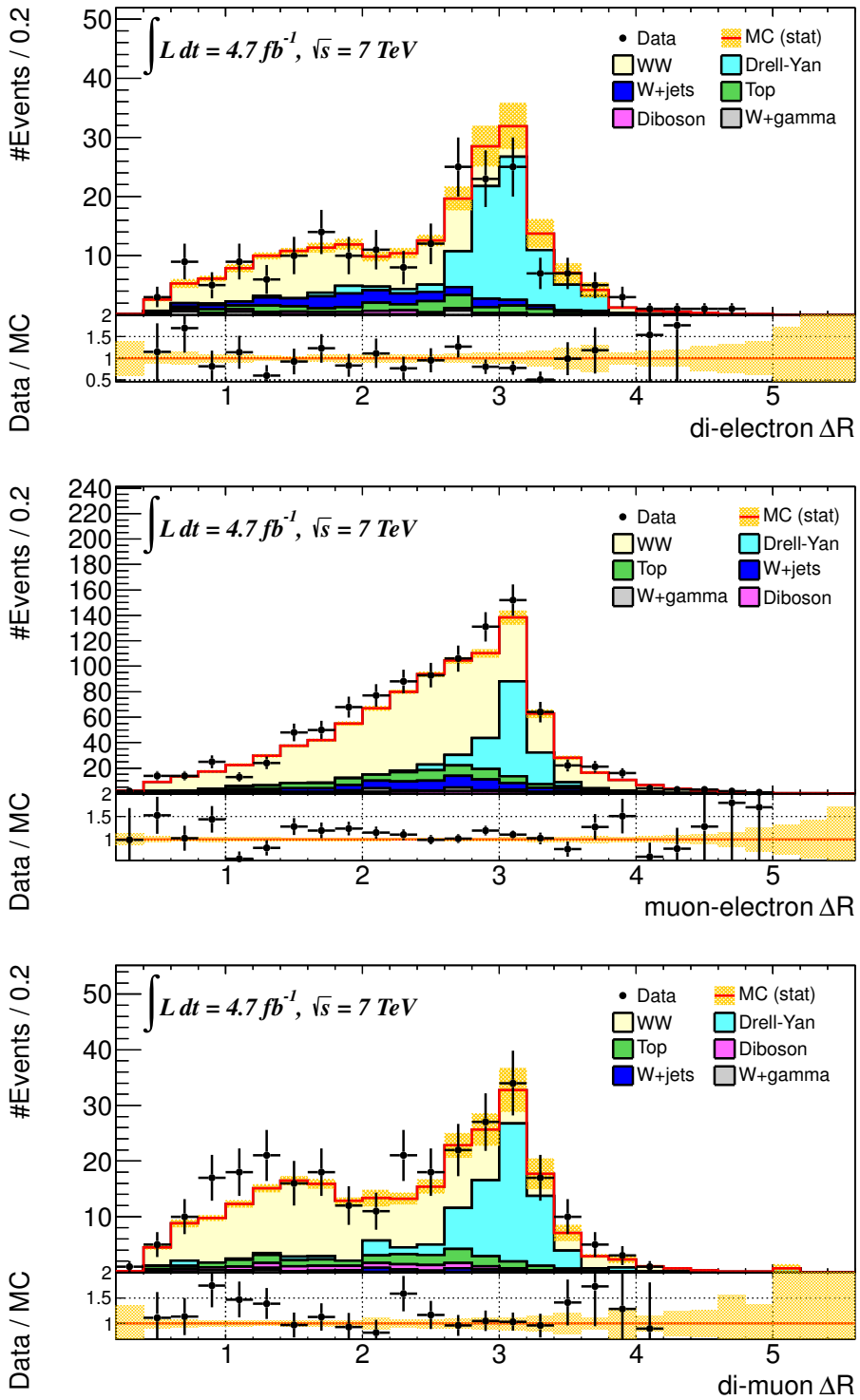


Figure 7.11: Di-lepton distance in the η-ϕ plane in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region.

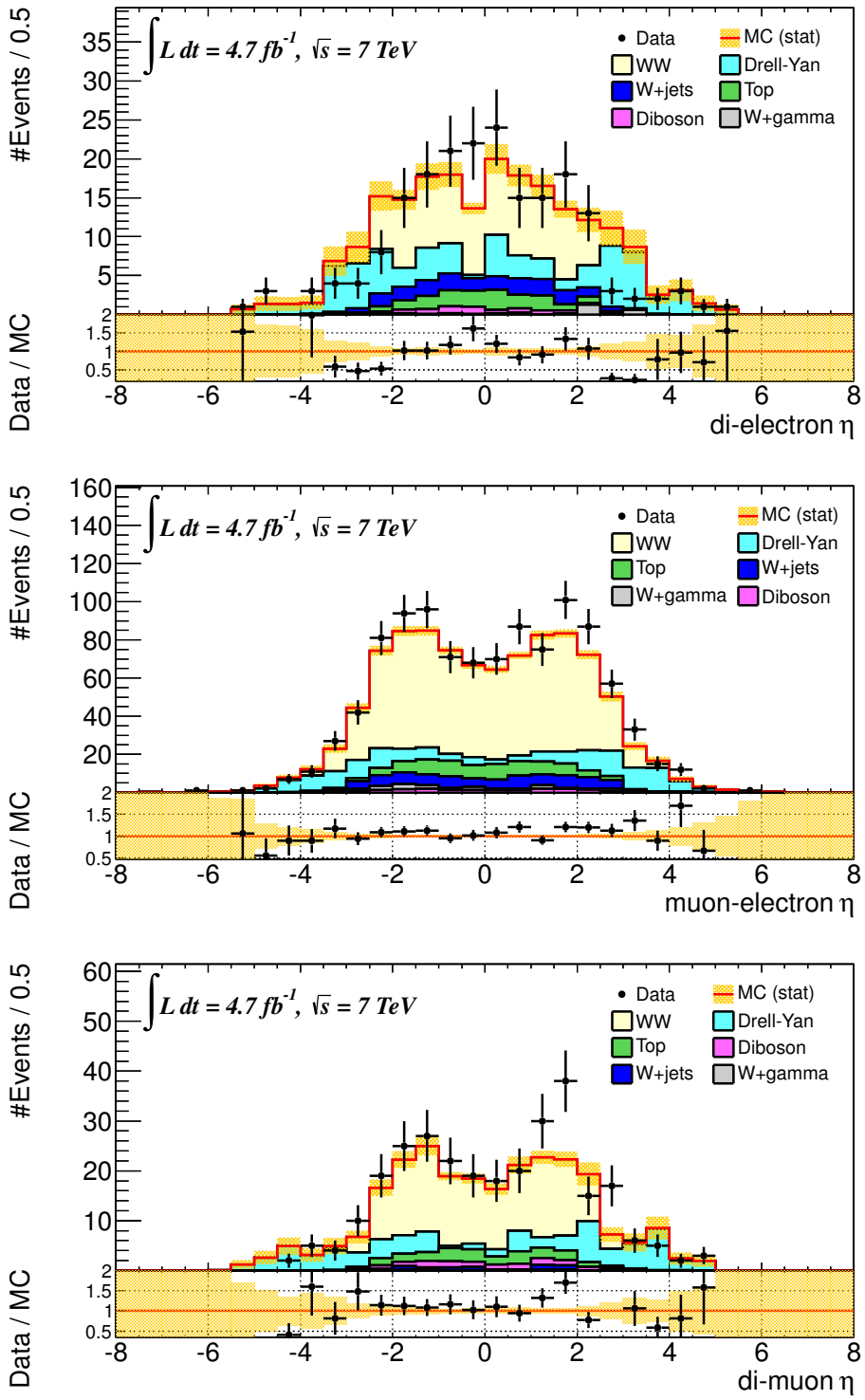


Figure 7.12: Pseudorapidity of the di-lepton system in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region.

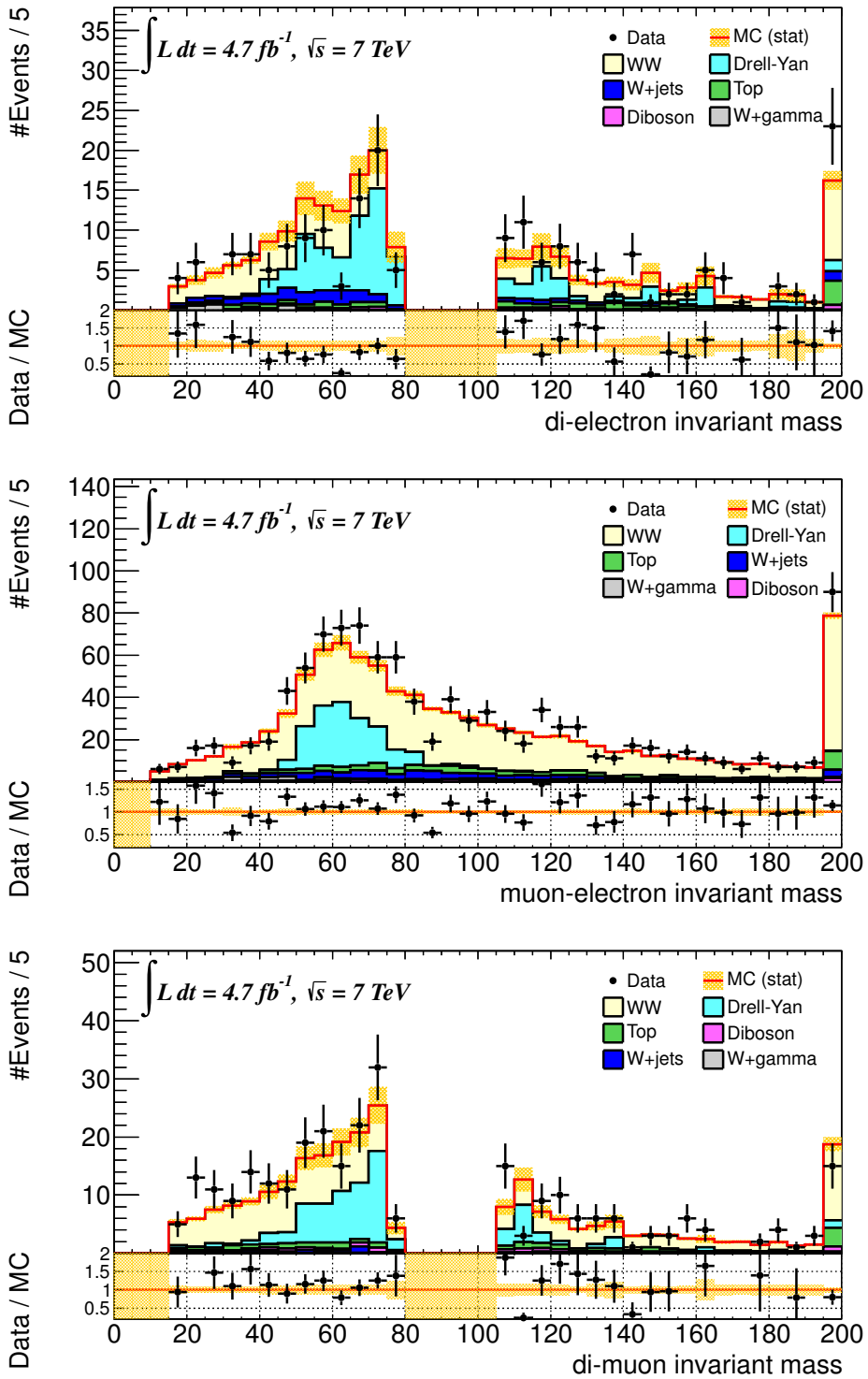


Figure 7.13: Invariant mass of the di-lepton system in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region.

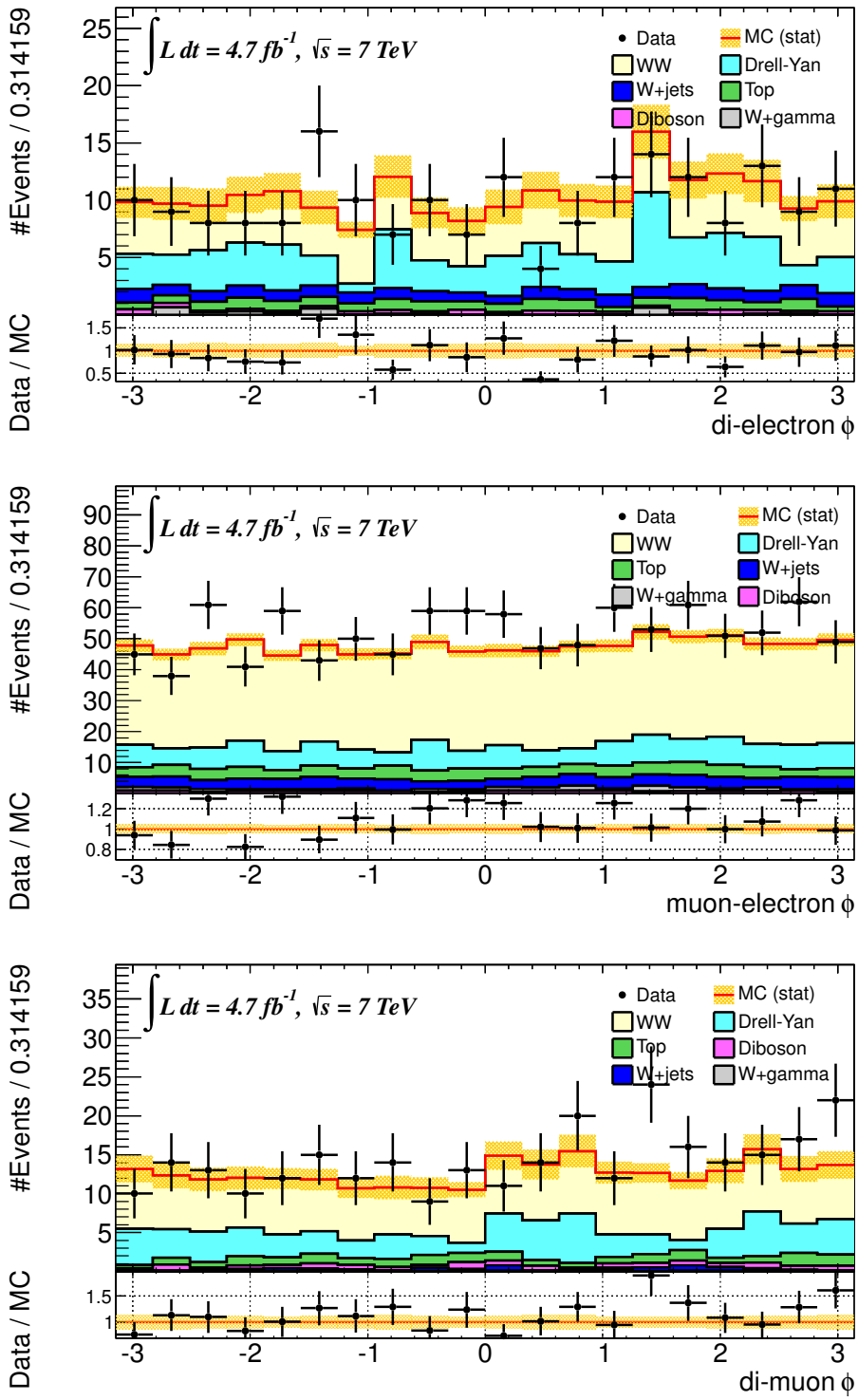


Figure 7.14: Di-lepton momentum azimuthal angle in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region.

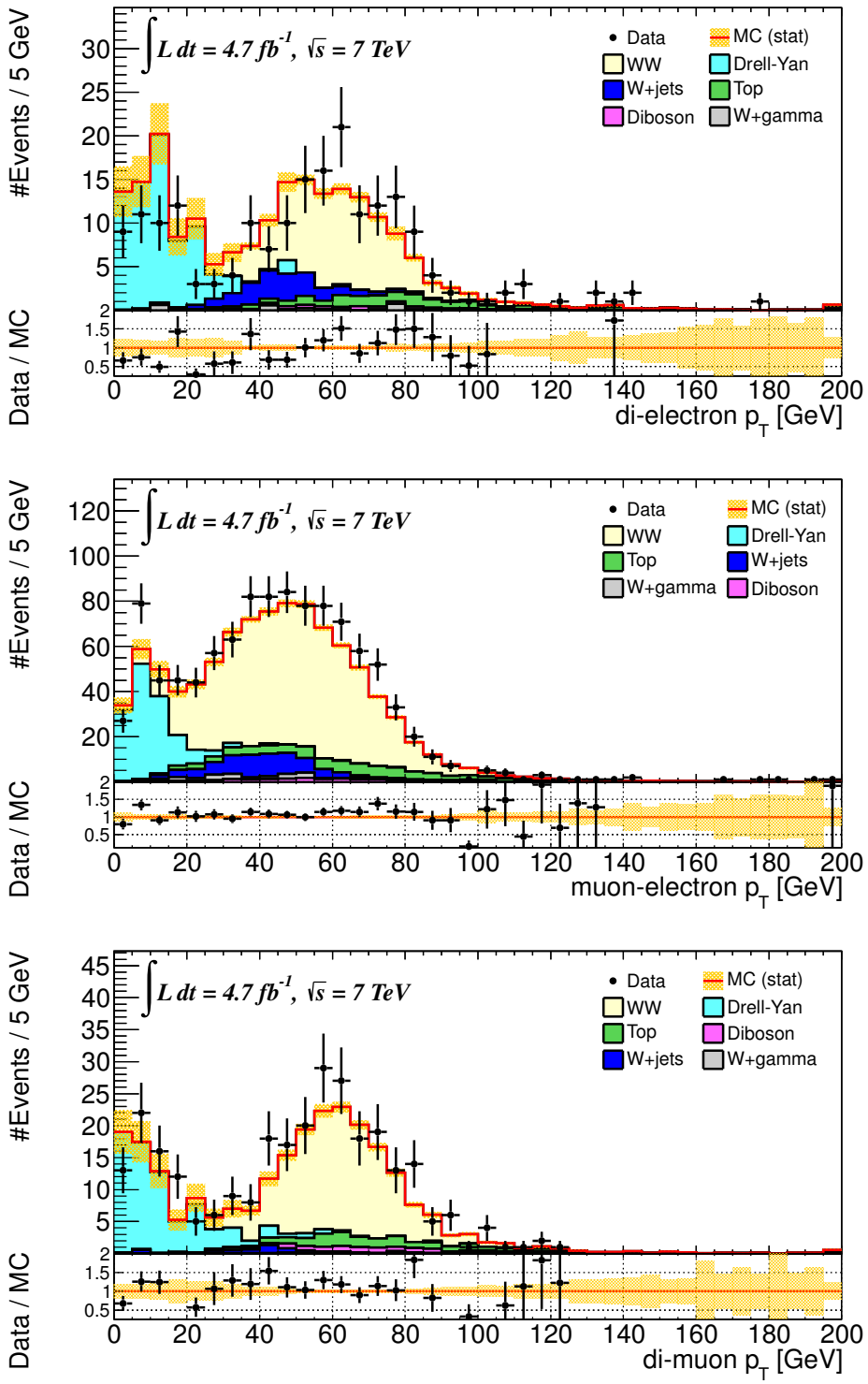


Figure 7.15: Di-lepton momentum p_T in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region.

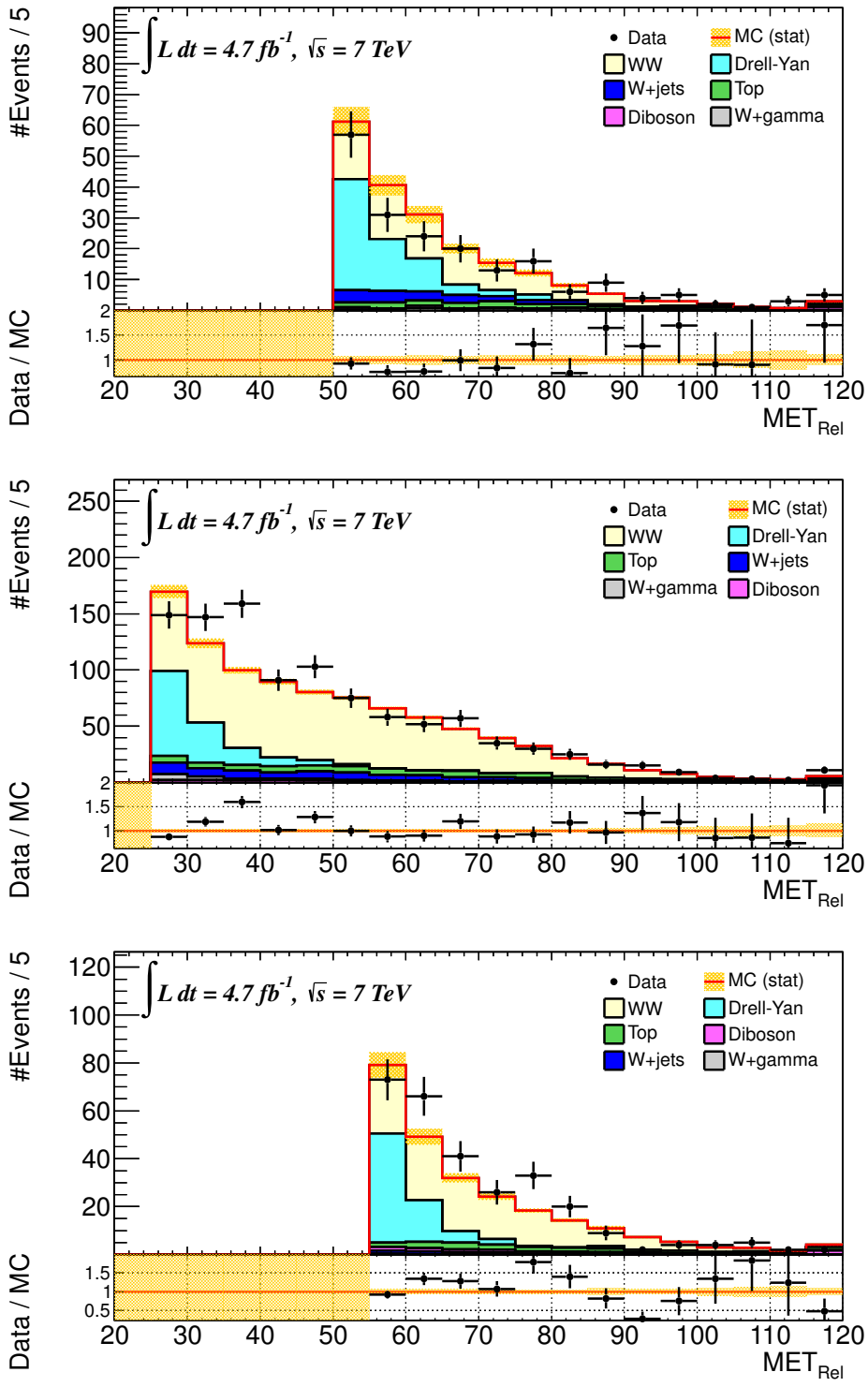


Figure 7.16: Relative missing transverse energy $\not\!\!E_{T,rel}$ in (top to bottom) the ee, $e\mu$ and $\mu\mu$ channels in the signal region.

Chapter 8

Background Estimation

In this chapter, the contributions of the background processes described in Section 2.2.3 to the signal region are estimated. For W+jets decays, the rate of jets misidentified as leptons and the rate of heavy flavor decays passing the isolation cuts are not expected to be well modelled in MC. Therefore, the W+jets contribution is estimated from a fully data-driven sideband selection described in Section 8.1. For the contribution from Drell-Yan (Z/γ^*) decays, both the jet veto efficiency and the tail of the $\not{E}_{T,rel}$ distribution is expected to be mismodelled in MC simulation. To correct for these effects, a data-driven scale factor is derived in Section 8.2. The contribution to the signal region from top quark decays is estimated using a data-driven technique from [77, 85] described in Section 8.3, and is compatible within systematic uncertainties with the prediction from MC simulation. Background contributions from other diboson processes are estimated from MC in Section 8.4, the uncertainties being small due to their small cross-section. Finally, a technique to estimate the number of background events from cosmic ray muons is developed and applied in Section 8.5.

8.1 W+jets

W bosons produced in association with one or more jets can give rise to background events in the W^+W^- signal region if one of the jets is misidentified as a lepton (a *fake lepton*). The rate at which this misidentification occurs is not expected to be accurately described in MC events, and is therefore estimated from data by defining a W+jets control region (CR) that is enriched in fake leptons, yet does not overlap with the signal region.

To define the control region, a looser lepton definition called *jet-rich leptons* is used, which relaxes some of the requirements listed in Section 6 but excludes leptons passing the original cuts. The definitions for jet-rich electrons and muons are listed in Table 8.1.

jet-rich e	Same p_T and η ranges as an identified electron Number of hits in the silicon and pixel detectors ≥ 4 Distance to primary vertex $\lvert z_0 \rvert < 1$mm Calorimeter Isolation in 0.3 cone $< 30\% \times E_T$ Track Isolation in 0.3 cone $< 13\% \times E_T$ Not an identified electron
jet-rich μ	Same p_T and η range as an identified muon Same ID track requirement as identified muon Distance to primary vertex $\lvert z_0 \rvert < 1$mm Calorimeter Isolation $< 30\% \times p_T$ Not an identified muon

Table 8.1: Definitions of jet-rich electrons and muons used for the W+jets control region. Note that no electron quality cut and no d_0 cut is applied for electrons and no track isolation, no d_0 cut and a looser calorimeter isolation is applied for muons.

The W+jets control region is defined identically to the signal region, but instead of requiring two identified leptons, exactly one identified and one jet-rich lepton is required. To extract a prediction for the signal region the *fake factor*, defined as the ratio between identified and jet-rich leptons in a sample of QCD jets, needs to be determined:

$$\mathrm{f}_l \equiv \frac{\mathrm{N}_{\text{identified lepton}}}{\mathrm{N}_{\text{jet-rich lepton}}} \tag{8.1}$$

It is defined and measured separately for electrons and muons in bins of p_T using di-jet samples with very low-threshold electromagnetic triggers, and cross-checked with estimates from Z+jets decays [86]. Systematic errors on the fake factors have also been estimated. The largest systematic uncertainty

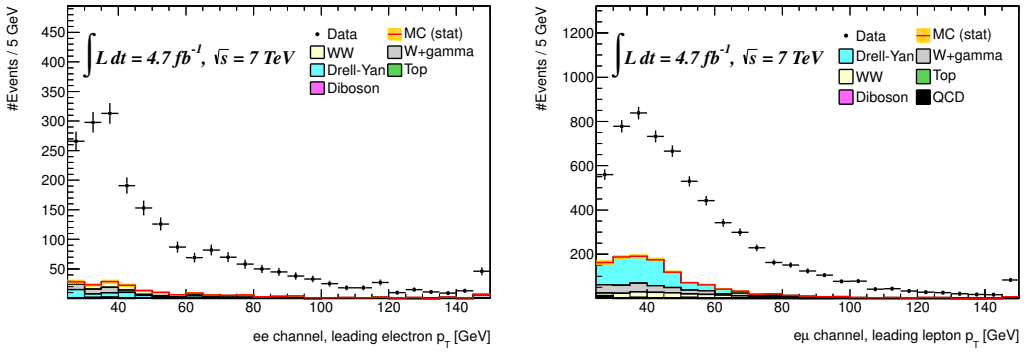


Figure 8.1: Leading lepton p_T for leptons in the signal region with one identified electron or muon and one jet-rich electron. All MC samples except the W+jets simulation are included.

is the *sample dependence*, arising from the difference in kinematics and flavor composition of the di-jet sample used to estimate the fake factor and the W+jets control region. The systematic uncertainty is estimated by applying the method both on di-jet and W+jets MC samples, using the difference of the results as an estimate of the uncertainty.

To obtain the background from W+jets in the ee and $\mu\mu$ channels, the corresponding fake factor is multiplied by the number of events in the control region $\mathrm{N}_{\text{one id + one jet-rich}}$:

$$\mathrm{N}_{\text{one id + one fake}} = \mathrm{f}_l \times \mathrm{N}_{\text{one id + one jet-rich}} \tag{8.2}$$

The $e\mu$ channel receives contributions from both fake electrons and fake muons:

$$\mathrm{N}^{\mathrm{e}\mu\text{-ch}}_{\text{one id + one fake}} = \mathrm{f}_e \times \mathrm{N}_{\text{one id } \mu \text{ + one jet-rich e}} + \mathrm{f}_\mu \times \mathrm{N}_{\text{one id e + one jet-rich } \mu} \tag{8.3}$$

The spectra of the leading (jet-rich) leptons in the W+jets control regions are displayed in Figure 8.1 for electrons and Figure 8.2 for muons. As can be observed in the plots, the contribution of other processes in the control region is small, and can therefore be safely subtracted using the MC predictions while being covered by the large systematic uncertainties on the fake factor.

Cross-Check in the Same-Sign Region

To cross-check and validate the fake-factor method, it is applied to the prediction of background in the signal region where the selection is changed to require two leptons with the same charge. This same-sign region is expected to be enriched with W+jets background, and is predicted to contain almost

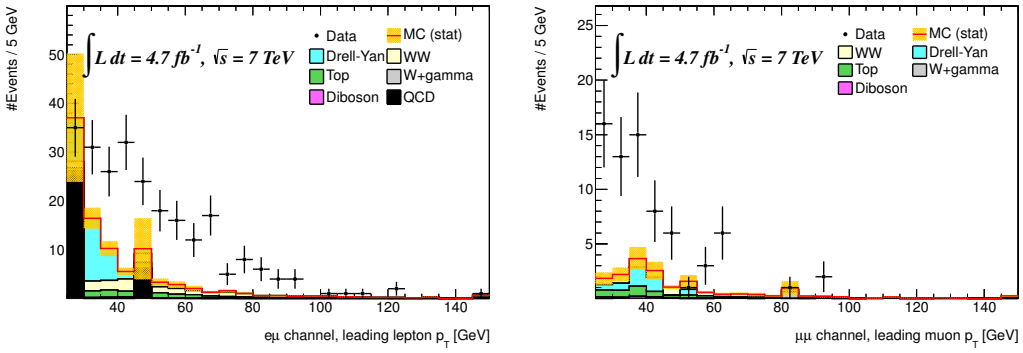


Figure 8.2: Leading lepton p_T for leptons in the signal region with one identified electron or muon and one jet-rich muon. All MC samples except the W+jets simulation are included. The ee channel is shown on the left, the $e\mu$ channel on the right.

	$W+\gamma$	Diboson	Drell-Yan	W+jets	Total	Obs.
ee	107.1 ± 7.5	60.5 ± 1.3	7013 ± 62	260.4 ± 3.5	$7487 \pm 63 \pm 146$	5820
$e\mu$	112.5 ± 7.2	98.0 ± 1.6	96.1 ± 6.4	239.0 ± 6.3	$609 \pm 13 \pm 83$	594
$\mu\mu$	19.3 ± 0.9	54.3 ± 1.2	7.0 ± 1.3	106.6 ± 6.8	$241 \pm 14 \pm 47$	243
$\sum$	239 ± 10	212.9 ± 2.3	7117 ± 63	606.1 ± 10.0	$8338 \pm 66 \pm 233$	6657

Table 8.2: W+jets data driven prediction after the preselection. The total is the sum of all other MC samples and the W+jets prediction. Only the most relevant MC samples are included. The second set of errors on the total are only systematic uncertainties of the W+jets fake factors.

no signal contamination. The predictions of this method compared to observation are given in Table 8.2 for the preselection and in Table 8.3 for the signal region.

Considering all the systematic and statistical uncertainties, the predicted and observed numbers agree reasonably well. The prediction number of events in the electron-electron channel seems to overshoot the observed number, but from Table 8.2 it is obvious that the electron charge misidentification rate in Drell-Yan events in data is lower than expected from simulations. Assuming the charge misidentification also affects other MC simulations, the excess yields in the signal region can be interpreted as overprediction by the MC simulation samples. The leading and subleading lepton spectra in the same sign region are displayed in Figure 8.3, including the data-driven W+jets prediction.

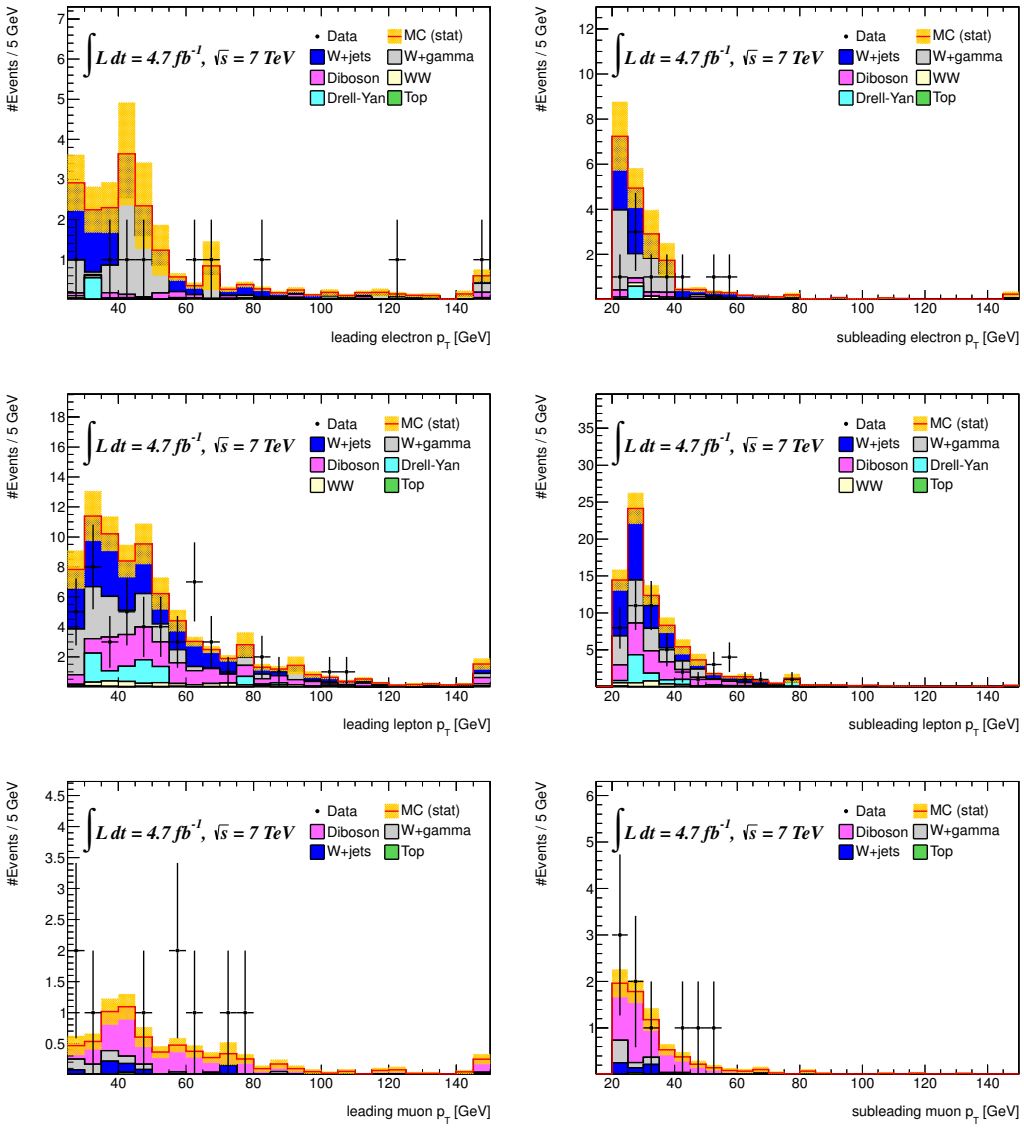


Figure 8.3: Leading and subleading lepton p_T for leptons in the same-sign control region after all cuts compared to predictions. From the top to the bottom row the ee, $e\mu$ and $\mu\mu$ channels are shown.

	$W+\gamma$	Diboson	Drell-Yan	W+jets	Total	Obs.
ee	7.3 ± 2.1	1.4 ± 0.2	0.5 ± 0.5	9.6 ± 0.4	$19.6 \pm 2.2 \pm 5.1$	9
$e\mu$	17.6 ± 2.6	17.4 ± 0.7	7.0 ± 1.8	31.4 ± 1.0	$76.9 \pm 3.4 \pm 14.7$	48
$\mu\mu$	0.88 ± 0.2	5.2 ± 0.4	0.0 ± 0.0	0.5 ± 0.3	$6.6 \pm 0.5 \pm 0.3$	9
$\sum$	25.8 ± 3.4	23.9 ± 0.8	7.5 ± 1.9	41.5 ± 1.1	$103.0 \pm 4.1 \pm 19.8$	66

Table 8.3: W+jets data driven prediction in the same-sign "signal" region. The total is the sum of all other MC samples and the W+jets prediction. Only the most relevant MC samples are included. The second set of errors on the total are only systematic uncertainties of the W+jets fake factors.

	$ee + \not{E}_{T,rel}$	$e\mu + \not{E}_{T,rel}$	$\mu\mu + \not{E}_{T,rel}$	Combined
fake e	$19.77 \pm 0.51 \pm 10.45$	$53.44 \pm 0.91 \pm 28.23$	-	$73.2 \pm 1.0 \pm 38.7$
fake μ	-	$14.9 \pm 2.3 \pm 10.5$	$5.07 \pm 0.86 \pm 2.00$	$19.9 \pm 2.5 \pm 12.5$
total	$19.77 \pm 0.51 \pm 10.45$	$68.3 \pm 2.5 \pm 30.1$	$5.07 \pm 0.86 \pm 2.00$	$93.2 \pm 2.7 \pm 40.7$

Table 8.4: Summary of the W+Jet background estimates with associated statistical and systematic uncertainties. Numbers given are expected events for 4.7 fb^{-1}.

Summary

In Table 8.4 the results of the W+jets background estimation method are shown. Systematic uncertainties are propagated from the systematic uncertainty of the fake factors. Since this method of estimating the W+jets background already includes all selection, reconstruction, scale and identification efficiencies, systematic uncertainties on these are not considered. In [77] a largely independent method is used to cross-check this estimate, and the results are compatible within systematic uncertainties.

8.2 Drell-Yan

Drell-Yan decays ($Z/\gamma^* \to \ell^+\ell^-$) are the most numerous background events to W^+W^- production, even after applying a di-lepton mass cut as described in Section 7.2. At this stage of the analysis, MC simulations for this process are in good agreement with the data, yet the high rate of the background events makes it necessary to put additional cuts into tails of distributions that are very sensitive to uncertainties in the detector simulation: Vetoing on jets with a transverse momentum $p_T > 25$ GeV depends on the jet energy scale and resolution, and requiring $\not{E}_{T,rel} > 55, 50$ GeV depends on all energy scales, energy resolutions and reconstruction efficiencies.

To reduce this uncertainty and to cross-check the MC prediction of the

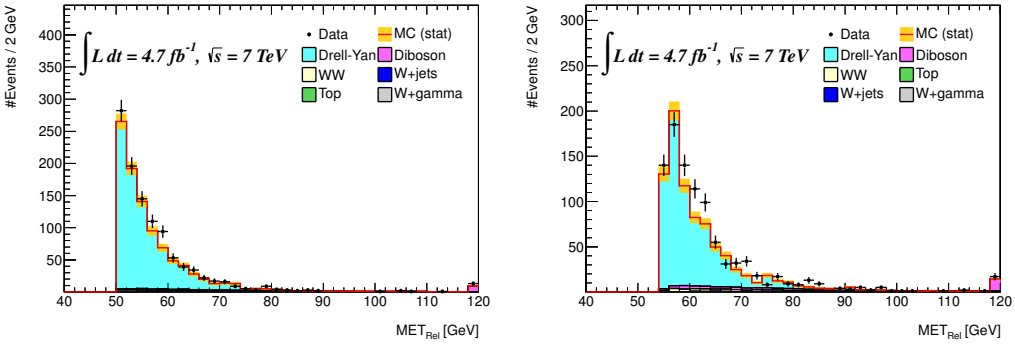


Figure 8.4: Comparison of data and MC prediction in the Drell-Yan control region after all cuts in the ee and $\mu\mu$ channels.

efficiencies in the tails of the distribution, a scale factor with systematic uncertainties for the ee and $\mu\mu$ channels is calculated using a sideband method similar to the one presented in Section 7.3. For the $e\mu$ channel, where $Z/\gamma^* \to \tau^+\tau^-$ decays are a major background, only a systematic uncertainty is derived from the sidebands in the ee and $e\mu$ channel.

The scale factors S for the ee and $\mu\mu$ channels are derived by comparing the agreement of data and MC simulation inside the Z mass peak where $|m_{\ell\ell} - m_Z| < 15$ GeV. It is conservatively assumed that Drell-Yan mismodelling is the sole cause for differences inside the Z mass peak, a reasonable assumption given that MC simulations predict that 91% (ee) and 86% ($\mu\mu$) of the events in the control region come from Drell-Yan decays. All other cuts are applied unchanged. The scale factor is then defined as

$$S = 1 + \frac{N_{data}^{ZCR} - N_{MC}^{ZCR}}{N_{MC,DY}^{ZCR}} \tag{8.4}$$

where N_{data}^{ZCR} is the number of data events in the control region, N_{MC}^{ZCR} the number of events predicted by MC simulation in the control region, and $N_{MC,DY}^{ZCR}$ the number of Drell-Yan events predicted in the control region. Using this definition, S can be used to correct the Drell-Yan MC prediction in the signal regions for differences in efficiencies between data and MC. The distributions for $\not\!\!E_{T,rel}$ in this control region are shown in Figure 8.4.

For the $e\mu$ channel this approach is not possible, since no invariant mass cut is applied. To obtain an estimate of the systematic uncertainty on the selection in this channel, the factor S is calculated in both the ee and $\mu\mu$ channels again, but substituting the $\not\!\!E_{T,rel}$ cut value normally used in the $e\mu$ channel. The two values obtained for S are averaged, assuming that the sources of fake $\not\!\!E_{T,rel}$ are the same in all channels. One indicator for this hypothesis is that the $\not\!\!E_{T,rel}$ distribution in MC simulation is very similar for

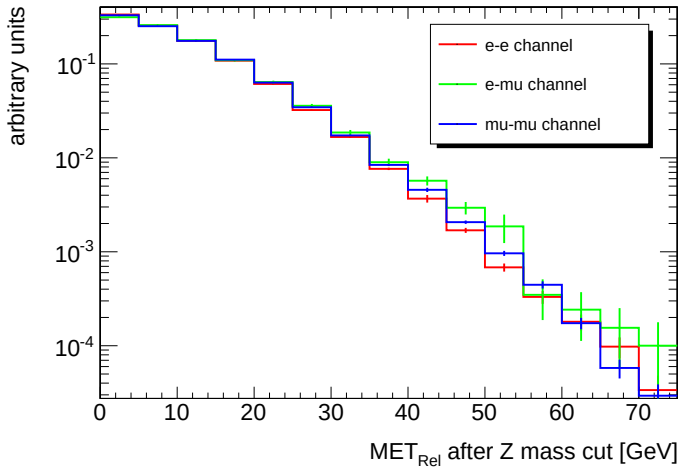


Figure 8.5: Comparison of relative missing transverse energy after the Z mass cut for Drell-Yan MC only in ee, $e\mu$ and $\mu\mu$ channels.

channel	$\not\!\!E_{T,rel}$ cut	N^{ZCR}_{data}	N^{ZCR}_{MC}	$N^{ZCR}_{MC,DY}$	S factor
ee	55 GeV	1066	992 ± 23	903 ± 23	1.082 ± 0.044
$\mu\mu$	50 GeV	953	851 ± 20	736 ± 20	1.139 ± 0.050
ee	25 GeV	779	845 ± 21	834 ± 21	0.921 ± 0.042
$\mu\mu$	25 GeV	1978	1960 ± 33	1940 ± 33	1.009 ± 0.029
$e\mu$	25 GeV	-	-	-	0.965 ± 0.051

Table 8.5: Drell-Yan background control region event yields. All errors are statistical only. The S factor is calculated using Equation 8.4. The result for the $e\mu$ channel is obtained by taking the difference between the S values for ee and $e\mu$ channels obtained with the $\not\!\!E_{T,rel}$ cut of the $e\mu$ channel as a systematic uncertainty.

$ee + \not\!\!E_{T,rel}$	$72.0 \pm 6.7(\text{stat}) \pm 3.2(\text{syst})$
$e\mu + \not\!\!E_{T,rel}$	$142.2 \pm 7.1(\text{stat}) \pm 12.5(\text{syst})$
$\mu\mu + \not\!\!E_{T,rel}$	$70.0 \pm 6.5(\text{stat}) \pm 3.5(\text{syst})$
Combined	$284.2 \pm 11.7(\text{stat}) \pm 17.2(\text{syst})$

Table 8.6: Summary of Drell-Yan background estimate from MC with scale factors and systematic uncertainties estimated from the Z control region. Numbers given are expected events for 4.7 fb^{-1}.

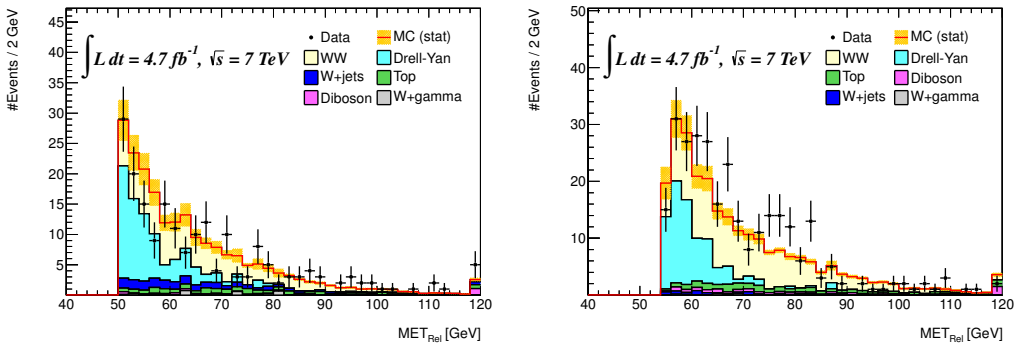


Figure 8.6: Comparison of data and MC prediction in the signal region after all cuts in the ee and $\mu\mu$ regions, using the scaled Drell-Yan contributions and the data-driven W+jets estimate.

all three channels, as can be seen in Figure 8.5.

The individual contributions for Equation 8.4 and the calculated S factor are listed in Table 8.5. Since the scale factors for ee and $\mu\mu$ are not compatible with unity, they are applied to the MC prediction. Their uncertainty is taken as a systematic uncertainty.

For the $e\mu$ channel, the average of the two predictions derived in the ee and $\mu\mu$ channel is calculated, and its difference from unity (3.5%) including $+1\sigma$ deviation (5.1%) is taken as a systematic uncertainty on the Drell-Yan modelling, resulting in an 8.6% systematic uncertainty. The more conservative approach used for the $e\mu$ channel is justified by the fact that the efficiencies from the ee and $\mu\mu$ channels have to be extrapolated.

The final predicted numbers of Drell-Yan background events with statistical and systematic uncertainties are shown in Table 8.6. The $\not\!\!E_{T,rel}$ distribution in the signal region with the scaled Drell-Yan and data-driven W+jets estimate from Section 8.1 is shown in Figure 8.6.

8.3 Top

Both single top decays $W^-t \to W^-bW^+/W^+\bar{t} \to W^+\bar{b}W^-$ and top pair decays $t\bar{t} \to bW^+\bar{b}W^-$ have a W^+W^- pair in the final state, and are therefore quite difficult to separate from W^+W^- decays. However, they are characterised by their high hadronic activity arising from the b quarks in the final state. To remove the majority of the top background events, jets and b-tagged jets are vetoed as described in Section 7.2. In [85] the efficiency of the jet vetoes on the top background is estimated using two different methods, which agree very well. The estimated values with statistical and systematic uncertainties are listed in Table 8.7.

$ee + \not{E}_{T,rel}$	$14.0 \pm 2.0(\text{stat}) \pm 2.9(\text{syst})$
$e\mu + \not{E}_{T,rel}$	$70.8 \pm 5.2(\text{stat}) \pm 14.4(\text{syst})$
$\mu\mu + \not{E}_{T,rel}$	$25.2 \pm 2.9(\text{stat}) \pm 5.1(\text{syst})$
Combined	$110.0 \pm 6.2(\text{stat}) \pm 22.4(\text{syst})$

Table 8.7: Summary of estimated top background events in the signal region [85]. Numbers given are expected events for 4.7 fb^{-1}.

8.4 Other Diboson Events

The contributions of other diboson decays, WZ, ZZ, $W\gamma$ and $W\gamma^*$, are estimated using MC simulation. WZ and ZZ decays can mimic W^+W^- production, if one or more leptons are missed in reconstruction or identification, leading to additional $\not{E}_{T,rel}$. $ZZ \to \ell^+\ell^-\nu\nu$ decays are effectively suppressed by the Z mass veto. In $W\gamma^{(*)}$ decays, converted photons can be mistaken as electrons. The MC predictions of the diboson decays are listed in Table 8.8. The systematic uncertainties arising from the use of the MC simulation are calculated using the methods described in 9 and are given here as well.

$ee + \not{E}_{T,rel}$	$8.6 \pm 1.2(\text{stat}) \pm 1.9(\text{syst})$
$e\mu + \not{E}_{T,rel}$	$36.2 \pm 2.9(\text{stat}) \pm 3.5(\text{syst})$
$\mu\mu + \not{E}_{T,rel}$	$12.8 \pm 0.6(\text{stat}) \pm 2.0(\text{syst})$
Combined	$57.6 \pm 3.2(\text{stat}) \pm 7.4(\text{syst})$

Table 8.8: Summary of estimated background events from other diboson decays in the signal region. Numbers given are expected events for 4.7 fb^{-1}.

8.5 Cosmic Muons

In rare cases, muons originating in air showers caused by high-energy cosmic rays in the atmosphere ("cosmic muons") can be background to W^+W^- decays. The rate of cosmic ray muons with $E_\mu > 1$ GeV is ≈ 160 Hz per square metre, with a majority over the threshold of 15 GeV [2]. In the ATLAS cavern, the rate is still considerable. To estimate the rate of muon triggers in coincidence with the LHC beam clock, a low luminosity run with eight low-intensity bunches from 2010 is used, where almost no muons from collisions are expected. The cosmic trigger rate of the lowest-threshold muon trigger from this data is at 4 Hz. Given the LHC frequency, the probability of a cosmic muon trigger is $\approx 4.4 \times 10^{-5}$ per bunch crossing. This gives a rate of

$\approx$ 500 Hz for typical LHC conditions in 2011[1], which is consistent with online observations during data taking.

This high trigger rate makes it difficult to use methods based on calculating cosmic muon cut efficiencies and multiplying them with the trigger rate, since many millions of reference cosmic muon events would have to be collected to obtain the expected small values with any statistical accuracy. Therefore, characteristic features of events with cosmic muons are used to estimate their number from the collected data itself.

There are two possible scenarios that lead to triggering and selecting an event involving a cosmic muon, both involving a high-energy cosmic muon passing close to a real interaction vertex during a collision. The first scenario - the only one considered by most analyses - is that a cosmic muon is reconstructed as two oppositely charged muons originating in the centre of the detector. In the second scenario, only the "lower half" of the cosmic muon track is reconstructed and identified as a muon track, and the track in the upper half of the detector lies outside the time window of the collision or is otherwise missed due to tracking and selection inefficiencies. This leads to a well reconstructed, isolated muon, accompanied by missing energy pointing in the opposite direction. If, by chance, such an event happens close to either a misidentified hadronic jet or a W decay, the event could possibly end up in the signal region.

To exclude the first scenario, it is used that a cosmic muon track is mostly a straight line in the inner detector. It is therefore expected that $\Delta\phi \approx \pi$, and $\eta_1 + \eta_2 \approx 0$. From data collected with cosmic muon triggers, 99.7% of cosmic muons that are reconstructed as two muons lie inside $|\eta_1 + \eta_2| < 0.002$ and $|\Delta\phi - \pi| < 0.0015$. In the W^+W^- cross-section measurement, just one event is found in this region after the preselection. This event subsequently failed the Z mass cut. Therefore, the cosmic muon contribution from this scenario can safely be neglected.

To estimate the background contribution from the second scenario, the analysis is repeated, leaving out the muon vertexing cuts described in Section 6.2 as well as the requirement to have at least three tracks at the primary vertex. Events where the difference in the z-positions between the two selected leptons is large $\Delta z_0(\mu, l) > 20$ mm are then considered as a background control sample. In Figure 8.7 the distribution of Δz_0 in this sample is shown. The transition from a peak around zero to an approximately flat background at around 20 mm is clearly visible.

This control region may also contain muons from other simultaneous collisions. This method therefore also constrains contributions from overlapping interaction vertices, which would yield a similar effect.

In the control region between 20 mm and 120 mm, a total of 37 events

[1]The bunch spacing is taken as 50 ns and half of the bunch crossings are marked as collisions.

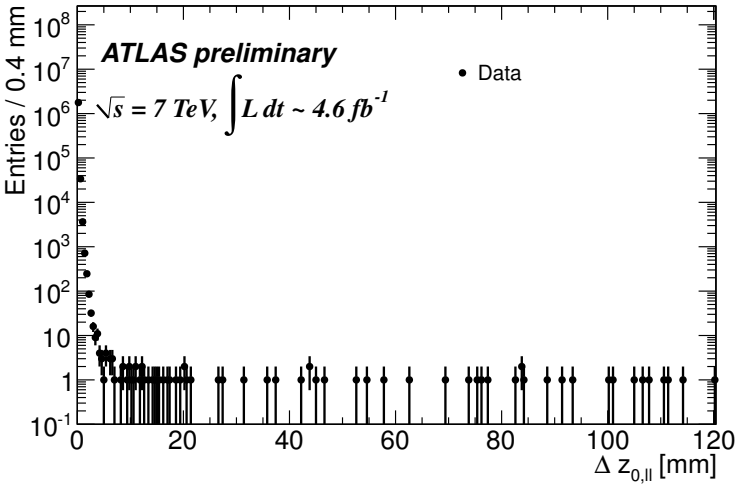


Figure 8.7: $\Delta z_0(\mu, l)$ after regular preselection in the cosmics control region with loosened vertexing cuts. The control region is defined as the range 20 mm $< \Delta z_0(\mu, l) <$ 120 mm.

that pass all selection cuts are found. Extrapolating to the accepted region of $\Delta z_0(\mu, l) < 1$ mm, this yields an estimate of ≈ 0.4 cosmic muon or pileup events in the signal region, an estimate that is expected to be further reduced by the d_0 requirement. This demonstrates that in this analysis, the cosmic muon background is negligible.

8.6 Higgs Boson Decays

A final and as yet hypothetical background to the W^+W^- cross-section measurement is the decay of the Higgs boson into a W boson pair. However, since the mass of the Standard Model Higgs boson has already been restricted to 115.5 GeV $-$ 127 GeV and must therefore be lighter than two W bosons, one of the W bosons must be virtual, strongly suppressing such decays at lower Higgs masses. To get a conservative estimate of the possible contribution, the MC simulation of Higgs decays with a Higgs mass of 130 GeV has been used. Any lower Higgs mass would result in a lower branching ratio into W pairs.

The resulting number of events in the signal region at 4.7 fb^{-1} were 2.9, 19.6 and 5.4 in the ee, $e\mu$ and $\mu\mu$ channels, with 27.9 combined events. This is only 2.8% of the expected W^+W^- signal in the signal region.

Since both the contribution of Standard Model Higgs decays with maximal allowed mass is very small, and since it is not yet established if a Standard Model Higgs boson exists, this background is not subtracted from the measurement.

8.7 Summary

In this chapter, all relevant backgrounds to di-leptonic decays of the W^+W^- signal process have been presented. For all backgrounds except diboson production, a data-driven estimate or cross-check is used, reducing systematic uncertainty due to MC mismodelling. For each background, the events predicted in the signal region have been provided, including an estimate of systematic and statistical uncertainties. The sum of all expected background contributions in the signal region is given in Table 8.9. In Chapter 10 this information is used to subtract the background contribution from the observed events in the signal region and to estimate the systematic uncertainty on the resulting cross-section.

$ee + \not{E}_{T,rel}$	$114.4 \pm 7.1(\text{stat}) \pm 11.5(\text{syst})$
$e\mu + \not{E}_{T,rel}$	$317.5 \pm 9.6(\text{stat}) \pm 35.8(\text{syst})$
$\mu\mu + \not{E}_{T,rel}$	$113.1 \pm 7.2(\text{stat}) \pm 6.8(\text{syst})$
Combined	$545.0 \pm 13.9(\text{stat}) \pm 50.1(\text{syst})$

Table 8.9: Summary of the total estimated background events in the signal region. Numbers given are expected events for 4.7 fb^{-1}.

Chapter 9

Systematic Uncertainties

In this measurement, the estimation of possible systematic uncertainties is very important. Even though the data was recorded in the first full year of running of the ATLAS detector, it is already possible to constrain most of the uncertainties in the detector performance to less than 3%. In the following, estimates for all considered systematic uncertainties on the W^+W^- signal acceptance and efficiency are presented and combined to an overall $\mathcal{A}\epsilon$ acceptance uncertainty.

9.1 Lepton Uncertainties

The passage of electrons and muons through the ATLAS detector in MC simulations is generally well modelled by the Geant4-based full detector simulation. The level of detail of the simulated geometry is, however, limited, and both mismodelling of components [32] as well as uncertainty on the alignment and the magnetic field can lead to systematic uncertainty on momentum and energy measurements. Therefore, the relevant single-lepton efficiencies, scales and resolutions have to be determined from data.

There are several lepton-related systematic uncertainties that need to be estimated for both muons and electrons:

- Uncertainty on the lepton trigger efficiency
- Uncertainty on the lepton reconstruction efficiency
- Uncertainty on the efficiency of lepton identification and quality requirements
- Uncertainty on lepton isolation
- Uncertainty on lepton energy scale and resolution

The estimation of these uncertainties and their effects on this analysis are treated in the following.

9.1.1 Efficiency Measurement by Tag and Probe

The main tool for estimating efficiencies for both muons and electrons is the *tag-and-probe* method, aiming at an unbiased selection of very loosely selected *probe* leptons using primarily other information in the event as *tag* requirements. For most measurements, the probe is a loose calorimeter cluster for electrons and an isolated track in either the inner detector or the muon spectrometer for muons. The tag criterion is usually one well-reconstructed, triggered lepton that combines with the probe to the invariant mass of either the Z or the J/ψ resonance. Probes in W boson decays are also used, tagged using the missing transverse energy $\not{E}_T$ trigger and requiring a minimum transverse mass for the combination of $\not{E}_T$ and the probe.

Once an unbiased sample of probes is selected, the efficiencies of reconstruction, identification and quality requirements can be evaluated. The accuracy and possible resolution in η, ϕ and transverse momentum p_T is limited by the available statistics, and by the accuracy of the subtraction of background processes, based on both MC simulation and sideband subtraction. The resulting efficiencies are then compared with the predictions from MC simulations, and sets of *scale factors* including estimates of systematic uncertainties are provided. The results using this method for electron reconstruction, identification and quality efficiencies are given in [74], for muon reconstruction and identification efficiencies in [87]. Updated numbers for 2011 data are provided internally by the ATLAS performance groups [79, 88]. The efficiency of the isolation requirements for leptons are estimated using tag-and-probe as well [77], using only Z boson decays.

The tag-and-probe method can also be used to measure the trigger efficiency. For this, both tag and probe are taken to be well-reconstructed leptons from Z or J/ψ decays, but the tag is also required to have fired the trigger. The fraction of probe muons firing the trigger in a certain region can then be used as an unbiased sample for efficiency estimation. The trigger efficiency measurements are described in [74, 89].

9.1.2 Lepton Energy Scale and Resolution

The uncertainty on the energy scale and resolution also contribute to systematic uncertainties. To improve and cross-check the test beam calibration, an in-situ calibration using Z and J/ψ decays is performed. The resolution of both scale and resolution is extracted from a fit of data to the peak of the respective resonance [74, 43, 44], yielding both a scale and resolution correction (see Chapter 6) as well as an estimate of uncertainties on this measurement.

Uncertainty	$ee + \not{E}_{T,rel}$	$e\mu + \not{E}_{T,rel}$	$\mu\mu + \not{E}_{T,rel}$	Total
Lepton Trigger	0.3%	0.5%	0.6%	0.5%
Electron recon. SF	1.6%	0.8%	-	0.8%
Electron ID SF	2.3%	1.0%	-	1.0%
Muon recon. & ID SF	-	0.4%	0.7%	0.4%
Electron Scale	0.9%	0.3%	-	0.3%
Electron Resolution	0.1%	0.02%	-	0.02%
Muon Scale	0.02%	0.2%	0.9%	0.3%
ID Muon Resolution	-	0.01%	0.02%	0.01%
MS Muon Resolution	-	0.01%	0.1%	0.01%
Lepton Isolation	4.0%	2.3%	2.3%	2.3%
Total Lepton uncertainties	5.0%	2.7%	2.6%	2.6%

Table 9.1: Summary of systematic uncertainties on the W^+W^- acceptance and efficiency $\mathcal{A}\epsilon$ due to leptons.

9.1.3 Combination

For each of the listed uncertainties, the analysis of the W^+W^- signal sample is repeated while varying the respective factor of the event weight (for efficiencies) or energy scale factor (for energy scale and resolution) by $\pm 1\sigma$ [77]. The differences of the $+1\sigma$ and -1σ acceptances to the nominal acceptance are then averaged. It has to be noted that if any momentum of a particle is changed by Δp, the transverse component of Δp is subtracted from the missing transverse energy $\not{E}_T$. This approach correctly propagates the impact of lepton momentum scale and resolution on the $\not{E}_T$. Moreover, it also avoids unwanted correlations: If the all uncertainty on the missing transverse energy is evaluated at once, the result would be correlated with all other scale and resolution uncertainties.

In Table 9.1 the resulting lepton uncertainties are given. The uncertainty on the lepton isolation (due to jet measurements and pileup) dominates in all channels with 4.00% (ee), 2.28% ($e\mu$), 2.30% ($\mu\mu$) and 2.25% (combined[1]). In total, the lepton-related systematic uncertainties are 4.97% (ee), 2.67% ($e\mu$), 2.56% ($\mu\mu$) and 2.63% (combined).

[1]The uncertainty for the combination is calculated using all channels at the same time, correctly taking inter-channel correlations into account. In this case, anticorrelations reduce the value for the combination slightly.

9.2 Jet Uncertainties

There are several important uncertainties in this analysis originating in the use of hadronic jets and b-jets. The main uncertainties arise from the description of jets in MC simulation, in the jet energy measurement, and in the b-jet identification.

9.2.1 MC Description

The description of additional jet radiation in W^+W^- production MC simulations is not uniform across different MC generators. Since a proper estimate of the uncertainty on the use of a model is difficult, the difference in signal acceptance between the Alpgen and MC@NLO generators is used as an estimate on the uncertainty. MC@NLO matches NLO calculations of hard QCD processes with parton showers, whereas Alpgen uses tree-level calculations for emissions of additional partons. In Section 7.3 the relative difference is found to be 5.0% for all channels.

9.2.2 Jet Energy Scale and Resolution

The uncertainty on the jet energy scale and resolution is currently one of the major contributors to the experimental uncertainty. While calorimeter intercalibration is achieved using the di-jet balance in minimum bias events, the absolute scale is refined using events where a Z boson recoils against a single jet and subsequently decays into electrons. The uncertainty on the energy scale calibration is $\approx 5\%$ [81].

However, since only the jet veto cut is affected by the jet resolution and scale in this measurement, the data-driven estimate of the jet veto efficiency from Section 7.3 can be used instead of propagating the measurement uncertainty. In this approach, the jet veto efficiency of W^+W^- signal events is corrected using the ratio of the jet veto acceptance of Z boson events in data to MC simulation, partially cancelling the effects of the jet energy scale uncertainty. The resulting uncertainty due to resolution and scale is 3.2% (ee), 1.9% ($e\mu$), 3.2% ($\mu\mu$) and 2.5% (combined). As for the lepton uncertainties, the effect of jet energy scale and resolution are propagated to the missing transverse energy $\not\!\!E_T$.

9.2.3 b-Jet Identification

The uncertainty on both the misidentification of jets as b-jets and of b-jets as regular jets also has to be taken into account. Due to the complexity of the b-tagging methods and the difficulty to select unbiased samples of b-jets, the uncertainty on the efficiencies is quite large, 5%-19% depending on the transverse jet momentum [82]. However, the b-tag rate is expected to be small

in the W^+W^- signal, and the propagation of the b-tag uncertainties to the acceptance results in an uncertainty of only $\approx 0.5\%$ in all channels.

9.3 Missing Transverse Energy Uncertainties

As described in Section 6.5, calorimeter clusters with high transverse momentum are replaced by analysis objects (e.g. electrons or jets) with improved calibration in the $\not\!\!E_T$ calculation. Uncertainties on analysis objects - lepton and jet scale and resolution uncertainties - have, however, already been considered in the respective sections. The remaining uncertainty arises from pileup (see Section 2.2.4) and the estimated measurement uncertainty on calorimeter deposits with low transverse momentum not associated with a jet. These uncertainties are estimated by varying MC pileup simulation settings, changing MC shower models and introducing additional detector material into the simulation [83]. Applying the estimated uncertainties on the W^+W^- acceptance results in a combined uncertainty of 1.2% due to pileup and 0.8% due to low transverse momentum deposits.

9.4 Uncertainties from PDF and Scale

To estimate the acceptance of the W^+W^- signal, events are simulated using the CT10 NLO parton distribution function [65] (see Sections 2.2.2 and 5.1). This PDF comes with a set of 52 error eigenvectors, which provide reweighting factors for individual events, depending on the momentum fraction of the incoming partons. To obtain a symmetric uncertainty, the prescription in [65] of separately adding positive and negative differences to the central values in quadrature and averaging them is used. This approach yields a PDF uncertainty of 0.8% on the combined acceptance.

In addition, the uncertainty between different PDF sets is estimated by reweighting the simulated W^+W^- events to the MRST 2008 NLO 68% CL PDF set [90] and re-estimating the efficiency. The resulting combined uncertainty is 0.9%. In total, the uncertainty on the PDF sets is taken to be the sum in quadrature of these uncertainties, resulting in an uncertainty of 1.2%.

A separate study determined the acceptance uncertainty from varying the renormalisation and factorisation scales of the MC simulation by a factor of two in either direction [22, 77], resulting in an additional 1.5% uncertainty.

9.5 Summary

In Table 9.2, the contributions of the previously discussed systematic uncertainties on the acceptance and efficiency $\mathcal{A}\epsilon$ (see Section 4.3) are summarised.

Uncertainty	$ee + \not{E}_{T,rel}$	$e\mu + \not{E}_{T,rel}$	$\mu\mu + \not{E}_{T,rel}$	Total
PDF	1.4%	1.4%	1.4%	1.2%
Scale (μR , μF)	2.1%	1.7%	0.6%	1.5%
Jet veto (MC modelling)	5.0%	5.0%	5.0%	5.0%
Lepton uncertainties	5.0%	2.7%	2.6%	2.6%
B tagging SF	0.4%	0.5%	0.5%	0.5%
JES & JER	3.2%	1.9%	3.2%	2.5%
MET pile-up	2.5%	0.7%	2.8%	1.2%
MET Cluster	1.8%	0.5%	1.8%	0.8%
Total acceptance uncertainty	8.7%	6.4%	7.5%	6.7%

Table 9.2: Summary of systematic uncertainties on the W^+W^- acceptance and efficiency $\mathcal{A}\epsilon$.

Since each contribution to the total uncertainty is derived from statistically independent sources and proper propagation to missing transverse energy is carried out, the uncertainties can be added in quadrature. The total systematic uncertainty on the acceptance $\mathcal{A}\epsilon$ is then 8.7% for ee, 6.4% for $e\mu$, 7.5% for $\mu\mu$ and 6.7% for the combination of all channels. In the next section, this uncertainty will be propagated to the measured value of the W^+W^- cross-section.

Chapter 10

Results

At this point, all the components presented in the previous chapters can be combined to a measurement of the W^+W^- production cross-section. We now have all necessary inputs to apply Equation 4.1 from Section 4.3:

$$\sigma_{W^+W^-} = \frac{N - B}{\mathcal{L} \cdot \mathcal{A}\epsilon \cdot \mathrm{BR}}$$

Using the number of expected signal events S and the NLO cross-section prediction $\sigma_{\mathrm{NLO}} = 47.0$ pb, this equation can be rewritten in a form more closely mirroring the experimental approach:

$$\frac{\sigma_{W^+W^-}}{\sigma_{\mathrm{NLO}}} = \frac{N - B}{S}$$

$$S = \sigma_{\mathrm{NLO}} \cdot \mathcal{L} \cdot \mathcal{A}\epsilon \cdot \mathrm{BR}$$

The number of observed events N is obtained in Section 7.4, the background estimate B and its uncertainty ΔB in Chapter 8, the integrated luminosity $\mathcal{L}$ and $\Delta\mathcal{L}$ in Section 5.2, the signal acceptance and efficiency $\mathcal{A}\epsilon$ in Section 7.3 and its uncertainty $\Delta\mathcal{A}\epsilon$ in Chapter 9. The W boson branching ratio to leptons, assuming lepton universality, is $\mathrm{BR} = 0.105 \pm 0.0018$ [2].

In Table 10.1, the estimated backgrounds $B \pm \Delta B$ and expected signal events S are summarised and compared to the observed events N. All expected numbers are normalised to $\mathcal{L} = 4.7$ fb^{-1}. Both systematic and statistical uncertainties are given for all predictions.

In the following, the statistical, systematic and luminosity uncertainties are collected and the result of the cross-section measurement is presented.

	$ee + \not\!\!E_{T,rel}$	$e\mu + \not\!\!E_{T,rel}$	$\mu\mu + \not\!\!E_{T,rel}$	Combined
Top (data-driven)	$14.0 \pm 2.0 \pm 2.9$	$70.8 \pm 5.2 \pm 14.4$	$25.2 \pm 2.9 \pm 5.1$	$110.0 \pm 6.2 \pm 22.4$
W+jets (data-driven)	$19.8 \pm 0.5 \pm 10.5$	$68.3 \pm 2.5 \pm 30.1$	$5.1 \pm 0.9 \pm 2.0$	$93.2 \pm 2.7 \pm 40.7$
Z+jets (MC + DD SF)	$72.0 \pm 6.7 \pm 3.2$	$142.2 \pm 7.1 \pm 12.5$	$70.0 \pm 6.5 \pm 3.5$	$284.2 \pm 11.7 \pm 17.2$
Other Dibosons	$8.6 \pm 1.2 \pm 1.9$	$36.2 \pm 2.9 \pm 3.5$	$12.8 \pm 0.6 \pm 2.0$	$57.6 \pm 3.2 \pm 7.4$
Sum of Background	$114.4 \pm 7.1 \pm 11.5$	$317.5 \pm 9.6 \pm 35.8$	$113.1 \pm 7.2 \pm 6.8$	$545.0 \pm 13.9 \pm 50.1$
W^+W^- MC Signal	$92.4 \pm 1.5 \pm 8.1$	$639.6 \pm 3.6 \pm 41.1$	$142.9 \pm 1.8 \pm 10.7$	$874.9 \pm 4.5 \pm 58.5$
Expected Events	$206.8 \pm 7.3 \pm 14.1$	$957.1 \pm 10.3 \pm 54.5$	$256.0 \pm 7.4 \pm 12.7$	$1419.9 \pm 14.6 \pm 77.0$
Observed Events	196	1041	287	1524

Table 10.1: Summary of observed and expected events together with the background estimates. The W^+W^- signal MC is scaled to 47 pb [25], all estimates are scaled to $\mathcal{L} = 4.7$ fb^{-1}.

10.1 Uncertainties on the Measurement

In all channels the number of observed events N is large enough that the Gaussian approximation holds. Therefore, the relative statistical uncertainty due to limited events is given by:

$$\frac{\Delta_{\text{stat}}\sigma_{W^+W^-}}{\sigma_{W^+W^-}} = \frac{\sqrt{N}}{N-B}$$

This uncertainty is largest (17%) in the low-statistics ee channel. The $\mu\mu$ and $e\mu$ channels have a statistical uncertainty of 9.7% and 4.5% respectively. Combining all channels reduces the uncertainty to 4.0%.

Since almost all systematic uncertainties on the background processes are estimated from data, it is assumed that they are not correlated with the uncertainties on the signal efficiency. Therefore, all systematic uncertainties on B, $\mathcal{A}\epsilon$ and BR are combined to a total systematic uncertainty using Gaussian error propagation:

$$\frac{\Delta_{\text{syst}}\sigma_{W^+W^-}}{\sigma_{W^+W^-}} = \sqrt{\left(\frac{\Delta A\epsilon}{A\epsilon}\right)^2 + \left(\frac{\Delta B}{N-B}\right)^2 + \left(\frac{\Delta\text{BR}}{\text{BR}}\right)^2}$$

The uncertainties ΔB and ΔA now include the (almost negligible) MC statistical uncertainties on background and signal samples, since they do not depend on the statistics of the measurement. Again, the ee channel is most strongly affected with an uncertainty of 18%, mainly due to the uncertainty on the W+jets estimate. The $\mu\mu$ channel has an uncertainty of 9.4%, the $e\mu$ channel an uncertainty of 8.3%. The systematic uncertainty on the combination of all channels is 8.6%.

To propagate the relative uncertainty on the luminosity $\Delta\mathcal{L}/\mathcal{L} = 3.7\%$ to the cross-section, two contributions have to be taken into account: The linear contribution on the explicit term $\mathcal{L}$ in Equation 4.1, and the effect of a changing luminosity estimate on the background estimates B. Since almost all background contributions are estimated or normalised using a data-driven approach, only the background from other di-boson events is affected. The total luminosity uncertainty is therefore:

$$\frac{\Delta_{\text{lumi}}\sigma_{W^+W^-}}{\sigma_{W^+W^-}} = \sqrt{\left(\frac{\Delta\mathcal{L}}{\mathcal{L}}\right)^2 + \left(\frac{B_{\text{other di-boson}}}{N-B}\frac{\Delta\mathcal{L}}{\mathcal{L}}\right)^2}$$

The additional uncertainty due to the di-boson prediction is $0.2\% - 0.4\%$, yielding 4.1% (ee), 3.9% ($e\mu$), 4.0% ($\mu\mu$) and 3.9% (combined) of uncertainty due to luminosity measurement.

10.2 Cross-Section Results

Combining all three channels of the measurement yields the result

$$\sigma_{W^+W^-} = 52.6 \pm 2.1(\text{stat}) \pm 4.5(\text{syst}) \pm 2.1(\text{lumi})\text{pb}$$

This result has a total uncertainty of 10% and is well compatible ($+1\sigma$) with the SM NLO prediction $\sigma_{NLO} = 47.0^{+2.0}_{-1.5}$ pb. In Table 10.2, the cross-sections extracted from each separate channel are listed, all compatible with the combined measurement.

Channel	Cross-Section
$ee + \not{E}_{T,rel}$	$41.5 \pm 7.1(\text{stat}) \pm 7.8(\text{syst}) \pm 1.7(\text{lumi})$ pb
$e\mu + \not{E}_{T,rel}$	$53.2 \pm 2.4(\text{stat}) \pm 4.4(\text{syst}) \pm 2.1(\text{lumi})$ pb
$\mu\mu + \not{E}_{T,rel}$	$57.2 \pm 5.6(\text{stat}) \pm 5.4(\text{syst}) \pm 2.3(\text{lumi})$ pb
Combined	$52.6 \pm 2.1(\text{stat}) \pm 4.5(\text{syst}) \pm 2.1(\text{lumi})$ pb

Table 10.2: Results of the cross-section calculation. All results are compatible with the SM NLO Prediction of $47.0^{+2.0}_{-1.5}$ pb within the uncertainties.

At this point, the transverse momentum of the di-lepton system $p_{T,\ell\ell}$ has not been used for selecting events. Its distribution for all analysis channels combined is shown in Figure 10.1.

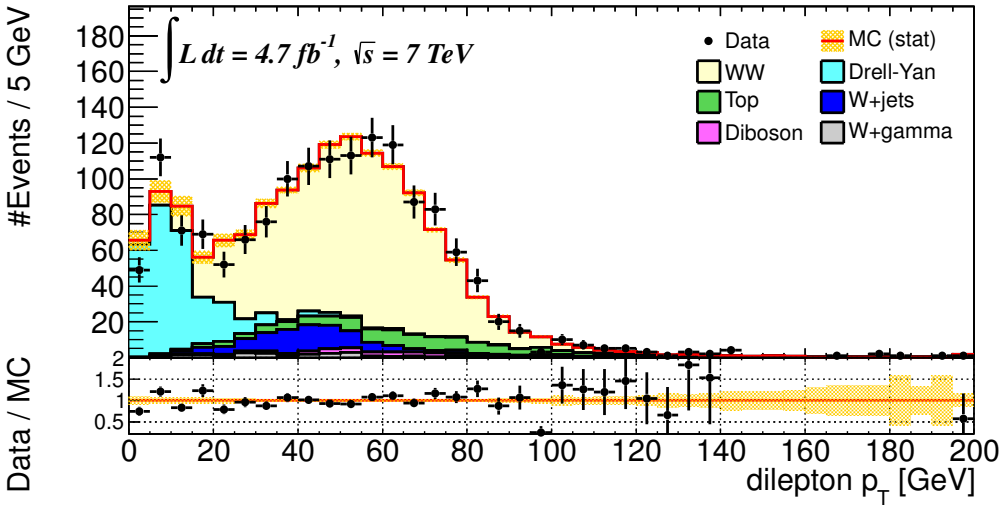


Figure 10.1: Transverse momentum of the di-lepton system in selected events, combining all analysis channels. The W^+W^- cross-section is scaled to the measured 52.6 pb. The W+jets contribution is estimated from data, the Drell-Yan Monte Carlo is scaled using a data-driven scale factor.

The W^+W^- production MC is scaled to the measured value to illustrate the result. Events from Drell-Yan decays are clustered at low $p_{T,\ell\ell}$, the agreement in that region validating the data-driven scale factor. W^+W^- decays dominate at $p_{T,\ell\ell} > 20$ GeV, the shape of the distribution being well modelled by the MC prediction. The other relevant backgrounds are W+jets and $t\bar{t}$ decays, with $t\bar{t}$ having a higher $p_{T,\ell\ell}$ than W+jets due to the high top mass. In total, very good agreement between data and MC simulation is observed.

10.3 Other Measurements of the W^+W^- Production Cross-Section

LEP2

The production of W^+W^- pairs has first been observed and measured at the LEP2 electron-positron collider at CERN in the years $1996 - 2000$ by the four experiments OPAL, ALEPH, DELPHI and L3 to a precision of $1 - 3\%$ at centre-of-mass energies from 161.3 GeV $< \sqrt{s} <$ 206.6 GeV [91]. From the kinematic cutoff at $2m_W = 161$ GeV the cross-section measured at LEP increases to 17.15 ± 0.25 pb at 206.6 GeV. All measurements at LEP agree very well with the SM predictions at the different energies, the theoretical uncertainty being at the 0.5% level.

TeVatron

At the TeVatron, the W^+W^- production cross-section in proton-antiproton collisions was measured at $\sqrt{s} = 1.96$ TeV by the D0 [92] and CDF [93] collaborations with an uncertainty of 20% and 15% respectively. On average, the parton-parton centre-of-mass energy for W^+W^- production has been $\sqrt{\hat{s}} = 245$ GeV. The resulting cross-sections are:

$$\sigma_{D0} = 11.5 \pm 2.1(\text{stat} + \text{syst}) \pm 0.7(\text{lumi}) \text{ pb}$$
$$\sigma_{CDF} = 12.1 \pm 0.9(\text{stat})^{+1.6}_{-1.4}(\text{syst}) \text{ pb}$$

Both measurements agree with the theoretical prediction of the Standard Model for the TeVatron conditions of $\sigma_{NLO} = 11.7 \pm 0.7$ pb [92].

LHC

At the LHC, the author has been involved in several measurements of the W^+W^- cross-section by the ATLAS experiment using proton-proton collisions at $\sqrt{s} = 7$ TeV [94, 95, 96, 85]. The average parton-parton centre-of-mass energy for W^+W^- boson production derived from MC@NLO simulation samples is $\sqrt{\hat{s}} = 277$ GeV. Parts of this thesis have been published in

the latest of these analyses [85]. A further paper with improved selection and updated uncertainties is in preparation.

The CMS experiment at the LHC has also published a measurement of the W^+W^- production cross-section [97], obtaining

$$\sigma = 52.4 \pm 2.0(\text{stat}) \pm 4.5(\text{syst}) \pm 1.2(\text{lumi}) \text{ pb}$$

This result is compatible with the one measured in this thesis and is obtained with very similar beam conditions. The similarity of the results suggests that either a common systematic error is affecting both experiments - e.g. MC modelling of the jet veto - or that the W^+W^- production cross-section really exceeds the SM prediction.

The question arises if this difference could conceivably be caused by the contributions of a Standard Model Higgs boson. However, a SM Higgs boson even at the highest not yet excluded mass of $m_H \approx 130$ GeV would contribute only 28 events to the current selection, corresponding to 2.8% or 1.5 pb of the W^+W^- cross-section. If the Higgs boson should be discovered at a later date, its contributions estimated from its mass should be subtracted from this measurement.

Chapter 11

Summary and Outlook

In the year 2011, the ATLAS experiment at the Large Hadron Collider at CERN in Geneva recorded proton-proton collisions at the centre-of-mass energy $\sqrt{s} = 7$ TeV. The recorded events where detector conditions were suitable for physics analysis correspond to a total integrated luminosity of $\mathcal{L} = 4.7$ fb^{-1}.

This thesis presents the first measurement of the W^+W^- di-boson production cross-section with the full data set of 2011, using W^+W^- decays to two electrons and neutrinos, two muons and neutrinos, and one electron, one muon and neutrinos, respectively. Regions of candidate events are defined for each decay channel, and data-driven techniques and Monte Carlo simulations are used to estimate contributions from background processes.

In total, 1524 candidate events with an estimated background of 545 events are used to determine the cross-section. A full set of systematic uncertainties is considered, and its effects are estimated.

The measured W^+W^- production cross-section is:

$$\sigma_{W^+W^-} = 52.6 \pm 2.1(\text{stat}) \pm 4.5(\text{syst}) \pm 2.1(\text{lumi}) \text{ pb}$$

The result is in good agreement with the Standard Model prediction at next-to-leading order of $47.0^{+2.0}_{-1.5}$ pb and shows a difference in total of $+5.6 \pm 5.2$ pb, or $+11.9\% \pm 11.1\%$.

A similar measurement at the CMS experiment at the LHC yields a cross-section of $\sigma = 52.4 \pm 2.0(\text{stat}) \pm 4.5(\text{syst}) \pm 1.2(\text{lumi})$ pb, indicating that either a common systematic error affects both measurements, or the cross-section really exceeds the prediction. Any Standard Model Higgs boson contribution to the cross-section measurement is estimated to be < 1.5 pb.

The main limitations of the measurement at present are systematic uncertainties, primarily the uncertainty on luminosity and the uncertainty associated with the jet veto cut. Further work in the ATLAS luminosity and performance groups will yield many improvements, but it is also necessary to search for new ways of reducing the uncertainties on jets, both for the theoretical estimate and in the experimental measurement. A paper presenting the ATLAS W^+W^- cross-section measurement with re-optimised cuts, using the latest systematic uncertainty estimates and including limits on anomalous triple gauge couplings is currently in preparation.

At the time of writing, the ATLAS experiment has just passed the mark of 6.2 fb^{-1} of recorded proton-proton collision data at the centre-of-mass energy $\sqrt{s} = 8$ TeV. It is expected to take at least 15 fb^{-1} of data in 2012, enough to either discover or exclude a Standard Model Higgs boson.

In the case of the discovery of a Higgs boson, its decays to W^+W^- pairs will be used for characterisation and for the investigation of its properties. As the main background process for Standard Model Higgs decays, the W^+W^- production cross-section will have to be measured and understood very well at higher energy. In addition, higher event statistics enable the extraction of differential cross-sections, with increased power to set limits on anomalous triple gauge couplings.

If a Standard Model Higgs boson should be excluded, the W^+W^- cross-section measurement becomes an essential tool in the investigation of the electroweak sector. Without a Standard Model Higgs boson, the predicted cross-section of the scattering of the longitudinal component $W_LW_L \to W_LW_L$ violates unitarity at 1.2 TeV [98]. This makes new physics at this scale necessary, new physics that could appear in a future measurement of the W^+W^- production cross-section.

Appendix

Appendix A

The GangaTasks Software Package

A.1 The Ganga Grid User Interface

For large-scale analyses similar to the one presented in this thesis, it is necessary to effectively use the computing resources of the LHC Computing Grid. The Grid user interface Ganga [52] is an IPYTHON-based program that assists physicists in this task by simplifying job configuration and submission, providing the ability to submit and monitor large numbers of similar jobs, and by providing an easily scriptable environment.

In the course of this thesis, the Ganga job repository was rewritten to an XML-based disk repository by the author to cope with increased demand in the LHC era [53]. The repository is a core component of Ganga, storing all job metadata. The new repository is more tolerant in the case of system instabilities and supports multiple concurrent sessions even on different machines, while at the same time improving startup performance by lazy loading, reading only requested information from disk when it is needed. The new repository architecture also enables the store of configured application, backend and dataset objects.

Ganga itself is scriptable, but at the time did not provide automatic resubmission in the case of grid problems and workflow automation. To improve this state the GangaTasks software package was developed by the author, first as a package for ATLAS users and later integrated into Ganga as a core feature.

```
app = Athena(option_file = './my_analysis.py')
app.prepare()
t = AnaTask()
t.analysis.application = app
t.analysis.files_per_job = 5
dsets = ['data11_7TeV.periodE.physics_Muons.PhysCont.AOD/',
         '..periodE...', ...]
t.initializeFromDatasets(dsets)
```

Figure A.1: Example of a GangaTasks configuration listing to process physics data.

A.2 GangaTasks

GangaTasks is an tool for automatic job submission, throttling, steering, and resubmission, providing plug-in hooks for job result checking and data management currently used for the ATLAS experiment. Event generation, simulation and reconstruction chains can be set up, job dependencies are automatically taken into account.

In Figure A.1 an example of a GangaTasks configuration and submission is given, running the script `my_analysis.py` on all datasets given in the list. The output of each job is checked and inserted into a separate ATLAS data set for easy retrieval. Throttling is controlled by setting the maximum number of jobs running at once, making it possible to test analyses on a small set of files first before using large amounts of computing time. If job dependencies become available, waiting jobs are automatically submitted. Jobs that fail will be automatically resubmitted a limited number of times to compensate for spurious problems on grid sites. If a majority of jobs fail, the task will be halted for user inspection. A command-line interface with visual elements (see Figure A.2) allows fast inspection of job and task status.

An overview over the GangaTask design is given in Figure A.3. Each task consists of one or more sub-tasks (transforms). Interdependencies can be configured between subtasks, with both $1 \rightarrow N$ and $N \rightarrow 1$ dependencies supported.

Due to its job configuration, submission and resubmission capabilities, GangaTasks is not only a very useful tool for private MC productions, but also for analyses where large numbers of jobs need to be managed. Since the bookkeeping of successful job results is automated, the user can concentrate on the macromanagement of processing and on dealing with persistent failures.

Figure A.2: Graphical overview of a running task. Blue fields indicate a completed "partition", green fields indicate partitions where jobs are still running.

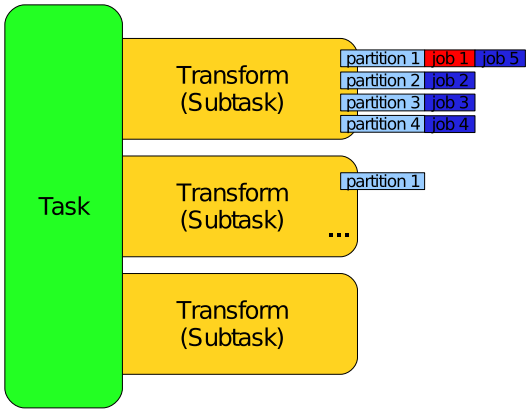


Figure A.3: Structure of a task in GangaTasks. Each task consists of one or more sub-tasks, which can have interdependencies. Each work unit (partition) is executed using one or more Ganga jobs. In case of job failure (e.g. job 1), a new job (job 5) is submitted and tries to process that work unit.

A.3 Summary

In the course of this thesis, the Ganga persistency layer and metadata repository was significantly improved in reliability and functionality. In addition, the GangaTasks package was added to the Ganga project, providing automation of MC production chains and large-scale analysis. The GangaTasks package has proven useful not only for this thesis, but also for the Ganga project, and was therefore adopted from the ATLAS part of Ganga into the core package.

Appendix B

The A4 Analysis Framework

In this section the High Energy Physics data format, processing toolset and analysis library A4 designed and written in the course of this thesis, is presented. It provides fast I/O of structured data using the GOOGLE protocol buffer library. The overall goal of A4 is to provide physicists with tools to work efficiently with billions of events, providing not only high speeds, but also automatic metadata handling, a set of UNIX-like tools to operate on A4 files, and powerful and fast histogramming capabilities.

B.1 Introduction

One common problem in High Energy Particle Physics computing is getting a reasonable balance between rapid and easy code development, usually the domain of scripting languages such as Python, and raw processing speed. However, new developments in computing outside High Energy Physics, in particular the adoption of the new C++11 standard and in general the availability of high-quality open source libraries, make it possible to improve the usability and readability of physics analysis codes without sacrificing processing speed.

The A4 project was started with the goal of processing and analyzing data for this thesis. Including Monte Carlo simulations, the processing of approximately one billion (10^9) events was necessary. From each event ≈ 6 kB of data were required for analysis, resulting in an expected dataset size of 6 TB. Criteria were a fast turn-around time for analyses, easy definition and generation of large numbers of diverse histograms and the possibility to quickly adapt the code to as yet unknown analysis requirements.

To achieve this, a file format with a standalone I/O library was designed (Sections B.2 and B.3). Additional libraries enable fast processing (Section B.4) and easy output handling (Sections B.5 and B.6). Conversion of results to the ROOT system is also provided (Section B.7). In the last section, some comparative numbers from basic benchmarks are presented.

```
message Lepton {
    optional double pt = 1;
    optional double eta = 2;
    optional double phi = 3;
    optional int32 charge = 4;
}
message PhysicsEvent {
    optional int32 run_number = 1;
    optional int32 event_number = 2;
    repeated Lepton electrons = 5;
    repeated Lepton muons = 6;
}
```

Figure B.1: Example of a protobuf message definition for a Physics event. The numbers are the field identifiers in the binary format. The fact that they are specified explicitly allows renaming the fields without changing the data on disk.

B.2 The Protocol Buffer Library

Motivated by reports [99] of higher performance and ease-of-use with respect to ROOT [51] trees, the GOOGLE protocol buffer (protobuf) library (used heavily at GOOGLE [100]) was chosen as a serialisation format. The protobuf library defines a fast binary format for *messages*. Message structures (so called 'descriptors') are defined in a simple C-like language in `.proto` files. An example is given in Figure B.1. From these, interfaces for different programming languages can be generated using the provided extensible compiler, `protoc`. A given interface contains the code describing a set of classes providing the (de)serialisation functionality tailored for a given `.proto` file and platform. The descriptors themselves can be serialised as protobuf messages, facilitating the inspection of arbitrary serialised messages without needing the descriptors at compile time.

In summary, the relevant features of the protobuf library are the following:

- Fast serialisation and de-serialisation of structured data in the form of *messages*
- Separation of data structure definition and code
- Extensible code generators for different programming languages[1]

[1] Java, C++ and Python are officially supported, others including C, Lisp, D, Go, Javascript, Matlab, Perl, and R, are available via add-ons

- Messages can be nested
- Repeated fields can store data in a manner similar to dynamic arrays
- Content can be omitted from optional fields, and does not take up space in this case
- Thread safety

One problem with multiple analysts working on the same dataset is ensuring that the files remain compatible when variables are added, removed or renamed. The descriptors of message fields also have some useful properties that assist in reusing files:

- Extendable with new fields
- Fields can be renamed or removed without breaking binary compatibility
- Descriptors are available at run-time if necessary, and can be stored as protobuf messages themselves
- They support metadata on field definitions, which allows for example describing conversions from ROOT trees or how metadata should be merged

The auto-generated protobuf code in C++ is well designed and exposes the contents of messages in a way that encourages efficient code without using error-prone pointers for data access. However, protobuf is primarily an interprocess message format, and cannot directly be used in files without extra work, described in the following section.

B.3 A4 I/O: Offline Storage of Protocol Buffer Messages

Individual protobuf messages are not suitable as a standalone offline data format. To make sense of them, they require external knowledge (usually compiled into the binary in the case of C++), they do not describe their own length, nor do they have ending delimiters.

Since no container format for protobuf messages was available, the A4 message file and stream format was defined. It is illustrated in Figure B.2. The following bullet points outline the primary design considerations:

- Store protobuf messages of arbitrary types

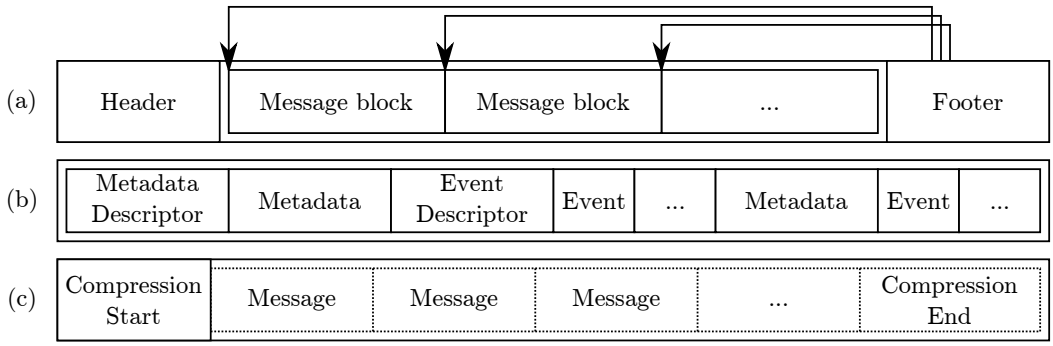


Figure B.2: The A4 data structure primitives. Each segment represents a protobuf message. An ellipsis is used to indicate repetition. (a) Overall logical file structure. The header contains the file type and version. It is followed by a number of message blocks. The footer contains information about the written messages, including byte offsets back to the header, to class descriptors and to metadata, making it possible to seek directly to the metadata. (b) An example message block, which may or may not be compressed (see (c)). The descriptors are only written once per protobuf class per output stream. (c) Example of a compressed series of messages. Message compression begins with an uncompressed message indicating the compression type. The compression is halted with an end-compression message. Compression is halted and immediately resumed when offsets are requested, e.g. to put descriptor offsets in the footer. Message compression is handled transparently in the reading library.

- Store the descriptors for messages, making the format self-describing and enabling the use of format-independent tools
- Store metadata for blocks of messages
- Transparent compression using different algorithms[2]
- Binary concatenation of A4 files yields a valid A4 file with all metadata quickly available, for trivial merging of large numbers of small files and efficient network transport
- Splitting of A4 files by a metadata field key (for example by data taking period) is possible with a command line tool
- Support a linear no-seeking mode of operation, suitable for network streaming

An experimental converter for ATLAS events to an "event message" was written in Python in the ATLAS Athena analysis framework[3] and run on ATLAS data using Ganga [52, 53] to submit jobs to the LHC computing grid. The resulting A4 files are stored on a dCache [101] system. These data were processed by distributing compiled executables using the A4 I/O and processing libraries via the local batch system. Both reduced or derived data-sets and complex sets of histograms have been produced, and enabled contributions to official ATLAS results.

Both C++ and Python interfaces are provided for the A4 format. The C++ interface can use the remote access libraries `rfio` and `dcap` to access grid storage elements. Experimental support for the Apache Hadoop distributed filesystem is also implemented. It also provides input and output classes that distribute files to any number of threads and combine output automatically.

Since the message structure is stored in the file, it does not need to be known at compile time if speed is not critical. Included in the A4 I/O module are the command-line tools `a4dump` and `a4info`, which print messages and metadata stored in A4 files in a human-readable format.

To summarise, the standalone A4 I/O module allows arbitrary protobuf messages to be written and read at high speed from multiple threads. A large amount of experimental data has been stored in A4 files, and was utilised to produce physics results. Building on this foundation, the A4 processing module described in the next section provides infrastructure elements to facilitate common tasks.

[2]Currently implemented are ZLIB, GZIP and SNAPPY.

[3]Python proved useful for rapidly prototyping a working converter at the expense of runtime speed. Faster conversion is now available with `root2a4` if the data is already in the form of ROOT trees.

```
import "a4/io/A4.proto";

message EventMetaData {
  optional bool simulation = 1 [(a4.io.merge)=MERGE_BLOCK_IF_DIFFERENT];
  repeated int32 mc_channel = 11 [(a4.io.merge)=MERGE_UNION];
  repeated string period = 3 [(a4.io.merge)=MERGE_UNION];

  optional int32 event_count = 6 [(a4.io.merge)=MERGE_ADD];
  optional double sum_mc_weights = 7 [(a4.io.merge)=MERGE_ADD];
  optional double reweight_lumi = 8 [(a4.io.merge)=MERGE_BLOCK_IF_DIFFERENT];
}
```

Figure B.3: An excerpt of the metadata descriptor used for the authors analysis. The A4 extensions to the protobuf field descriptors, `a4.io.merge`, describe how two metadata messages can be combined into one. The presence of `MERGE_BLOCK_IF_DIFFERENT` prevents histograms from data and simulation being summed. The `event_count` contains the sum of events which were processed in the original files, before any events were removed from the file during skimming.

B.4 Processing and Automatic Book-Keeping

In High Energy Physics, data are typically analysed in an event loop: each event is loaded in turn and processed by an analysis function. The A4 processing module attempts to make it as easy as possible to write powerful, configurable programs that analyse files containing events using multiple threads which run simultaneously. Metadata is also stored in a protobuf message class, which can be annotated to define how the metadata may be merged as illustrated in Figure B.3. The merging of the metadata itself is illustrated in Figure B.4. This allows automatic propagation of arbitrary information such as run numbers, Monte Carlo IDs and initial event counts through to the final histograms. This automatic metadata handling drastically simplifies the bookkeeping necessary to produce physics results from data.

The A4 processing module provides a `processor` base class. This class is available as a C++ template, which allows custom event and metadata protobuf message classes to be specified. The source code of an example processor is given in Figure B.5. This code compiles into an executable program. Command-line arguments can be used to set input and output files, disable A4 output entirely, set the number of threads, limit the number of events processed for testing purposes and control the metadata management. Additional program options can be added by the analysis writer. The popular C++ Boost library [102] is used to provide this feature, making it possible to provide `.ini` files to specify command line options.

Using this library, the common tasks of skimming (selecting only specific

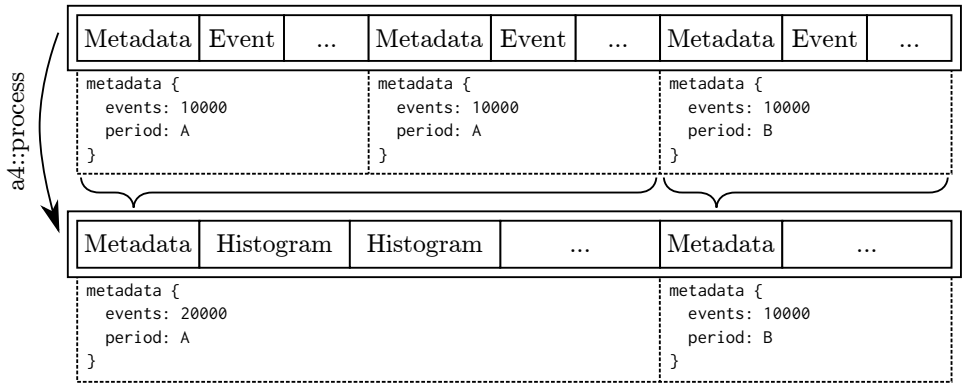


Figure B.4: A simplified example of metadata propagation. Analysis code using `a4::process` generates histograms or processed events from the input events using the `-per=period` command-line switch. The metadata is automatically combined according to the definitions on the metadata descriptor (see Figure B.3 for examples). If the metadata key had instead been `-per=simulation` (or another field which was uniform across files), the resulting histograms would contain entries for all events described by a single metadata: `{ events: 30000, period: A, period: B }`.

events to copy), slimming (selecting only specific physics objects to copy) and thinning (dropping unneeded variables from physics objects) can be performed simply by copying the event object or parts of it, modifying it, and calling `write`. In addition, new fields of the event can be filled, or even a different message type written out. One advantage of A4 in this approach is that the event definition does not have to be changed in the case of slimming or thinning, since in protobuf messages fields that are not set do not use any space[4]. In all cases, the metadata will be preserved and passed on to the output files. In addition, the metadata applicable to the current event is always available in the processor.

In short, the processor provides a quick way to define common tasks or more complex processing steps. Automatic handling of metadata reduces manual bookkeeping. Programs are compiled into executables which accept customisable command-line options, listed by `-h/--help`, and can additionally be stored in an `.ini` file.

Histograms are the primary output of High Energy Physics analyses. Since defining, filling and storing histograms makes up a large part of typical analysis codes by line count, simplifying this process is a worthwhile task. In the following section, the A4 store and A4 histogramming modules are presented, and their use in the processor architecture described.

[4]It is possible to check if fields have been set.

```
#include <a4/application.h>
#include "Event.pb.h"
#include "smear.h"
class SkimSlimThinSmear : public a4::process::ProcessorOf<Event, MetaData> {
  public:
    void process(const Event & event) {
      // Cut on at least two muons (skim)
      if (event.size_muons() < 2) return;
      // Remove some fields (slim)
      Event new_event = event; new_event.clear_tracks();
      foreach(auto & muon, *new_event.mutable_muons()) {
        // Smear muons before writing them
        if (metadata().simulation()) smear(muon);
        muon.clear_id_hits(); // Remove hits from muons (thin)
      }
      write(new_event); // Write modified event to output file
    }
};

int main(int argc, const char * argv[]) {
  return a4::process::a4_main_process<SkimSlimThinSmear>(argc, argv);
}
```

Figure B.5: Listing of an example program utilising the A4 processor class, skimming events with two muons, slimming these by removing tracks, removing the hits on the muons, and writing the events to an output file. The functions referring to the physics objects are generated automatically by `protoc`. The "smear" function is a hypothetical analysis function which modifies the contents of the written muon objects.

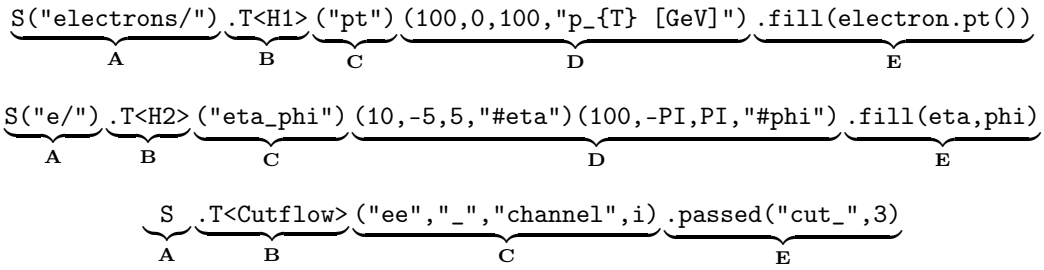


Figure B.6: Examples of the A4 store invocation rules. `S` is an object of the type `ObjectStore`, representing a directory or prefix. **A** represents the location where the histogram should be stored, and its return type is the same as the original type of `S` which is cheap to copy. This may be efficiently passed through to a function accepting an `ObjectStore` as a parameter, allowing reuse of histogram definitions. **B** requests a particular object type, in this case a one- or two-dimensional histogram or a histogram indexed by label. **C** names the object in the store, and may contain a variable number of string or integer arguments. **D** specifies the axis range, and if available, may specify variable binning with C++11's new initializer lists. The axis label specified in this example can be omitted if desired. **D** is repeated once for each dimension of the target histogram. **E** fills the histogram at the desired quantity or label. The resulting code has a performance close to that of the **E** call after the histogram has been encountered for the first time.

B.5 Reusing Histogram Definitions with the A4 Histogram Store

The A4 store uses features of the new C++11 standard to provide a way to define, initialize, fill and store histograms and similar objects on one line inside the event loop, greatly simplifying the necessary bookkeeping. For this, a *store* object, usually called `S` is defined, representing a directory or a prefix to the name of an object intended to be saved as a result. In Figure B.6, the use of the A4 store using the lightweight A4 histogram classes `H1`, `H2` and `Cutflow` is illustrated[5].

These one-line definitions can not only be used to define single histograms, but can be prefixed with different subdirectories and reused in different functions. It is possible to conditionally add a prefix to all of the following histograms in the current event using, for example, the statement `if (in_control_region) S = S("control_region/");`.

All histograms from this point will be created and filled in the subdirectory "control_region". It is also possible to pass a store object to a function in which common distributions are filled, enabling the reuse of common plots at

[5]A tool to convert A4 histograms to ROOT histograms is included.

different points in the analysis.

A common approach to simplify histogram management - used e.g. in the ATLAS Athena analysis framework - is to use a map keyed by the name of the histogram, making it necessary to do expensive string operations in the event loop. In A4, the store object uses a specially designed hash table instead, which uses the numerical value of the `const char *` pointer to the given strings instead of comparing characters. On first insertion of any pointer, it is checked to see if it points into a memory region designated read-only by the operating system and rejected otherwise. This protects the user of a library against errors if a slower dynamic string is inadvertently used. Common operations usually requiring dynamic strings - concatenation and numbering - can be done instead using the store itself, e.g. `S.T<H1>("hist_", "nr_", 4)` or `S("subdir_",3,"/")` are valid calls. This fast string lookup is also used for the `Cutflow` class to provide a histogram class where the bins are indexed by label: `S.T<Cutflow>("cf").passed("my_cut")` indicates that this event has passed "my_cut".

When using a processor object, the contents of the A4 store are written as protobuf messages to an A4 file specified on the command line. The results are written into blocks with the same metadata key as specified on the command line with the `--per` switch. This again reduces the bookkeeping effort, since results can be obtained per data taking run even if files contain multiple runs, or runs are spread over multiple files. The command-line tool `a4merge` is provided to merge stores (and the histograms in them). A file containing multiple keys can be split into multiple files with the `--split-per` parameter. In the final files, each set of histograms is associated with the metadata of all the events used to produce it. Histograms can then be re-weighted depending on metadata using `a4reweight`. This tool uses cross-section information as well as metadata to re-weight histogram entries to a desired luminosity.

B.6 Handling of Systematic Variations and Multiple Channels

Using the A4 store, the processor implements another key feature: handling of *channels* and *systematics*. The function call `channel("electron");` later causes a re-run of the whole processing function, with the same initial event, but with an additional store prefix of `channel/electron/`. Only during this analysis pass, the return value of the function is `true`. This is useful to obtain histograms created for example in object selection just for events that pass selection criteria further on, without having to copy and paste these histograms. This function can also be used to study the effect of alternative cuts on the analysis.

The `systematics("scale_up")` function has similar semantics, but does

not trigger a re-run by itself. If a re-run is scheduled with a command line option, this function returns true during that rerun, and the prefix `systematic/scale_up/` is added to the store. This enables conditional application of systematic uncertainty factors, and evaluation of these factors without recompilation.

To summarise, the A4 histogram store enables fast definition and filling of histograms on one line. Store objects as prefixes make it possible to reuse histogram definition and fill code in functions, and fill them e.g. after each cut. The processor functions `channel` and `systematic` simplify optimisation and the evaluation of systematic uncertainties. The histograms are stored in A4 files, which keep metadata information about the events the histograms were filled with. The A4 store can also be used in ROOT analysis using an adapter for ROOT histograms, with a limited set of features.

B.7 A4 ROOT

The A4 ROOT module contains programs to convert both A4 event data and A4 histograms to and from the popular ROOT file format. The command-line tool `a4results2root` converts histogram stores, whereas `a42root` auto-generates a ROOT tree structure to hold event data[6]. Conversion from ROOT trees to A4 messages in a flat format would also be automatically possible, but in general, flat ROOT trees usually used in analysis do not contain sufficient information to reconstruct the object structure. Adding such information to a special `.proto` file is necessary, see Figure B.7. For all ATLAS flat ROOT data formats (D3PDs), this file can be generated by the ATLAS Athena analysis framework with D3PDMakerA4 which is included in a4root. The `root2a4` program then converts any ROOT tree using the given structure.

B.8 Benchmarks

One of the primary design goals of A4 is high processing speed. To test this, a set of benchmarks using synthetic events is performed. For each event, all fields are filled with a random number generated using the GLIBC `random()` function. The events are "processed" by calculating the sum of all their fields. As a comparison, The same procedure is performed using ROOT version 5.32.

For all these benchmarks it must be considered that ROOT trees are also designed to quickly retrieve all instances of one or a few fields for each event, an advantage that is effectively disabled by the requirement to load all fields. In the present case, it is assumed that the relevant information was already selected and processed into a smaller file for local analysis.

[6]It is planned to merge these two tools in the near future

```
package a4.atlas.ntup.photon;
import "a4/root/ROOTExtension.proto";

message Photon {
    optional float E = 1;
    optional float px = 7; optional float py = 8; optional float pz = 9;
    extensions 100000 to max;
}

message Event {
    optional uint32 run_number = 1;
    optional uint32 event_number = 2 [(root_branch)="EventNumber"];
    repeated Photon photons = 100 [(root_prefix)="ph_"];
    extensions 100000 to max;
}
```

Figure B.7: An illustrative `.proto` file which can be used to convert an existing flat ROOT tree to an A4 event file. The A4 extensions to the protobuf field descriptors, `root_branch` and `root_prefix` are used to indicate the names of the leaves on the ROOT side. For example, this input ROOT file has a `std::vector<float>` branch called `ph_px[i]`, corresponding to `event.photons(i).px()` in the resulting protobuf class structure. The protobuf `extensions` keyword reserves numbers to be used in user extensions to the class, allowing the format to be extended at runtime.

The benchmark setup is similar to the flat ntuple format used in ATLAS analysis. The number of single, non-repeated fields of a certain type in the synthetic message is denoted by $n_{flat,float}$. The number of repeated fields - represented by `std::vector<T>` objects in ROOT - is $n_{rep,float}$. The number of entries in a repeated field is $n_{nfill,float}$. All benchmarks are performed on an unloaded system on a RAM disk. The reported runtimes are the minimum value measured over three runs, and the error bars shown are the standard deviation. For the first part of the benchmarks, no compression was used. The ROOT basket size was left at the default of 32k, but was varied in a separate run from zero to 1MB, resulting in speed differences of $< 3\%$. No other attempts to tune the performance of ROOT have been made, representing common non-expert usage.

There are two features of interest in the benchmark data: A speedup with respect to ROOT of a factor 3 up to a factor of 6 for events with > 100 flat or array-like fields of any type (Figure B.8) and a slowdown with respect to ROOT for large array-like fields to 0.9 (double) 0.6 (float), and 0.5 (integers) (Figure B.9). The observed behaviour is not yet fully understood, but indicates that gains in speed are possible, in situations where high bandwidth is available.

In the Appendix, the corresponding plots for all types are shown. One un-

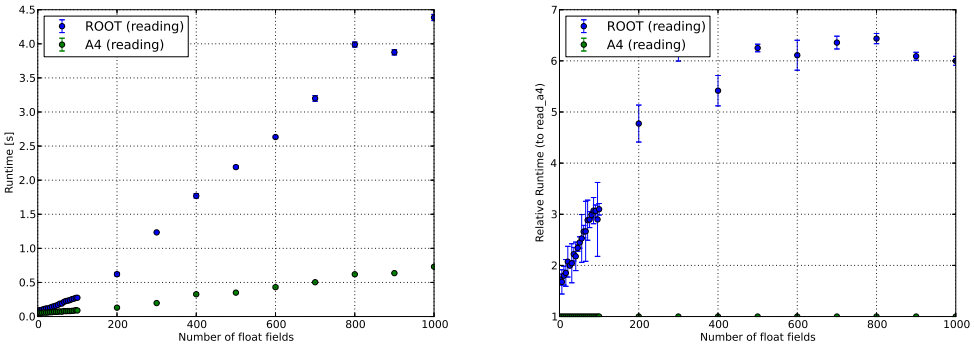


Figure B.8: Processing time in seconds for 100000 events versus $n_{flat,float}$. On the right-hand side, the time is normalized to the A4 runtime.

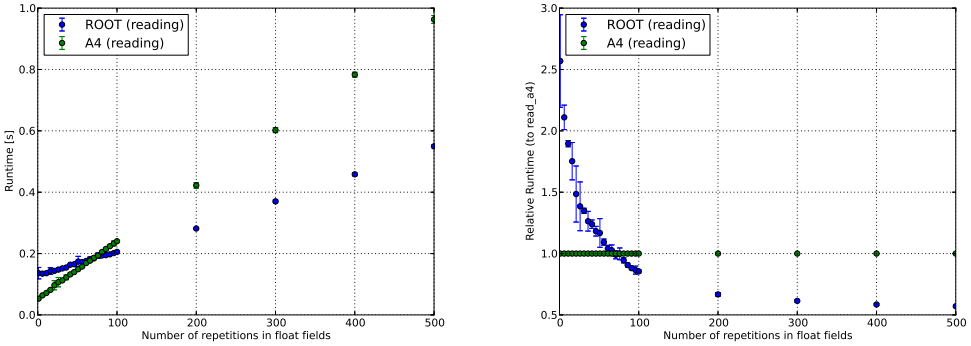


Figure B.9: Processing time in seconds for 100000 events with $n_{rep,float} = 4$ versus $n_{nfill,float}$. On the right-hand side, the time is normalized to the A4 runtime.

expected result is the relatively inefficient handling of integers, where further investigation into protobuf performance is necessary. In addition, plots with ZLIB level 1 compression enabled are shown. Since the same compression is used in ROOT and A4, the difference in speed decreases. The maximum speedup of A4 with respect to ROOT is now ≈ 1.8, except for boolean variables where it remains at ≈ 4.

To check the behavior under more realistic conditions, a typical ATLAS ntuple file based on ROOT trees was converted to an A4 file[7], and a minimal event loop for both ROOT and A4 was written. The speedup of A4 in this case was 2.6. After this, the branches of the ROOT tree were manually disabled, until the same processing speed was reached. At this point, 40% of

[7]The A4 file in this case did not have the structure described in the benchmark above, but had more complex structured definitions for physics objects.

the branches were enabled, indicating approximately linear scaling of runtime with the number of active branches in ROOT.

It can be concluded that in situations where a majority of data in an event are used in an analysis, A4 can provide significant speedups. Even in other cases, the simple slimming and thinning can quickly lead to a situation where again the majority of data in an event is required. Large arrays are not yet handled efficiently, and need further attention.

B.9 Summary and Outlook

This section presented an overview of the A4 library, a toolkit for data analysis in High Energy Physics. A fast I/O format is described, and it is demonstrate that it is able to perform comparably to ROOT trees for flat event analysis. The bookkeeping requirements at any step in data analysis are minimised by automated treatment of metadata for events and histograms. Creation of histograms and common procedures such as evaluation of systematic uncertainties are simplified by the A4 store in combination with the A4 processor class. Interoperability with ROOT is achieved by a set of conversion tools.

Many features of the architecture[8] have not yet been fully exploited. Consolidation and documentation of the existing codebase and extending the test suite is currently a focus of development. Finally, no attempt has yet been made to modify the protobuf library message parsing code for our specific problem, an approach that might improve performance or usability in some cases even further, and could then be submitted for inclusion in the protobuf library.

[8] e.g. chaining processors, threading single files

Bibliography

[1] J. Hittorf, Ueber die Elektricitätsleitung der Gase, *Annalen der Physik* **212**(2), 197–234 (1869).

[2] K. Nakamura et al. (Particle Data Group), 2010 Review of Particle Physics, *J. Phys. G* **075021**(37) (2010).

[3] S. L. Glashow, Partial Symmetries of Weak Interactions, *Nucl. Phys.* **22**, 579–588 (1961).

[4] A. Salam, Weak and Electromagnetic Interactions, The Nobel Symposium held at Lerum, Sweden, 1968.

[5] S. Weinberg, A Model of Leptons, *Phys. Rev. Lett.* **19**, 1264–1266 (Nov 1967).

[6] D. Gross and F. Wilczek, Ultraviolet Behavior of Non-Abelian Gauge Theories, *Phys. Rev. Lett.* **30**(26), 1343–1346 (1973).

[7] H. Politzer, Reliable Perturbative Results for Strong Interactions?, *Phys. Rev. Lett.* **30**(26), 1346–1349 (1973).

[8] A. Einstein, Zur allgemeinen Relativitätstheorie, *Ständiger Beobachter der Preussischen Akademie der Wissenschaften* , 778–786, 799–801 (1915).

[9] CDF Collaboration, Precise Measurement of the W-Boson Mass with the CDF II Detector, *Phys. Rev. Lett.* **108**, 151803 (Apr 2012).

[10] F. Abe et al. (CDF Collaboration), Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab, *Phys. Rev. Lett.* **74**(14), 2626–2631 (Apr 1995).

[11] ATLAS Collaboration, Combined Search for the Standard Model Higgs Boson using up to 4.9 fb^{-1} of pp Collision Data with the ATLAS Detector at the LHC, *Phys. Lett. B* **710**(1), 49 – 66 (2012).

[12] CMS Collaboration, Combined Results of Searches for the Standard Model Higgs Boson in pp Collisions at $\sqrt{s} = 7$ TeV, *Phys. Lett. B* **710**(1), 26 – 48 (2012).

[13] Q. R. Ahmad et al. (SNO Collaboration), Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **87**(7) (Jul 2001).

[14] M. Hobson, *General Relativity : An Introduction for Physicists*, Cambridge University Press, Cambridge, UK New York, 2006.

[15] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory*, Westview Press, Boulder, U.S.A., 1995.

[16] E. Noether, Invariante Variationsprobleme, *Nachrichten der Königlichen Gesellschaft der Wissenschaften* , 235–257 (1918).

[17] R. P. Feynman, Space-Time Approach to Non-Relativistic Quantum Mechanics, *Rev. Mod. Phys.* **20**, 367–387 (Apr 1948).

[18] S. Dürr, Z. Fodor, J. Frison, C. Hoelbling, R. Hoffmann, S. Katz, S. Krieg, T. Kurth, L. Lellouch, T. Lippert, et al., Ab Initio Determination of Light Hadron Masses, *Science* **322**(5905), 1224–1227 (2008).

[19] B. R. Stella and H.-J. Meyer, Υ(9.46 GeV) and the gluon discovery (a critical recollection of PLUTO results), *Eur. Phys. J. C* **H36**, 203–243 (2011), 1008.1869.

[20] T. Hambye and K. Riesselmann, SM Higgs Mass Bounds from Theory, *Preprint* (1997), hep-ph/9708416.

[21] ATLAS Luminosity Group, Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector in 2011, Technical Report ATLAS-COM-CONF-2011-130, CERN, Geneva, Jul 2011.

[22] M. Dobbs, S. Frixione, E. Laenen, K. Tollefson, H. Baer, E. Boos, B. Cox, R. Engel, W. Giele, J. Huston, et al., Les Houches Guidebook to Monte Carlo Generators for Hadron Collider Physics, *Preprint* (2004), hep-ph/0403045.

[23] A. Sherstnev and R. Thorne, Different PDF Approximations useful for LO Monte Carlo Generators, *Preprint* (2008), 0807.2132.

[24] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, MadGraph 5 : Going Beyond, *JHEP* **1106**, 128 (2011), 1106.0522.

[25] J. Campbell, R. Ellis, and C. Williams, Vector Boson Pair Production at the LHC, *JHEP* **2011**, 1–36 (2011).

[26] K. J. F. Gaemers and G. J. Gounaris, Polarization Amplitudes for $e^+e^- \to W^+W^-$ and $e^+e^- \to ZZ$, *Z. Phys. C* **1**, 259–268 (1979).

[27] J. Butterworth, E. Dobson, U. Klein, B. Mellado Garcia, T. Nunnemann, J. Qian, D. Rebuzzi, and R. Tanaka, Single Boson and Diboson Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV, Technical Report ATL-COM-PHYS-2010-695, CERN, Geneva, Aug 2010.

[28] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. Seymour, and B. Webber, HERWIG 6: An Event Generator for Hadron Emission Reactions with Interfering Gluons (including Supersymmetric Processes), *J. High Energy Phys.* **2001**, 010 (2001).

[29] T. Sjöstrand, S. Mrenna, and P. Skands, PYTHIA 6.4 Physics and Manual, *JHEP* **2006**(05), 026 (2006).

[30] J. Ebke and P. Waller, MCViz Home Page, `http://mcviz.net`.

[31] European Organization for Nuclear Research (Geneva), CERN Home Page, `http://www.cern.ch`.

[32] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, *J. Instrum.* **3**, S08003 (2008).

[33] European Organization for Nuclear Research (Geneva), ATLAS Home Page, `http://atlas.ch`.

[34] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*, CERN, Geneva, 2004.

[35] ATLAS Collaboration, *ATLAS Magnet System: Technical Design Report*, CERN, Geneva, 1997.

[36] G. Aad, M. Ackers, F. Alberti, M. Aleppo, G. Alimonti, J. Alonso, E. Anderssen, A. Andreani, A. Andreazza, J. Arguin, et al., ATLAS Pixel Detector Electronics and Sensors, *J. Instrum.* **3**(07), P07007 (2008).

[37] J. R. Pater for the ATLAS Collaboration, The ATLAS SemiConductor Tracker Operation and Performance, *J. Instrum.* **7**(04), C04001 (2012).

[38] The ATLAS TRT collaboration, The ATLAS Transition Radiation Tracker (TRT) Proportional Drift Tube: Design and Performance, *J. Instrum.* **3**(02), P02013 (2008).

[39] H. Wilkens for the ATLAS LArg Collaboration, The ATLAS Liquid Argon Calorimeter: An Overview, *J. Phys. Conf. Ser.* **160**(1), 012043 (2009).

[40] ATLAS Collaboration, Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions, *Eur. Phys. J. C* **70**, 723–753 (2010).

[41] ATLAS Collaboration, Readiness of the ATLAS Tile Calorimeter for LHC collisions, *Eur. Phys. J. C* **70**, 1193–1236 (2010).

[42] ATLAS Collaboration, ATLAS Muon Spectrometer Technical Design Report, *CERN/LHCC* **97**, 22 (1997).

[43] ATLAS Collaboration, Muon Momentum Resolution in First Pass Reconstruction of *pp* Collision Data Recorded by ATLAS in 2010, Technical Report ATLAS-CONF-2011-046, CERN, Geneva, Mar 2011.

[44] ATLAS Public Documentation, `https://twiki.cern.ch/twiki/pub/AtlasPublic/MuonPerformancePublicPlots/ATL-COM-PHYS-2011-1504.pdf`.

[45] G. Chiodini, N. Orlando, and S. Spagnolo, ATLAS RPC Time-of-Flight Performance, Technical Report ATL-MUON-PROC-2012-002, CERN, Geneva, Apr 2012.

[46] V. Cindro, D. Dobos, I. Dolenc, H. Frais-Kölbl, A. Gorišek, E. Griesmayer, H. Kagan, G. Kramberger, B. Macek, I. Mandić, M. Mikuž, M. Niegl, H. Pernegger, D. Tardif, W. Trischuk, P. Weilhammer, and M. Zavrtanik, The ATLAS Beam Conditions Monitor, *J. Instrum.* **3**(02), P02004 (2008).

[47] S. Jakobsen, Calibration, Testing, Commissioning and First Data of ALFA at LHC, *Nucl. Instrum. Meth. A* (2011).

[48] S. White, The ATLAS Zero Degree Calorimeter, *Nucl. Instrum. Meth. A* **617**(1), 126–128 (2010).

[49] G. Barrand, I. Belyaev, P. Binko, M. Cattaneo, R. Chytracek, G. Corti, M. Frank, G. Gracia, J. Harvey, E. Herwijnen, et al., GAUDI - A Software Architecture and Framework for building HEP Data Processing Applications, *Comput. Phys. Commun.* **140**(1), 45–55 (2001).

[50] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, et al., GEANT4 – A Simulation Toolkit, *Nucl. Instrum. Meth.* **506**(3), 250–303 (July 2003).

[51] R. Brun and F. Rademakers, ROOT - An Object Oriented Data Analysis Framework, *Nucl. Instrum. Meth. A* **389**(1), 81–86 (1997).

[52] J. Mościcki, F. Brochu, J. Ebke, U. Egede, J. Elmsheuser, K. Harrison, R. Jones, H. Lee, D. Liko, A. Maier, et al., Ganga: A Tool for Computational-Task Management and Easy Access to Grid Resources, *Comput. Phys. Commun.* **180**(11), 2303–2316 (2009).

[53] J. Elmsheuser, F. Brochu, I. Dzhunov, J. Ebke, U. Egede, M. Jha, L. Kokoszkiewicz, H. Lee, A. Maier, J. Mościcki, et al., Reinforcing User Data Analysis with Ganga in the LHC Era: Scalability, Monitoring and User-Support, *J. Phys. Conf. Ser.* **331**, 072011 (2011).

[54] ATLAS Collaboration, The ATLAS Simulation Infrastructure, *Eur. Phys. J. C* **70**, 823–874 (Dec. 2010), 1005.4568.

[55] S. Frixione and B. R. Webber, Matching NLO QCD Computations and Parton Shower Simulations, *JHEP* **0206**, 029 (2002).

[56] T. Binoth, M. Ciccolini, N. Kauer, and M. Kramer, Gluon-Induced W-Boson Pair Production at the LHC, *JHEP* **0612**, 046 (2006), hep-ph/0611170.

[57] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, ALPGEN, A Generator for Hard Multiparton Processes in Hadronic Collisions, *JHEP* **0307**, 001 (2003), hep-ph/0206293.

[58] ATLAS Internal Documentation, Heavy Flavor Overlap Removal Tool, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HforTool?rev=10`.

[59] B. P. Kersevan and E. Richter-Was, The Monte Carlo Event Generator AcerMC Version 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5, *Preprint* (2004), hep-ph/0405247.

[60] M. Smizanska, PythiaB: An Interface to Pythia6 dedicated to Simulation of Beauty Events, Technical Report ATL-COM-PHYS-2003-038, CERN, Geneva, 2003.

[61] R. C. Gray, C. Kilic, M. Park, S. Somalwar, and S. Thomas, Backgrounds To Higgs Boson Searches from $W\gamma^* \to \ell\nu\ell(\ell)$ Asymmetric Internal Conversion, *Preprint* (2011), 1110.1368.

[62] ATLAS Internal Documentation, The Common Parameters used for ATLAS MC Production 2011, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/McProductionCommonParameters`.

[63] ATLAS Collaboration, New ATLAS Event Generator Tunes to 2010 Data, Technical Report ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr 2011.

[64] ATLAS Collaboration, ATLAS Tunes of PYTHIA 6 and Pythia 8 for MC11, Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.

[65] H. Lai, M. Guzzi, J. Huston, Z. Li, P. Nadolsky, J. Pumplin, and C. Yuan, New Parton Distributions for Collider Physics, *Phys. Rev. D* **D82**, 074024 (2010), 1007.2241.

[66] ATLAS Internal Documentation, `https://atlas-datasummary.cern.ch/lumicalc/`.

[67] ATLAS Public Documentation, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`.

[68] ATLAS Collaboration, Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector in 2011, Technical Report ATLAS-CONF-2011-116, CERN, Geneva, Aug 2011.

[69] ATLAS Internal Documentation, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets2011`.

[70] ATLAS Internal Documentation, `https://savannah.cern.ch/task/?14970`.

[71] J. Morel, Status of the Atlas Liquid Argon Calorimeter and its Performance after One Year of LHC Operation, Technical Report ATL-COM-LARG-2011-041, CERN, Geneva, Dec 2011.

[72] ATLAS Public Documentation, `https://twiki.cern.ch/twiki/pub/Atlas/MuonTriggerPhysicsTriggerRecommendations2011/trigger.pdf`.

[73] ATLAS Internal Documentation, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/EfficiencyMeasurements?rev=71`.

[74] ATLAS Collaboration, Electron Performance Measurements with the ATLAS Detector using the 2010 LHC Proton-Proton Collision Data, *Eur. Phys. J. C* **72**, 1–46 (2012).

[75] ATLAS Internal Documentation, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/IsEMIdentification?rev=24`.

[76] ATLAS Internal Documentation, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/WZElectroweakCommonTopics2011?rev=108`.

[77] T. Dai, J. Ebke, J. Elmsheuser, J. Gao, D. Hall, C. Hays, R. Kehoe, O. Kepka, S. Li, J. Liu, L. Liu, Y. Liu, E. Monnier, H. Peng, M.-A. Pleier, M. Schott, Y. Wu, H. Yang, Z. Zhao, B. Zhou, J. Zhu, and Y. Zhu, Measurement of the W^+W^- Production Cross Section in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector, Technical Report ATL-COM-PHYS-2012-145, CERN, Geneva, Feb 2012.

[78] B. Resende, Muon Identification Algorithms in ATLAS, Technical Report ATL-PHYS-PROC-2009-113, CERN, Geneva, Sep 2009.

[79] ATLAS Internal Documentation, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MCPAnalysisGuidelinesRel17MC11a`.

[80] M. Cacciari, G. P. Salam, and G. Soyez, The Anti-k_T Jet Clustering Algorithm, *JHEP* **2008**(04), 063 (2008).

[81] ATLAS Collaboration, Validating the Measurement of Jet Energies with the ATLAS Detector using Z+Jet Events from Proton-Proton Collisions at $\sqrt{s} = 7$ TeV, Technical Report ATLAS-CONF-2011-159, CERN, Geneva, Dec 2011.

[82] ATLAS Collaboration, Measurement of the b-Tag Efficiency in a Sample of Jets Containing Muons with 5 fb^{-1} of Data from the ATLAS Detector, Technical Report ATLAS-CONF-2012-043, CERN, Geneva, Mar 2012.

[83] ATLAS Collaboration, Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS, *Eur. Phys. J. C* **C72**, 1844 (2012), 1108.5602.

[84] N. E. Adam, V. Halyo, and S. A. Yost, Evaluation of the Theoretical Uncertainties in the $Z \to \ell^+\ell^-$ Cross Sections at the LHC, *JHEP* **0805**, 062 (2008), 0802.3251.

[85] ATLAS Collaboration, Measurement of the W^+W^- Production Cross Section in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector, Technical Report ATLAS-CONF-2012-025, CERN, Geneva, Mar 2012.

[86] R. Aben et al., Background Estimates in the $H \to W^+W^- \to \ell\nu\ell\nu$ Analysis with 4.7 fb^{-1} of Data collected with the ATLAS Detector at $\sqrt{s} = 7$ TeV, Technical Report ATL-COM-PHYS-2011-1728, CERN, Geneva, Dec 2011.

[87] ATLAS Collaboration, Muon Reconstruction Efficiency in Reprocessed 2010 LHC Proton-Proton Collision Data recorded with the ATLAS Detector, Technical Report ATLAS-CONF-2011-063, CERN, Geneva, Apr 2011.

[88] ATLAS Internal Documentation, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ElectronGamma`.

[89] ATLAS Internal Documentation, `https://twiki.cern.ch/twiki/bin/viewauth/Atlas/MuonTriggerPhysicsTriggerRecommendations2011`.

[90] A. Martin, W. Stirling, R. Thorne, and G. Watt, Parton Distributions for the LHC, *Eur. Phys. J. C* **C63**, 189–285 (2009), 0901.0002.

[91] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, LEP Electroweak Working Group, A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model, *Preprint* (2005), hep-ex/0511027.

[92] D0 Collaboration, Measurement of the W^+W^- Production Cross Section with Dilepton Final States in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV and Limits on Anomalous Trilinear Gauge Couplings, *Phys. Rev. Lett.* **103**, 191801 (2009), 0904.0673.

[93] CDF Collaboration, Measurement of the W^+W^- Production Cross Section and Search for Anomalous $WW\gamma$ and WWZ Couplings in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV, *Phys. Rev. Lett.* **104**, 201801 (2010), 0912.4500.

[94] ATLAS Collaboration, Measurement of the W^+W^- Cross Section in $\sqrt{s} = 7$ TeV pp Collisions with ATLAS, *Phys. Rev. Lett.* **107**, 041802 (Jul 2011).

[95] J. Ebke for the ATLAS Collaboration, Electroweak Di-boson Production in ATLAS, *Preprint* (2012), 1201.0665, To be published in Proceedings of the PLHC2011 Conference, Perugia, Italy, June 6-11, 2011.

[96] ATLAS Collaboration, Measurement of the W^+W^- cross section in $\sqrt{s} = 7$ TeV pp Collisions with the ATLAS Detector and Limits on Anomalous Gauge Couplings, *Phys. Lett. B* (2012), 1203.6232.

[97] CMS Collaboration, Measurement of WW Production Rate, Technical Report CMS-PAS-SMP-12-005, CERN, Geneva, Apr 2012.

[98] J. M. Butterworth, B. E. Cox, and J. R. Forshaw, WW Scattering at the CERN LHC, *Phys. Rev. D* **65**(9), 096014 (2002).

[99] S. Khalatyan, Non-ROOT I/O, https://indico.cern.ch/contributionDisplay.py?confId=141309&contribId=29.

[100] Protocol Buffers, GOOGLEs Data Interchange Format, http://code.google.com/p/protobuf.

[101] A. Agarwal, R. Enge, K. Fransham, E. Kolb, C. Leavett-Brown, D. Leske, K. Lewall, H. Reitsma, E. Rempel, and R. Sobie, dCache with Tape Storage for High Energy Physics Applications, *J. Phys. Conf. Ser.* **219**, 072024 (2010).

[102] BOOST C++ Libraries, http://www.boost.org/.

Acknowledgements

At this point, I would like to give thanks to all those who have contributed to the successful completion of this thesis.

My grateful thanks go to my advisor Prof. Dr. Dorothee Schaile for giving me the opportunity to write this thesis, her thematic suggestions and her continuous support over the years. Her generous support for summer schools, for a nine-month long stay at CERN, for collaboration meetings and conferences - even outside my main field of study - has been extraordinary.

I also would like to thank Prof. Dr. Wolfgang Dünnweber for agreeing to write the second review of my thesis, and Prof. Dr. Gerhard Buchalla and Prof. Dr. Jochen Schieck for participating in the thesis defense.

A lot of thanks go to Dr. Johannes Elmsheuser for supervising this thesis, for many discussions on computing and physics, his helpful suggestions and his good advice.

I thank Peter Waller for the amazing and fruitful collaboration in physics and software over the last two years, for the great time at CERN, and not least for the many things he taught me about software and computing.

Many thanks also to Prof. Dr. Otmar Biebel for the refreshing discussions over coffee, his always open door for questions, and not least for his support as the head of the postgraduate program "particle physics at the energy frontier of new phenomena" that graciously supported me for three years.

During my nine-month stay at CERN, my colleagues and friends Jonas Will und Johannes Erdmann shared a flat with me. I am very grateful for their companionship, patience, and also their willingness to share excellent food and drink. At CERN, many people helped me find a home away from home, and made my stay a wonderful experience. I specifically want to thank Peter Loscutoff, Lindsey Gray, Andre Bach, Rob and Helen Lambert, Pierpaolo Valerio, Hugo Day, Luke Thompson, and many others from the CERN games club - where the rules of physics do not apply. From the ATLAS run coordination and management team, I especially want to thank Stephanie Zimmerman for giving me the chance to become ATLAS shift leader, and Daniel Dobos and Thilo Pauly for their help in difficult situations.

For support, patience and advice I also thank the my analysis group, especially the former and current convenors Joao Guimaraes da Costa, Marc-Andre Pleier and Matthias Schott. For many compared cutflows and great work in the analysis group I also give my thanks to Shu Li.

From my officemates over the years I want to give special thanks to Philippe Calfayan for sharing his office at CERN, with a wonderful view of the Jura. Thanks also to Benjamin Ruckert, Christoph Bussenius, Steven Beale, Florian Rättich and especially Cristina Galea for many interesting conversations and for a consistently great office atmosphere.

I want to thank Günter Duckeck for a lot of help, discussions and experiments on software, the grid and physics, and Rod Walker, John Kennedy and Christoph Mitterer for the wonderfully working combination of the German LCG cloud and the LRZ dCache system and analysis cluster.

Thanks also go to Michiel Sanders for connecting me with the "Netzwerk Teilchenwelt", giving me the chance to convey the excitement of particle physics reseach to local grammar school students.

At the institute, I was fortunate to have a group of wonderful colleagues too numerous to name. Special thanks go to Stefanie Adomeit, Jona Bortfeld, Bonnie Chow, Marie-Helene Genest, Julien de Graat, Ralf Hertenberger, Federica Legger, Christian Meineck, Thomas Müller, Christopher Schmitt and Josephine Wittkowski.

All administrative problems were superbly handled by our formidable group secretary Herta Franz, and her successor in office Elke Grimm-Zeidler - many thanks.

Finally, I want to thank my family, first of all my parents, for their unwavering support in all my studies, their guidance and advice at all times. My greatest thanks go to my wife Amelie for her constant encouragement, her patience with me spending long nights estimating backgrounds on short deadlines, and for her wonderful support during my long stay at CERN.