

**Inclusive and differential cross-section
measurements of tW single top-quark production
at $\sqrt{s} = 13$ TeV with the ATLAS detector**

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Rui Zhang
aus
Shanxi, China



Bonn, 25.02.2019

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

1. Gutachter:	Prof. Dr. Ian C. Brock
2. Gutachterin:	Prof. Dr. Norbert Wermes
Tag der Promotion:	07.06.2019
Erscheinungsjahr:	2019

Contents

1	Introduction	1
2	Theoretical and experimental basics	5
2.1	Introduction to the quantum field theory	5
2.1.1	The Lagrangian density	5
2.1.2	(Local) Gauge invariance	5
2.1.3	Interaction of fields	6
2.2	The Standard Model of particle physics	7
2.2.1	Ingredients of the SM	7
2.2.2	Electroweak unification	9
2.2.3	Higgs field and its boson	10
2.2.4	Quarks in electroweak theory	11
2.2.5	Quarks in the strong interaction	12
2.2.6	Resonances	13
2.2.7	Flaws and limitations of the SM	14
2.3	Collider physics	14
2.3.1	Basic concepts	14
2.3.2	Hadron collider	16
2.3.3	Detectable particles	18
2.4	Top-quark physics	18
2.4.1	Discovery	18
2.4.2	Top-quark properties	18
2.4.3	The role of the top quark in the EWSB	18
2.4.4	Top-quark production at hadron colliders	19
2.4.5	Decay of the top quark	23
2.5	Interference between tW and $t\bar{t}$ productions	24
3	Particle detection and reconstruction	27
3.1	The Large Hadron Collider	27
3.1.1	Accelerator complex	27
3.1.2	Main ingredients	27
3.1.3	Experiments at the LHC	28
3.1.4	Operation energies and integrated luminosities	29
3.2	The ATLAS experiment	29
3.2.1	The inner detector	31
3.2.2	The calorimeter	33

3.2.3	The muon spectrometer	34
3.2.4	The magnet system	37
3.2.5	The trigger system	38
3.3	Object reconstruction	38
3.3.1	Charged tracks and vertices	38
3.3.2	Calorimeter clusters	40
3.3.3	Electrons	41
3.3.4	Muons	42
3.3.5	Jets	44
3.3.6	Tagged jets	47
3.3.7	Missing transverse momentum	48
4	Analysis setup	51
4.1	Collision data	51
4.2	Monte Carlo (MC) simulation	52
4.2.1	Detector simulation	52
4.2.2	Signal tW channel	52
4.2.3	Major background	53
4.2.4	Other (small) backgrounds	54
4.3	Event selection	57
4.3.1	Preselection	57
4.3.2	Jet and tag selection	57
4.3.3	Overlap removal	58
4.3.4	Cut flow	58
4.4	Fake-lepton estimation	59
4.5	Selection performance	59
4.5.1	Analysis regions and event yields	59
4.5.2	Control plots	60
5	Multivariate analysis	65
5.1	Boosted decision trees technique	65
5.1.1	Decision tree	65
5.1.2	Training a decision tree	65
5.1.3	Boosting	66
5.1.4	Bagging	68
5.1.5	Figure of merit	68
5.1.6	Overtraining and undertraining	68
5.2	Application of BDT	69
5.2.1	Variables selection	69
5.2.2	Hyperparameters optimisation	70
5.2.3	BDT performance	70
6	Uncertainty estimation and statistical analysis	79
6.1	General concepts	79

6.2	Maximum likelihood method	79
6.2.1	Multi-region and multi-bin likelihood function	80
6.2.2	Incorporating systematic uncertainties	81
6.2.3	Hypothesis testing	82
6.2.4	Validation of the statistical model	84
6.3	Systematic uncertainty estimation	84
6.3.1	Theory uncertainties	84
6.3.2	Detector and algorithm performances uncertainties	85
6.3.3	Cross-section normalisation on small backgrounds	87
6.4	Application of maximum likelihood fit	87
6.4.1	Simultaneous fit	87
6.4.2	Validations and discussions	90
6.4.3	Results	95
7	Differential cross-section measurements	97
7.1	Motivation	97
7.2	Unfolding	97
7.2.1	Deconvolution problem	97
7.2.2	Bin-by-bin corrections	98
7.2.3	Matrix inversion	98
7.2.4	Regularised unfolding	99
7.2.5	Iterative unfolding	99
7.2.6	RooUnfold package	100
7.3	Cross-section determination	101
7.3.1	Particle-level objects and fiducial region	101
7.3.2	Choice of variables and binning	101
7.3.3	Signal-enriched region	102
7.3.4	Number of iterations	102
7.3.5	Unfolding and cross-section calculation	104
7.3.6	Validation of unfolding methods	104
7.3.7	Uncertainty estimation	106
7.4	Results	107
8	Adversarial neural network training	113
8.1	Introduction	113
8.1.1	Information theory	113
8.1.2	Feed-forward networks	114
8.1.3	Model capacity	116
8.1.4	Optimisation	117
8.1.5	Back-propagation	117
8.2	Invariant training on nuisance parameters	117
8.3	Implementation	119
8.3.1	Setup and training	119
8.3.2	ANN response	121
8.3.3	Fit results and discussions	121

9	Conclusions	127
A	Generator information of samples used in the fake-lepton events estimates	131
B	Rest BDT input variables	133
C	Pull and impact of the NPs (BDT)	139
D	Detailed uncertainties on the differential cross-sections	141
E	Pull and impact of the NPs (ANN)	149
	Bibliography	151
	Acronyms	167
	Acknowledgements	173

Introduction

Humanity's interest in the formation of the universe and the constituents of matter has been ubiquitous and enduring. Since the discoveries of cathode rays in 1838 [1] and the electron in 1897 [2], human beings have stepped into the subatomic scale epoch and quantum era. As a consequence, two major questions emerge: what are the fundamental building blocks of matter and what are the forces to hold them together. Until the early 1930s, the known fundamental particles were the electron (1897), the proton (1919) [3], and the neutron (1932) [4]—all necessary building pieces of an atom. By then, quantum mechanics and general relativity had also been well established. However, solving problems are sometimes accompanied by observing novel phenomena that are hard to be understood, which drives humans to explore the unknown, push the boundaries of the scientific and technical limits, and then push further. In 1928, Paul Dirac realised that the positively-charged electrons, named positrons, were required by his equation to unify quantum mechanics and special relativity [5]. And about four and half years later, the positron, as the first antimatter, was experimentally discovered by Carl Anderson [6]. Loss of energy and momentum in radioactive beta decays provoked the hypothesis of neutrino by Wolfgang Pauli in 1931 [7]. After almost three decades, Clyde Cowan and Fred Reines found a particle that fit the description of the proposed neutrino when studying the particles created by a nuclear power plant [8]. This particle turned out to be the electron neutrino. During the period of time, people also gained plenty of skills on how to accelerate particles. Indeed, accelerators are the driving tool of particle physics research, and acquisition of particle physics knowledge has in turn provided more possibilities for the development of accelerators.

From late 1940s to late 1960s, hundreds of strongly interacting particles were found, and believed to be distinct elementary particles in their own right. The confusion was gradually cleared up when people discarded the idea that not all the hundreds of particles were fundamental. There may be substructures of most if not all these observed particles such that fewer elements could form those particles. In 1964 the quark model [9–11] was proposed to classify those strongly interacting particles, hadrons as we know now. The tight structure organises the large number of lighter hadrons (composed of three quarks), resulting in a further simplification. By that time people had also made great advances in the relativistic quantum electrodynamics (QED) [12–14] describing electrically charged particles interacting with matter. It was not only successful and accurately predictive, but also inspiring so that the same basic concepts were applied to the other forces of nature. In 1960s, the weak interaction, first observed in beta decay, was successfully unified with QED, forming electroweak interaction [15–17]. Thus two of the four known fundamental interactions of nature: electromagnetism and the weak

interaction, do share the same basic concepts and mathematical formulation. This is the first step towards the Standard Model (SM) of particle physics [18], which is currently the most successful theory of fundamental particles and how they interact.

At the same time, not only were many new types of accelerators discussed and built, but also the energy and intensity of accelerators were pushed higher and higher [19, 20]. By the 1940s, three types of accelerators using different mechanisms had been demonstrated: direct current acceleration [21], resonant acceleration [22] and the betatron mechanism [23]. All types were limited in a few hundred MeV barrier by various reasons. The breakthrough happened in 1945, when the principle of phase stability were discovered [24, 25], resulting in the invention of synchrotron. Later came a number of technical inventions, such as strong focusing [26], storage ring collision, stochastic cooling [27] and so on, allowing dramatic moves of the energy frontier. The synchrotron overshadowed other types of accelerators in the race for high energy. Even until today, the record holder of high energy—the Large Hadron Collider (LHC) [28], which generates the data used in this thesis—is a synchrotron-type accelerator.

From 1970 till the end of the 20th century, many breakthroughs were made to complete the elementary particles list and to refine the SM. The J/ψ meson was first identified in November 1974 [29, 30], with the profound, rapid changes in particle physics causing the “November Revolution”. The discovery demonstrated the existence of the 4th quark. The 3rd generation of charged leptons, τ , was discovered in 1975 [31], which, together with electron and muon (1937) leptons [32], form three generations of charged leptons. In 1977 the Υ meson was observed [33], indicating the existence of the 5th quark. The particles that carry the forces—the gluon (1979) [34] and W and Z bosons (1983) [35–38]—were also observed later on. Finally the top-quark, the 6th and also the last quark predicted by SM, was discovered in 1995 [39, 40]. In the final year of the last century, the tau neutrino, counterpart of τ , was observed [41] and completed the three generations of neutrinos as counterparts of three charged leptons. So far, the six types of quarks, three generations of leptons, and three out of four force carriers have been discovered and identified. Mankind, with all these great achievements, entered in the 21st century.

The LHC was built from 1998 to 2008 [42], aiming to answer some of the fundamental open questions in physics concerning the four fundamental forces and all elementary particles. One of the unsolved mysteries was the origin of particle mass. The SM kept W and Z bosons massless to be self-consistent, while measurements showed that these bosons actually have relatively large masses. The Brout-Englert-Higgs mechanism was introduced to solve the conundrum in 1964 by Englert and Brout [43], followed independently by Higgs [44]. The simplest description of the mechanism added a Higgs field that permeates all of space, to the SM. A Higgs boson was predicted as the excitation of the Higgs field. On 4 July 2012, two experiments at the LHC announced the observations of a new particle [45, 46] which was confirmed to be the Higgs boson [47, 48]. The successful operation of LHC delivers a lot of data to not only profoundly study the properties of the Higgs, but also test the SM predictions to all possible processes. As the heaviest particle found so far, the top quark plays an important role in understanding the structure of the SM and the Brout-Englert-Higgs mechanism. Before the LHC era, most of the experiments were limited by the small number of available top-quarks. In contrast, the LHC is a top-factory that produced 35 million of top-quark events for a single experiment in 2016 alone. The large samples of top-quarks collected offer the opportunity to scrutinise the physics involving top-quark.

The topic of this thesis is related to the top-quark physics. It is organised as follows. The theory of particle physics and experimental high energy physics are briefly summarised in Chapter 2.

Chapter 3 describes how particles are detected and reconstructed experimentally, in particular at the LHC. Chapter 4 presents data and simulation samples used in the thesis, and furthermore, from reconstructed particles, how one can identify an event of interest. In order to estimate signal yields in data, a discriminator that can separate signal and background is necessary. Chapter 5 details a multivariate analysis to achieve better separation. The sources of systematic uncertainties are presented in Chapter 6. All these are combined in a maximum likelihood fit used for extracting the results. In addition, differential cross-section measurements are performed and described in Chapter 7, including techniques and results. Last but not least, a new deep learning technique to further reduce systematic uncertainty is tested and presented in Chapter 8. The final chapter gives a short summary of this thesis.

Theoretical and experimental basics

This chapter introduces the theoretical and experimental basics of particle physics. Section 2.1 briefly describes quantum field theory, and Section 2.2 summarises the most successful model in particle physics so far, the Standard Model. Some basic information on particle colliders is presented in Section 2.3, while Section 2.4 and Section 2.5 detail the physics related to the topic of this thesis, the top quark and its properties.

2.1 Introduction to the quantum field theory

Quantum field theory (QFT) is a theoretical framework to combine quantum mechanics, classical fields, and special relativity. In this theory, particles are quanta of the corresponding fields, represented by wave functions, and interactions between particles are mediated by the exchange of particles. QFT is developed from the Lagrangian formalism¹ of classical field theory, and quantised by imposing canonical (anti-)commutation relations on it.

2.1.1 The Lagrangian density

The Lagrangian density, or simply the Lagrangian, is a function of the fields and their derivatives. The time and spatial integral of the Lagrangian, the *action*, is stationary with respect to the changes of paths, known as the principle of least action. Physical insights can be revealed by the terms of the Lagrangian. For example, a product of three or more fields denotes an interaction between the corresponding particles, in other words these particles couple to each other; the coefficient of the product illustrates the intensity of the coupling (*coupling strength*); a product of one field with itself and its derivative indicates the mass and the kinetic energy of the corresponding particle, respectively.

2.1.2 (Local) Gauge invariance

An essential feature of QFT is its invariance under *local* gauge transformations. Unlike global transformation that applies the same phase transformation at each point in space-time, local transformation implies different phase transformations at different points. It is possible to generalise these gauge

¹ The Lagrangian formalism manifests Lorentz-invariance, thus is preferred in order to be consistent with relativistic field theory.

transformations to more sophisticated groups of transformations, leading to a large class of so-called gauge theories. The required local gauge invariance (or gauge symmetry) is encoded to the invariance of the Lagrangian under Lie groups², via replacing the partial derivative in the Lagrangian by the gauge covariant derivative.

The gauge covariant derivative introduces additional fields, called *gauge fields*. Interactions between particles are introduced as well, mediated by the gauge particles excited from gauge fields. These particles are then said to be charged under the gauge fields. Within a certain type of interaction, the more charges particles carry, the tighter they couple. The coupling strength indicates how strong this interaction is. Conversely, charges create a gauge field surrounding it. The dynamics of the gauge fields, such as the kinetic energy and self-interactions of gauge bosons if they exist, also enter into the Lagrangian.

2.1.3 Interaction of fields

Perturbative expansion

Physical quantities usually cannot be analytically solved via the equations of motion that arise from the Lagrangian. To overcome this problem, an approach is to expand the observable in a power series of the coupling strength. The power series is called a perturbative expansion. The individual terms of a perturbation series are known as *orders*, the first order of a series is called leading order (LO), the second order is called next-to-leading order (NLO), the third order is called next-to-next-to-leading order (NNLO), etc. Higher order calculation requires dramatically more efforts as the complexity grows exponentially. When the coupling strength is sufficiently small, i.e. $\ll 1$, the power series converges fast and the higher orders can be treated as perturbations of the lower orders. When it is not sufficiently small, other techniques are developed to accelerate the convergence. For instance, if the calculation contains logarithm patterns where the logarithm is smaller than 1, the series can be reordered and resummed by the power of the logarithms, and converge faster. Similarly, the first term of such logarithm series is called leading log, the second term is called next-to-leading log, the third term is called next-to-next-to-leading log (NNLL).

Invariant Matrix Element

When calculating the probability of a process to happen, it is convenient to write in bra-ket notation with a “bra” of outgoing (final) particles and a “ket” of incoming (initial) particles. To manifest the interaction, a unitary operator expressed as a matrix sits in between, which is called the scattering matrix. If the incoming particles miss each other or do not interact at all, the scattering matrix is simply the identity matrix. Thus, to isolate the reaction, an identity matrix is deducted from the scattering matrix. Furthermore, stripping the energy and momentum conservations that are satisfied by any process, it ends up with the invariant Matrix Element (ME), $\mathcal{M}$, which conveys the information of interaction, and is different for any process.

² In mathematics, group theory studies the algebraic structures of groups. It is a perfect tool for studying symmetry since symmetry forms a group. Gauge symmetries can be represented by Lie groups, which are continuous groups.

Feynman diagram

A Feynman diagram (or Feynman graph) is a formal tool in perturbation theory, serving to evaluate the ME. It consists of external lines representing incoming and outgoing particles, internal lines representing intermediate particles, and vertices representing the coupling strength of the interaction occurring. Arising uniquely from the Lagrangian, all these components have precise associations to mathematical expressions, defined by Feynman rules. The full expression of a diagram corresponds to the multiplication of all components. Each order of the perturbation series may contain multiple Feynman diagrams. A physical observable is evaluated by the square of the sum of all possible Feynman diagrams contributing to it. Likewise, to calculate the observable up to a certain order, the diagrams that contribute up to that order need to be included. This can be identified by counting the vertices in a Feynman diagram, as the number of vertices reflects the order of the coupling strength. The calculation, in general, contains not only the squares of the individual graphs, but also interference terms between the graphs, stemming from quantum interference.

2.2 The Standard Model of particle physics

The Standard Model (SM) is an assemblage of theories that embodies all our current understandings of fundamental particles and forces, supported by a great deal of experimental observations. According to the SM, quarks and leptons are the building blocks of the matter, and forces act through carrier particles exchanged between the matter particles. The complete picture of the SM requires 18 parameters³, including nine quark and lepton masses, three strengths of force, four parameters from the quark mixing model (Section 2.2.4), and two directly related to the Higgs sector (Section 2.2.3).

2.2.1 Ingredients of the SM

Fermions and their antiparticles

Matter is composed of elementary particles with a spin⁴ of one half, called fermions. There are twelve fermions: six quarks and six leptons. The quarks and leptons are arranged in three generations, where each generation is identical to the others except for their masses and flavours. An overview of the particles in the SM is given in Fig. 2.1. The electron (e), the electron neutrino (ν_e), the up quark (u), and the down quark (d), known as the first generation, represent the basic building blocks of the low-energy universe. For instance, the proton consists of two up quarks and one down quark while the neutron is made up of one up quark and two down quarks. Protons and neutrons make up the nuclei, and surrounding with one or more electrons, they form different types of atoms. When encountering a higher energy scale, however, the other generations are revealed. Every generation of quarks consists of one “up-type” quark (u , the charm quark c , or the top quark t), carrying electric charge of two thirds, and one “down-type” quark (d , the strange quark s , or the bottom quark b), having negative one third electric charge. All quarks have a unit colour charge. Each lepton generation consists of one charged lepton (e , muon μ , or tau τ) with one negative electric charge and an associated electrically neutral partner called neutrino (ν_e , muon neutrino ν_μ , or tau neutrino ν_τ). No colour charges are carried

³ QCD vacuum angle is not counted.

⁴ Spin is an intrinsic angular momentum carried by the particles. According to the spin-statistics theorem in relativistic quantum field theory, particles with integer spin are bosons, while particles with half-integer spin are fermions.

by leptons. Except for the massless neutrinos⁵, the masses of the other fermions are undetermined in the theory and are measured by experiments. Among the quarks, three of them (u , d , and s) with masses below 1 GeV are called *light* quarks.

Every fermion has an antiparticle of itself with the same mass but opposite charges. Particle–antiparticle pairs can annihilate each other, by the simultaneous creation of both a particle and its antiparticle. Antiparticles are denoted by putting a bar on top of the symbol of the corresponding particle, or switch the sign of its electric charge, i.e. the antiparticle of X being $\bar{X}$ or the antiparticle of X^- being X^+ . Sometimes though, its anti-partner is implied by using the name of the particle when it is clear from the context. For instance if a W boson decays to a up-down pairs, one of the quarks must be an antiparticle, which can be derived from the electric charge conservation.

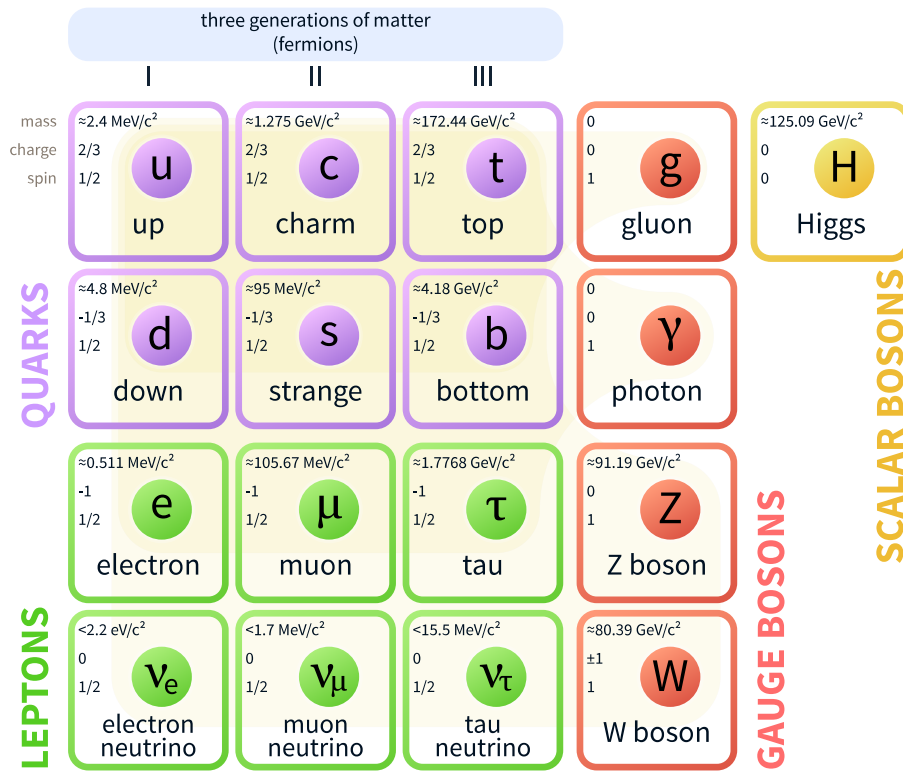


Figure 2.1: Overview of the particles in the SM and some of their properties. The quarks (in lavender) and the leptons (in green) are the building blocks of matter. The gauge bosons (in red) are force carriers. The scalar boson (in yellow) is the excitation of the Higgs field that introduces mass to the elementary particles, and will be discussed in Section 2.2.3. The quarks and leptons are grouped into three generations. Their antiparticles are not shown. Figure is taken from [49].

⁵ The absence of the right-handed components makes it difficult for the neutrinos to acquire masses within the SM. However, the discovery of the neutrino oscillation indicates that their masses are non-zero.

Gauge bosons and fundamental interactions

The interactions between quarks and leptons are mediated by the exchange of particles with a spin of one, known as the gauge bosons or vector bosons (see Fig. 2.1). There are four fundamental interactions: the gravitational, electromagnetic, strong and weak interactions. Except for the gravitational interaction, the other three are incorporated in the SM. The electromagnetic force is carried by the photon (γ) and is responsible for the attraction between orbital electrons and atomic nuclei, chemical bonding and electromagnetic waves. The weak force, carried by $W^\pm$ (or W to refer to both) and Z^0 (or Z to ignore the charge) bosons, acts between subatomic particles. It causes radioactive decay and plays an essential role in nuclear fission and fusion. The strong interaction is mediated by a particle called a gluon (g), which binds quarks together to form hadrons, such as protons and neutrons. Its residual creates the nuclear force that binds protons and neutrons to atomic nuclei. Fermions couple to bosons by charges that are distinctive for different interactions. The electric charge, the colour charge, and the weak hypercharge and isospin are the source of the electromagnetic force, the strong force, and the weak force, respectively. The masses of the bosons arise after the spontaneous symmetry breaking by the Brout-Englert-Higgs mechanism [43, 44].

2.2.2 Electroweak unification

Although the two forces appear to be very different at low energies, the electromagnetic and weak interactions can be unified above the unification energy in electroweak theory.

Gauge bosons in electroweak theory

In quantum electrodynamics (QED), the local gauge invariance of the Lagrangian is represented by $U(1)$ group, where there are a single gauge coupling e and a single gauge boson, γ . It turns out that the weak isospin and hypercharge are more fundamental than the electric charge. The weak isospin of a particle describes a transformation under $SU(2)$, while the hypercharge describes how it transforms under $U(1)$, which is different from $U(1)$ in QED. Therefore, the electroweak model is based on a more complicated group $SU(2) \otimes U(1)$. The Lagrangian of the model introduces three bosons $W^{(i)}$ ($i = 1, 2, 3$) associated with $SU(2)$, and one boson B associated with $U(1)$. However, these states are not the bosons observed in the experiments. The physical $W^\pm$ bosons can be identified as the linear combinations of $W^{(1)}$ and $W^{(2)}$. And the weak mixing angle, or the Weinberg angle θ_W , is introduced to describe a linear mixture of $W^{(3)}$ and B , giving rise to Z and γ . The coupling strengths of the γ , W , Z fields, g_e , g_W , g_Z , respectively, are related to each other as follows:

$$g_e = g_W \sin \theta_W = g_Z \sin \theta_W \cos \theta_W,$$

Problems in the electroweak model

The electroweak model, beautiful as it is, has some serious shortcomings. All vector bosons involved are required to be massless since their mass terms violate gauge symmetry. However, in reality, the W and Z bosons are so massive that the weak force only acts at very small distances. In addition, fermion masses are absent as well. Left-handed fermions with spins anti-parallel to their momenta form isospin doublets of $SU(2)$, while right-handed fermions with parallel spins and momenta form isospin singlets. Doublets and singlets transform differently under $SU(2)$. As a result, the mass term

of a fermion, which can be written as a product of left-handed and right-handed terms, does not respect gauge invariance. Furthermore, several processes in the SM such as WW -scattering violate unitarity at high energy. These problems are solved by the Brout-Englert-Higgs mechanism.

2.2.3 Higgs field and its boson

The Brout-Englert-Higgs mechanism and the associated Higgs field are an essential part of the SM. The Higgs mechanism, in short, allows the W and Z bosons to acquire masses without breaking the local gauge symmetry. Leptons and quarks acquire their masses by interacting with the Higgs field.

Higgs mechanism

The Higgs mechanism adds an additional field to the theory, the Higgs field, which has a Mexican-hat-like potential. This potential is chosen such that its vacuum expectation value, i.e. the position of its minimum, is non-zero. Although the Lagrangian is still symmetric in the field, the perturbations around the minimum (the vacuum) are not symmetric any more, which is known as *spontaneous symmetry breaking* or electroweak symmetry breaking (EWSB). As a complex field, its perturbative expansion about the vacuum state contains perturbations both in the radial direction and in the tangential directions along the trough of the potential. The former describes a massive scalar field, the Higgs field, while the latter describes one or more⁶ massless scalar bosons, named Goldstone bosons. Except for the masses and dynamics of these scalar fields, the Lagrangian introduces the mass terms of the gauge bosons, and interaction terms between the Goldstone fields and the gauge fields. By choosing the unitary gauge, the interaction and Goldstone dynamics terms can be absorbed by the gauge transformation of the gauge fields. Sometimes it is said that the Goldstone bosons are “eaten” by the gauge fields. As a result, a massive Higgs field and massive gauge fields remain in the Lagrangian; while the Goldstone fields no longer appear.

Boson masses

The Higgs field of the SM is described by a complex doublet, which has four independent degrees of freedom. Three degrees of freedom create three Goldstone bosons “eaten” by the three electroweak mediators, $W^\pm$ and Z^0 . From this, the W and Z bosons acquire their masses (m_W , m_Z , respectively), connected with the vacuum expectation value (v) of the Higgs field and the coupling constants (g_W and g_Z) of the $SU(2)$ gauge interactions:

$$m_W = \frac{1}{2} \cdot g_W v,$$

$$m_Z = \frac{1}{2} \cdot g_Z v.$$

⁶ The new introduced field may contain N degrees of freedom. It leaves $N - 1$ massless fields and needs $N - 1$ gauge fields to absorb them. For example, in the SM the Higgs field is a complex doublet containing four degrees of freedom. So three massless fields are generated.

With measured values of m_W and g_W , v is found to be 246 GeV. An unexpected bonus is that the masses of the W and Z bosons are related by the Weinberg angle:

$$\frac{m_W}{m_Z} = \cos \theta_W.$$

The remaining degree of freedom gives rise to the Higgs boson. The mass of the Higgs boson, determined by the coefficients of the Higgs potential, is not predicted by the theory. In July 2012, the ATLAS and CMS collaborations announced the discovery of a new particle with a mass of 125 GeV [45, 46] and a year after it was confirmed as the Higgs boson of the SM [48, 50]. The Higgs boson interacts with itself and with the W and Z bosons. Adding these interactions also gives a solution to the aforementioned unitarity problem in the electroweak model. The photon, on the contrary, stays massless due to a remaining $U(1)$ symmetry of the QED.

Fermion masses

For each fermion, a doublet Higgs field is inserted in the production of the left-handed doublet and the right-handed singlet of the fermion. Such a term is gauge invariant and can be added to the Lagrangian. By choosing the unitary gauge, each fermion leaves two terms in the Lagrangian: one with the coupling of the fermion to the Higgs field through its non-zero vacuum expectation value, the other one with the interaction between the fermion and the Higgs boson itself. The fermion masses (m_f) arise from the first term, while the second term gives the coupling strength (g_f) of the fermion to the Higgs boson, proportional to the fermion mass:

$$g_f = \sqrt{2} \cdot \frac{m_f}{v}.$$

There are nine independent parameters g_f , or equivalently m_f , for each fermion, excluding neutrinos whose masses are set to zero.

2.2.4 Quarks in electroweak theory

The leptonic weak vertices only connect leptons in the same generation. Therefore, the individual conservations of electron number, muon number, and tau number are guaranteed. In the quark sector, however, transitions between different generations are possible, even though the strengths are weaker than those between the same generations. These phenomena triggered the invention of the Cabibbo–Kobayashi–Maskawa (CKM) matrix [51, 52]. It takes the assumption that weak eigenstates of the quarks that couple to the W are not identical to the mass eigenstates that couple to the Higgs. The weak eigenstates (d', s', b') are defined as linear combinations of the mass eigenstates (d, s, b), related by the CKM matrix V_{CKM} :

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}.$$

The matrix element V_{ij} squared gives the probability of a transition from quark i to quark j or vice versa. In Feynman diagrams, the matrix elements are associated with the corresponding vertices, for

example, the Wtb vertex includes a factor of V_{tb} .

The CKM matrix must be unitary as it relates two orthonormal bases. It is also complex in order to explain CP (symmetry) violation⁷ in the SM. To isolate the contribution to CP violation, a useful approximation is given by the Wolfenstein parametrisation [53]. It reads as follows, together with the current best experimental values [54]:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)$$

$$= \begin{pmatrix} 0.97434^{+0.00011}_{-0.00012} & 0.22506 \pm 0.00050 & 0.00357 \pm 0.00015 \\ 0.22492 \pm 0.00050 & 0.97351 \pm 0.00013 & 0.0411 \pm 0.0013 \\ 0.00875^{+0.00032}_{-0.00033} & 0.0403 \pm 0.0013 & 0.99915 \pm 0.00005 \end{pmatrix}.$$

A , ρ , and η are real numbers intended to be of order unity, and $\lambda \approx 0.22$. From the Wolfenstein parametrisation or the experimental values, one can see that the CKM matrix has a clearly hierarchical structure. The diagonal values are close to one, while the off-diagonal values are reduced by about a factor of λ for each step away from the diagonal. It illustrates that quark mixing in general is small.

2.2.5 Quarks in the strong interaction

Quarks not only participate in the electroweak interactions, but are also involved in the strong interaction, which is represented by the theory of quantum chromodynamics (QCD).

Colour charge of quarks and gluons

The underlying theory of QCD is similar to that of QED: instead of requiring local $U(1)$ gauge invariance, posing the transformation under $SU(3)$. The electric charge of QED is then replaced by three *colour* charges, r , b , and g . Quarks carry one unit of colour charge while antiquarks carry an opposite colour (anti-colour) charge, $\bar{r}$, $\bar{b}$, and $\bar{g}$. Due to the natural structure of $SU(3)$, the colour charges are exchanged by eight different gauge fields, corresponding to eight massless gluons. The gluons form an colour octet and all of them carry colour change. This feature enables gluon self-interaction. Quarks and gluons are the only particles participating the strong interaction.

Colour confinement

Colour confinement, or simply *confinement*, is the phenomenon that colour charged particles are always confined to colour neutral (colourless) states, and thus isolated particles with non-zero colour charge cannot be directly observed. It can be qualitatively explained by the QCD coupling strength and gluon self-interaction. The interaction between quarks can be thought as exchanging virtual gluons. Since the QCD coupling strength is the order of 1, these gluons can easily radiate other gluons. The colour field is then squeezed in a tube between the two quarks. The potential energy stored in the tube is not only dense, but also nearly constant along the distance. When the two quarks are moving apart,

⁷ CP symmetry violation plays a key rule in the dominance of matter over antimatter in the present universe. The initial proposal by Kobayashi and Maskawa was motivated by it, when only three types of quarks had been discovered experimentally. They concluded that imaginary parts of some CKM matrix elements can be used to explain it.

the potential energy increases almost linearly, thus requires input energy⁸. When the input energy to the bi-quark system is high enough, another pair of quarks can be created from the vacuum to keep the system at a low energy level. This can recursively happen to the newly born quark pairs. As a result, coloured particles arrange themselves into colourless bound states that are known as *hadrons*. The process to form a hadron from a bare quark or gluon is called *hadronisation*. Due to the confinement, the strong interaction is a short distance interaction.

Asymptotic freedom

The QCD coupling strength, also called the QCD coupling constant in many places, is *not* constant at all. Its value depends on the energy scale, or equivalently, the distance of the interaction being considered.⁹ This feature can be derived theoretically, but here will be demonstrated pictorially. The field surrounding a colour charged particle gets polarised due to a cloud of virtual particle-antiparticle-pairs and mediators continuously popping in and out of existence. Although the virtual anti-coloured quarks screen the original quark's colour charge, the virtual gluon cloud has an anti-screening effect that prevails. As a result, the effective colour charge becomes larger with a larger distance. At high energies (short distances), the QCD coupling strength becomes sufficiently small and the theory is asymptotically free. For instance, at an energy scale above 100 GeV, the QCD coupling strength is of the order of 0.1 and perturbation theory can be used, although the higher-order corrections cannot be neglected.

2.2.6 Resonances

In particle physics, resonance particles, or resonances, refer to those extremely short-lived particles, whose lifetime is of the order of 10^{-23} seconds. Even travelling at almost the speed of light, these particles could only pass about 10^{-15} meters, about the diameter of a proton, before decaying. If we fill the invariant mass of a resonance's decay children into a histogram, the histogram will peak at the mass of the resonance.¹⁰ In scattering processes, the incoming particles with total energy near the invariant mass of a resonance are keen to bind for a while, increasing the scattering cross-section at that energy.

Hadrons

Most of the resonances in particle physics are hadrons. The quarks are held together by the strong force, which also contributes to the hadron mass in the form of potential energy. Hadrons are grouped into two categories: baryons, made of three quarks, and mesons, made of one quark and one antiquark¹¹. Protons and neutrons are examples of baryons, while pions ($\pi^\pm, \pi^0$) and kaons ($K^\pm, K_S^0$, etc.) are examples of mesons. Among all hadrons, protons are stable, and neutrons bound within atomic nuclei are stable. Other hadrons are unstable under ordinary conditions and will decay to other particles.

⁸ From experimental measurements, the potential in the colour field is $\sim 1 \text{ GeV fm}^{-1}$.

⁹ The statement is also true for all other coupling constants, even though the dependencies on the energy scale are different for each of them.

¹⁰ The invariant mass, $m = \sqrt{E^2 - \vec{p}^2}$, is Lorentz-invariant. Thus it is unchanged before and after the decay.

¹¹ Hadrons containing more than three valence quarks (exotic hadrons) have been discovered in various experiments [55–57]. However, their features do not match the expectation of QCD, and whether they are hadrons is debatable.

2.2.7 Flaws and limitations of the SM

Being the best theoretical description in particle physics, though, the SM does not explain the complete picture of the universe. One major problem is that the gravity is omitted in the theory. In addition, the amount of matter in the universe predicted using the CKM matrix falls several orders of magnitude short of what is observed by astrophysicists [58]. Therefore, other CP violation sources must contribute. Moreover, discoveries of neutrino oscillations from many sources and over a wide energy range require neutrinos to be massive [59], which contradicts the SM assumption. Last but not least, the SM only describes visible matter, that is all matter that makes up the things around us: the Earth, the stars and galaxies we have observed. But cosmological observation indicates that the universe contains about five times more *dark matter* [60], participating only gravitational force. Among all the particles in the SM, none has the properties of dark matter. To explain the drawbacks of the SM, a lot of theories beyond the SM [61], known as new physics, are developed and tested. However, no new physics has been discovered experimentally so far.

2.3 Collider physics

In order to test the SM and search for new physics, colliders are built to accelerate particles to very high energy and let them interact. Analysing outcomes of the collisions reveals the structure of the subatomic world and the laws that govern it. These become apparent at high energies, and therefore high energy colliders are powerful tools for the research.

2.3.1 Basic concepts

Natural units

The natural units system is widely used in high energy physics. It sets the speed of light c , the reduced Planck constant $\hbar$, and the Boltzmann constant k_B to unity, so that they can be dropped from all equations. In this way, all quantities are expressed in powers of eV:

$$[\text{energy}] = [\text{momentum}] = [\text{mass}] = [\text{temperature}] = [\text{width}] = \text{eV},$$
$$[\text{time}] = [\text{length}] = \text{eV}^{-1}.$$

The natural units system not only largely reduces the complexity of the equations and calculations, but also makes dimensional analysis easier.

Cross-section and decay rate

Cross-section, σ , describes the probability of two particles react during a collision. It is an effective size of an area transverse to their relative motion, within which they undertake the process. The unit used in particle physics is barn (b), where $1 \text{ b} = 10^{-28} \text{ m}^2$. More commonly used are its submultiples: $1 \text{ pb} = 10^{-12} \text{ b}$ or $1 \text{ fb} = 10^{-15} \text{ b}$. It is calculated as an integration of the ME over the allowed phase space with additional factors reflecting four-momentum conservation, coupling strength, normalisation, etc. When the phase space covers all scattering angles (and possibly other variables), it is called a total (or *inclusive*) cross-section. When a cross-section is specified as a function of some certain

observables, such as an angle or energy (of final-state particles), it is called a *differential* (or exclusive) cross-section.

Decay rate, or decay width, reflects the probability per unit time that a given particle decays. It is associated with the statistical behaviour of populations of identical particles. For an ensemble of sufficiently large amount of the particles, the expected number surviving as a function of time follows an exponential decay. The exponential decay constant is called the decay rate, denoted by Γ . The time after which the ensemble is expected to be reduced to $1/e$ of the original size is called the lifetime, $\tau = 1/\Gamma$. The higher the decay rate is, the shorter the particle lives. When multiple decay modes are available, as is often the case, individual decay rates (Γ_i) can be associated with each mode, and the total rate (Γ_{tot}) is the sum of the individual rates:

$$\Gamma_{\text{tot}} = \sum_{i=1}^n \Gamma_i.$$

The particle lifetime is then given by

$$\tau = \frac{1}{\Gamma_{\text{tot}}}.$$

The unit of the lifetime is eV^{-1} or GeV^{-1} in natural units, while the decay width is in eV or GeV . The ratio of the partial decay width to the total decay width defines the branching fraction of an individual decay mode:

$$\mathcal{B}_i = \frac{\Gamma_i}{\Gamma_{\text{tot}}}.$$

Event rate and luminosity

The event rate of a physical process in a collision experiment is proportional to its cross-section:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma.$$

The factor $\mathcal{L}$ is called the *instantaneous* luminosity, which reflects the property and capability of an accelerator. For two Gaussian beams colliding head-on in the z -direction, it can be calculated by

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

where f denotes the crossing frequency, n_1, n_2 the number of particles in the two beams, σ_x, σ_y the Gaussian shaped beam sizes.

One can get the total event number of the process by integrating the rate over time:

$$N = \sigma \cdot \int \mathcal{L} dt \equiv \sigma \cdot \mathcal{L}_{\text{int}}.$$

$\mathcal{L}_{\text{int}}$ is called the *integrated* luminosity.

In practice, the instantaneous luminosity is hard to measure precisely. Therefore, the integrated luminosity is determined by the event number of some accurately known processes, such as QED

processes:

$$\mathcal{L}_{\text{int}} = \frac{N_{\text{QED}}}{\sigma_{\text{QED}}}.$$

However, at the Large Hadron Collider (LHC) such processes are hard to find. Therefore, van der Meer scans [62] are adopted to measure the beam parameters under special running conditions and extrapolate to normal running conditions to calculate the instantaneous luminosity.

2.3.2 Hadron collider

A hadron collider smashes two high energy hadron beams into each other. It is able to reach higher centre-of-mass energy than electron-positron (e^+e^-) colliders. For example, the LHC accelerates protons to the energies of ~ 6.5 TeV. In contrast, the energies of electrons and positrons in the LEP reached around 100 GeV. Thus a hadron collider is a great place to search for new heavy particles.

Parton distribution function

In a proton, the valence and virtual quarks, and gluons are collectively called *partons*. In a proton–proton inelastic collision, this parton–parton collision is denoted as the *hard* collision/scattering. Since the parton carries a part of the energy and momentum of the proton, the effective collision energy is lower than the sum of the beam energies, and the flavour of the initial parton is unknown. Both pieces of information are encoded in the parton distribution function (PDF) of the proton, which is defined as the probability density for finding a certain parton with a momentum fraction $x = \frac{p_{\text{parton}}}{p_{\text{proton}}}$ inside a proton. The cross-section to produce a final state X in the collision of two protons at a centre-of-mass energy $\sqrt{s}$ can be expressed as

$$\sigma_{p_1 p_2 \rightarrow X}(s) = \sum_{i,j=\text{partons}} \int dx_i dx_j \cdot f_1(x_i, \mu^2) f_2(x_j, \mu^2) \sigma_{ij \rightarrow X}(\hat{s}, \mu^2).$$

In this equation, i, j refer to the partons in two protons p_1, p_2 ; f_1 and f_2 are the PDFs of the protons defined at the energy scale μ ; $\sigma_{ij \rightarrow X}$ is the cross-section for i and j to create X at the centre-of-mass squared $\hat{s} = x_i x_j s$.

Kinematics at hadron colliders

Transverse momentum. Since the partons are boosted along the beam direction (z direction), the sum of their projection to the transverse plane¹², the transverse momentum, is close to zero. The total transverse momentum of the final objects then vanishes due to momentum conservation. To access the information of undetectable particles, such as neutrinos, one can reconstruct all the other objects, and deduct them from zero as *missing* transverse momentum. If there are two or more neutrinos in the final states, however, the missing transverse momentum reflects the vector sum of all missing momenta.

¹² The transverse plane is perpendicular to the beam direction.

Rapidity. Because of the unknown boost of the partons, the rapidity y is used, defined as the following:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right),$$

where E is the energy and p_z is the projection of the momentum along the z axis. It can be shown that upon a Lorentz transformation $\beta = \frac{v}{c}$ along the z axis, the transformation on rapidity is

$$y' = y - \tanh^{-1} \beta.$$

Thus, rapidity differences are invariant under boosts along the beam direction.

Pseudorapidity. The disadvantage of the rapidity is that both the energy and the z -component of the momentum need to be measured. In the limit of high energy, $|\vec{p}| \simeq E$, the pseudorapidity η is a lot easier to calculate:

$$\begin{aligned} \eta &= \frac{1}{2} \ln \left(\frac{|\vec{p}| + p_L}{|\vec{p}| - p_L} \right) \\ &= -\ln \tan \left(\frac{\theta}{2} \right), \end{aligned}$$

where θ is the polar angle. Paired with the azimuthal angle ϕ , a boost-independent angular distance between two particles can be defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2},$$

where $\Delta\eta$ and $\Delta\phi$ are the differences of pseudorapidity and the azimuth, respectively.

Transverse mass. The transverse mass, m_T , is another quantity that is invariant under Lorentz boosts along the beam axis. It is defined as

$$m_T^2 = E_T^2 - p_T^2 = \sum p_T^2 - \sum \vec{p}_T^2,$$

where $E_T = \sqrt{p_T^2 + m^2}$ is the transverse energy and becomes $\sqrt{p_T^2}$ in the high energy limit.

Underlying event and pile-up

The underlying event (UE) consists of tracks and energy that come from other partons in the same beam particles, from non-scattering proton remnants, and from other beam particles which happen to interact. Pile-up (PU) comes from the objects outside of the hard collision. There are two sources of PU: in-time and out-of-time. In-time PU is the result of multiple hadron interactions in the same bunch crossing. Out-of-time PU comes from earlier and later bunch crossings, which leave electronic signals in the detector.

2.3.3 Detectable particles

Detectable particles of a collision must live long enough to travel several metres. They include e , μ , $K^\pm$, $\pi^\pm$, p , γ , n . Nuclei could also be counted if they are produced. Other particles decay before reaching even the innermost subdetector, with a few exceptions that fly a short distance in the detector, such as K_S^0 , Λ , B mesons, and τ leptons.

A highly energetic parton undergoes QCD radiation thereby losing energy. It leads to the formation of parton showers (PSs). Hadronisation also takes place afterwards. The hadrons (including their decay products) tend to cluster concentrated in a narrow cone around the original parton momentum direction, which is called a *jet*. On average, about 60 % of the jet energy is from charged particles (mostly $\pi^\pm$) and 30 % is from γ via π^0 decays. The remaining 10 % is carried by neutral hadrons.

2.4 Top-quark physics

2.4.1 Discovery

The top quark, as the weak isospin partner of the bottom quark, was predicted by Kobayashi and Maskawa in 1973, when they were trying to explain the observed CP violation. In 1975, the tau lepton was discovered [31]. This gave more confidence in the existence of the third generation. Shortly after that, in 1977, the Υ meson was discovered at Fermilab [33] as a resonance in the $\mu^- \mu^+$ invariant mass spectrum, interpreted as the bound state of $b\bar{b}$. Lots of effort was put on measuring the bottom-quark properties. For example, the electric charge was measured as $-1/3 e$ and weak isospin being $-1/2$. These properties strongly suggest the existence of its weak isospin partner, the top quark. After about two decades of searching for it, it was finally discovered in 1995 by the CDF and DØ experiments [39, 40] at the Fermilab Tevatron.

2.4.2 Top-quark properties

The top quark is the “up-type” quark of the third generation with one unit colour charge and two thirds positive electric charge. It is the heaviest elementary particle in the SM, with a mass $m_t \sim 173 \text{ GeV}$ [63], ~ 35 times heavier than the next fermion, the bottom quark, and also heavier than the Higgs boson ($m_H \sim 125 \text{ GeV}$). The width of the top quark is $\Gamma_t \approx 1.41 \text{ GeV}$, corresponding to a very short lifetime of $\tau_t \approx 5 \times 10^{-25} \text{ s}$. In fact, it is smaller than the hadronisation time $\tau_{\text{QCD}} \sim 10^{-23} \text{ s}$. This means that top quarks decay before they form hadrons. It provides a great opportunity to explore the interactions of a bare quark at a few hundred GeV. Furthermore, top-quark decay also preserves the spin polarisation and spin-spin correlations which manifest themselves in characteristic angular distributions and correlations of the final states, providing direct access to the properties of the bare top quark.

2.4.3 The role of the top quark in the EWSB

The top quark has an unquestionable leading role with a mass close to the EWSB. The top-quark mass enters the electroweak precision observables as an input parameter via loop corrections. For example, it dominates the main production mechanism $gg \rightarrow H$ (Fig. 2.2(a)) and contributes in an important

way to one of the decay channels $H \rightarrow \gamma\gamma$ (Fig. 2.2(b)). It is sensitive not only to the absolute value of the top quark to Higgs coupling but also to its sign.

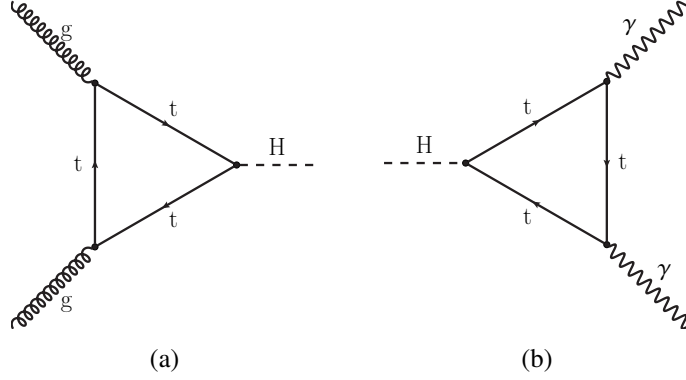


Figure 2.2: SM contributions to $gg \rightarrow H$ via a top-quark loop (a) and $H \rightarrow \gamma\gamma$ via a top-quark loop (b).

In addition, the request of the vacuum expectation value v to be the true minimum of the Higgs potential, up to a cut-off scale, Λ , implies a relation between Λ , the Higgs boson mass and the top-quark mass [64]. At scales larger than Λ , the Higgs potential becomes negative, and then it develops a new minimum, potentially leading to an unstable electroweak vacuum. According to the recently measured Higgs mass and top-quark mass, the vacuum state is at the boundary of being stable and metastable. It would be a bit odd if the vacuum is not certainly stable with a lifetime of the age of the universe.

Furthermore, the Higgs mass receives radiative corrections from each existing fermion and gauge boson, and other types of corrections. The corrections from the fermions (mainly from the top quark) are massive and of order λ . As λ is normally chosen as $10^{16} - 10^{19}$, it is unnatural for the Higgs mass to be the currently measured value (~ 125 GeV). In some scenarios beyond the SM, heavier particles are introduced and almost cancel the contributions from fermions [65]. Studying detailed properties of top quarks and its coupling with the Higgs boson can give handles on new physics.

In a word, understanding the top quark is crucial to give answers to these open questions and help us to understand the nature better.

2.4.4 Top-quark production at hadron colliders

At hadron colliders, top quarks are produced via two production mechanisms: in pairs $pp (\bar{p}) \rightarrow t\bar{t}$, where a virtual gluon splits into a top-antitop-quark pair via strong interaction, and as singly coupled with a W boson and a bottom quark via the electroweak force. In this section, hadron colliders refer to the Tevatron or the LHC. The Tevatron [66], located at Fermilab in the United States, was a proton-antiproton collider, which made the top-quark discovery as mentioned earlier. The LHC located at CERN near Geneva, Switzerland, is a proton-proton collider with higher centre-of-mass energy, which produced the data used in the thesis.

Top-quark pair production

The pair production ($t\bar{t}$ production) can be produced either by gluon-gluon fusion (gg fusion) or by two quarks annihilation ($q\bar{q}$ annihilation). The Feynman diagrams at LO are shown in Fig. 2.3. At

the LHC, gluon-gluon fusion becomes the dominant process. For comparison, at the Tevatron with centre-of-mass energy $\sqrt{s} = 1.96$ TeV the $q\bar{q}$ annihilation contributes $\sim 85\%$ of the total $t\bar{t}$ production, while at the LHC with $\sqrt{s} = 14$ TeV gg fusion contributes $\sim 90\%$ [54].

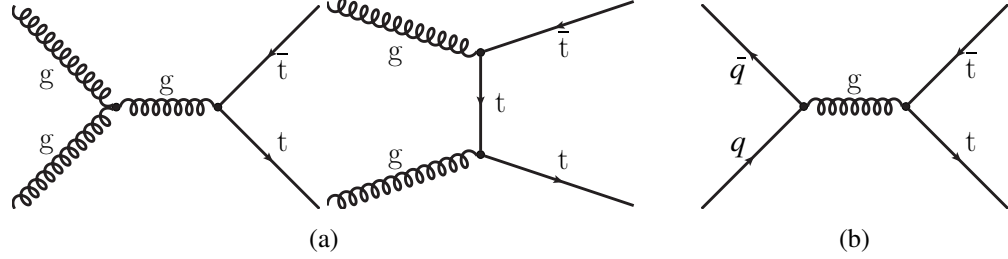


Figure 2.3: Example LO Feynman diagrams of the $t\bar{t}$ production by gg fusion (a) and by $q\bar{q}$ annihilation (b).

Theoretical calculations of top-quark pair production using perturbation theory are now available at NNLO with NNLL soft gluon resummation [67]. Assuming a top-quark mass of 172.5 GeV¹³, the pair production cross-section is computed as [68, 69]

$$\sigma_{t\bar{t}}(\sqrt{s} = 8 \text{ TeV}) = 252.9^{+6.4}_{-8.6} \pm 11.7 \text{ pb},$$

$$\sigma_{t\bar{t}}(\sqrt{s} = 13 \text{ TeV}) = 832^{+20}_{-29} \pm 35 \text{ pb}.$$

The first uncertainty comes from the renormalisation scales, while the second one is associated with the PDF choice and strong interaction coupling. Theoretical predictions and experimental measurements of the $t\bar{t}$ production cross-sections are summarised in Fig. 2.4.

Single top-quark production

Electroweak single top-quark production includes three different signatures: t -channel, s -channel, and associated production of a top quark and a W boson (tW ¹⁴). These production modes have smaller cross-sections compared with $t\bar{t}$ production. Single top-quark production was first observed in 2009 by the DØ [71] and CDF [72] experiments at the Tevatron via t -channel and s -channel combined. Example Feynman diagrams at LO for all three channels are presented in Fig. 2.5.

The t -channel is the dominating process where a bottom quark scatters off a light quark via a W boson and turns into a top quark. The bottom quarks in the initial states can be generated either by gluon splitting ($g \rightarrow b\bar{b}$) or directly from the proton according to its PDF. The next dominant process at the Tevatron is s -channel production, where light quark-antiquark-pair annihilate each other to a virtual W , thereby creating a top-bottom quark pair. In contrast, this process has the smallest cross-section at the LHC, where finding an initial antiquark is more difficult. In tW production, the most distinct feature is the fermionic propagator for the top-quark creation. The tW channel can be drawn as t -channel-like as Fig. 2.6(a) or s -channel like as Fig. 2.6(b). In the t -channel like diagram, the bottom quark first emits a W , turning to a virtual top quark, then interacts with a gluon. In the s -channel like diagram, the bottom quark first gets stuck by a gluon, becoming a virtual bottom quark, then “decays” to a W and a top quark. Similarly as in the t -channel, it requires the existence of a

¹³ The input top-quark mass gives the best description of the ATLAS data.

¹⁴ It is also known as Wt in some literature.

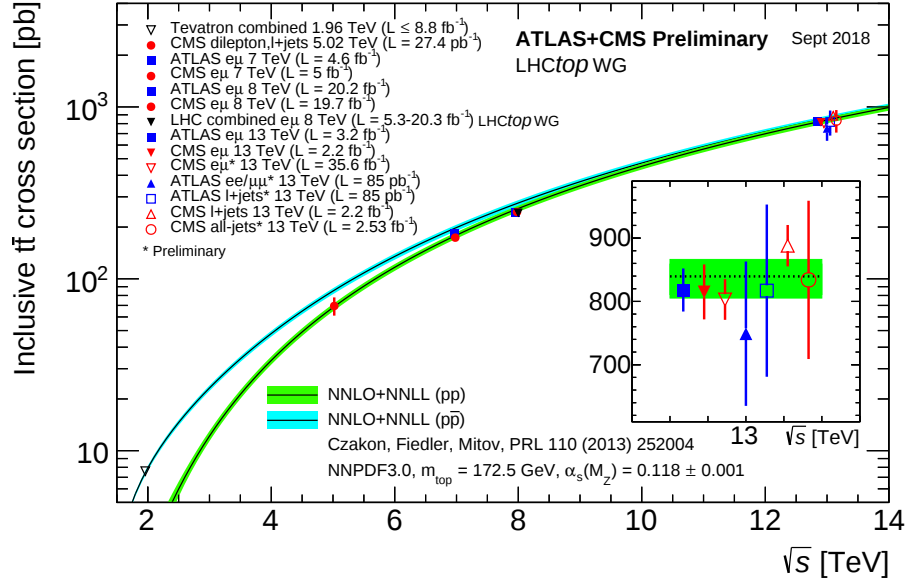


Figure 2.4: Theoretical predictions and experimental measurements of $t\bar{t}$ production cross-sections as a function of the centre-of-mass energy. Theory curves and uncertainties are generated at NNLO with NNLL soft gluon resummation for $m_t = 172.5$ GeV. ATLAS and CMS are two of the experiments at the LHC and will be introduced in Chapter 3. Figure is taken from [70] updated in September 2018.

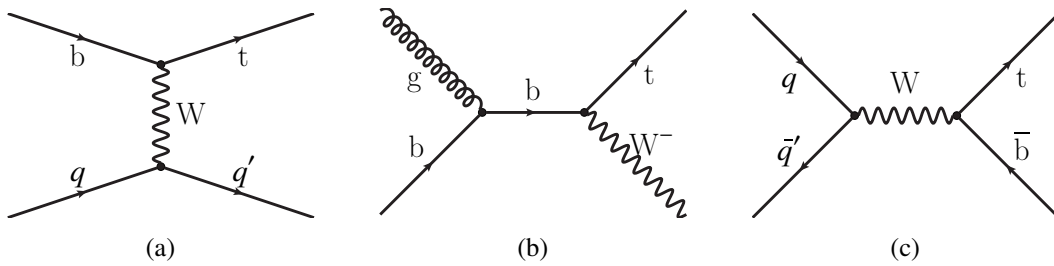


Figure 2.5: Example LO Feynman diagrams of the single top-quark productions via t -channel (a), tW channel (b), and s -channel (c).

bottom quark inside the proton, which gets solved by introducing a bottom-quark PDF. It is known as the five-flavour scheme theoretical description. As the name suggests, five flavours (u , d , s , c , and b) are treated as active flavours inside the proton. Without a bottom-quark PDF, the so-called four-flavour scheme, provides another theoretical description. In this scheme, a bottom quark is instead produced from gluon splitting and another bottom quark flies along the beam direction (see Figs. 2.6(c) to 2.6(d)). By doing this, the mass of the bottom quark could be non-zero while in five-flavour scheme it is set to zero.

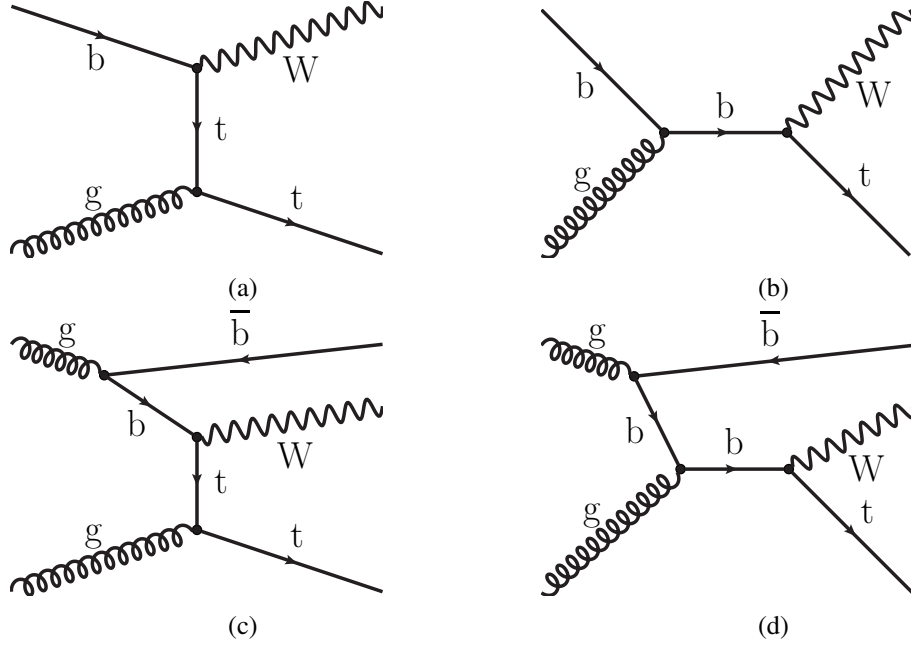


Figure 2.6: LO Feynman diagrams for the tW production of t -channel like (a) and s -channel like (b) in the five-flavour scheme, t -channel like (c) and s -channel like (d) in the four-flavour scheme.

With the same top-quark mass of 172.5 GeV assumption as is done in the top-quark pair production calculation, the cross-sections of single top-quark productions at the LHC at $\sqrt{s} = 13$ TeV are calculated as [73, 74]

$$\begin{aligned}\sigma_{t\text{-channel}}(\sqrt{s} = 13 \text{ TeV}) &= 217.0^{+6.6}_{-4.6} \pm 6.2 \text{ pb}, \\ \sigma_{tW\text{-channel}}(\sqrt{s} = 13 \text{ TeV}) &= 71.7 \pm 1.8 \pm 3.4 \text{ pb}, \\ \sigma_{s\text{-channel}}(\sqrt{s} = 13 \text{ TeV}) &= 10.32^{+0.29}_{-0.24} \pm 0.27 \text{ pb}.\end{aligned}$$

The first uncertainty comes from the renormalisation scales, while the second one is associated with the PDF choice and strong interaction coupling. Theoretical predictions of the cross-sections of the single top-quark production are summarised in Fig. 2.7, together with some of the existing experimental measurements up to June 2016¹⁵. The cross-sections of the tW channel and s -channel at 13 TeV were not available by then, which is one of the motivations of the topic in this thesis.

¹⁵ The time is chosen intended to motivate the topic of this thesis.

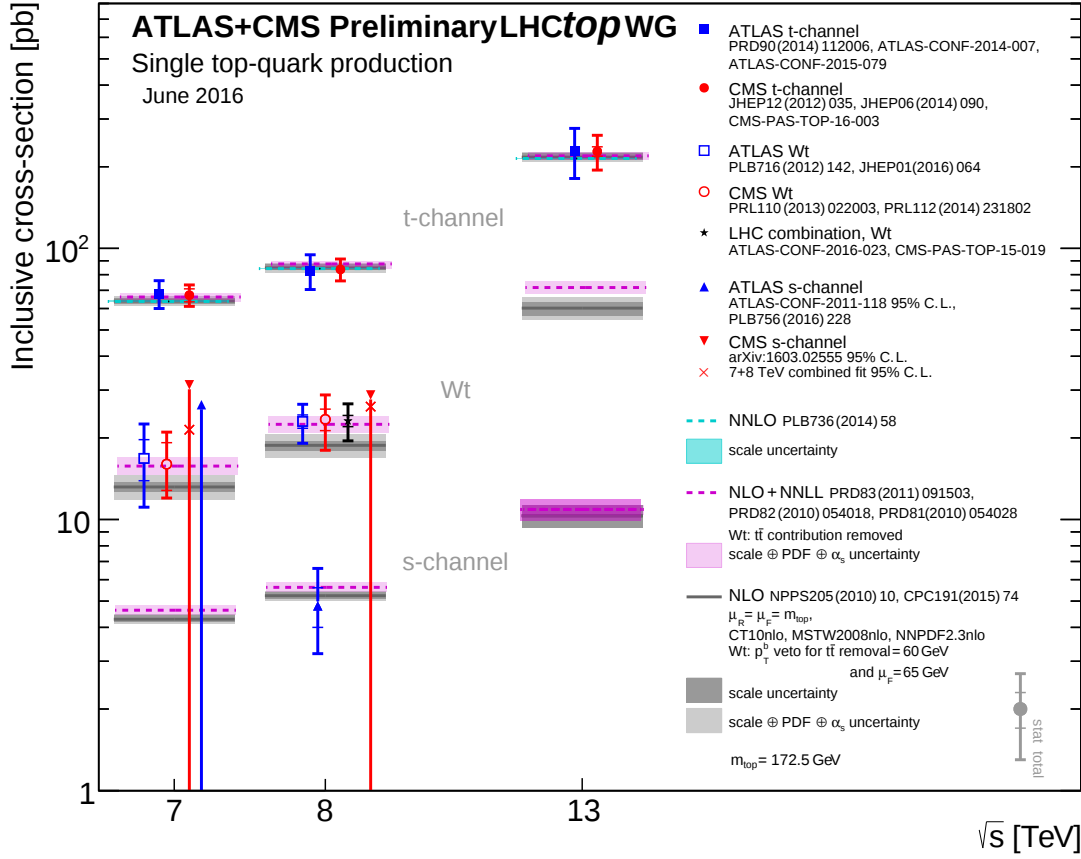


Figure 2.7: Theoretical predictions and experimental measurements of the single top-quark production cross-sections as a function of the centre-of-mass energy. Theory bands and uncertainties are generated at NLO with NNLL soft gluon resummation for $m_t = 172.5$ GeV (except for t -channel the calculation is at NNLO with NNLL). ATLAS and CMS are two of the experiments at the LHC and will be introduced in Chapter 3. Figure is from LHC top-physics working group, and updated in June 2016.

2.4.5 Decay of the top quark

In the SM, the top quark decays into a “down-type” quark and a W boson. According to the CKM matrix, the decay mode is nearly exclusively to a bottom quark and a W as the $|V_{tb}|$ is close to 1. Hence, it reduces the complexity of the final states and increases the efficiency of searching for a top quark. The W has several decay modes via the electroweak interaction. It can decay leptonically to a charged lepton with its corresponding neutrino, or hadronically to different flavours of quarks. Figure 2.8 shows the full chain of a top-quark decay.

For the production channels that have two W bosons in the decay chain, such as the tW or $t\bar{t}$ productions, there are three combinations of the decays: all-hadron channel, where both W bosons decay hadronically, corresponding to a branching fraction $\mathcal{B} = 45.3\%$; single-lepton channel, where one of the W bosons decay hadronically and the other leptonically¹⁶, corresponding to $\mathcal{B} = 44.0\%$;

¹⁶ Here $\ell = e, \mu, \tau$.

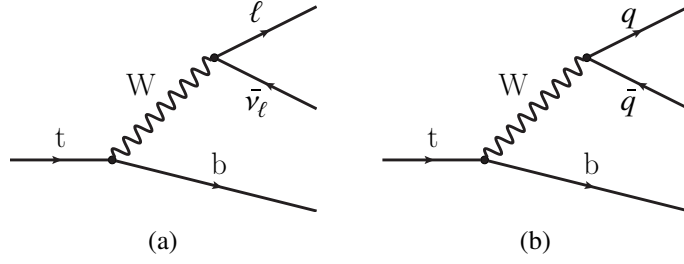


Figure 2.8: LO Feynman diagrams of the top-quark decays where the W decays leptonically (a) or hadronically (b).

and the dilepton channel, where both W bosons decay leptonically, corresponding to $\mathcal{B} = 10.7\%$ ¹⁷.

Despite the lowest branching fraction, dilepton channel gives the cleanest signature. Because of two neutrinos in the final state, there is certain amount of missing transverse momentum, which is also a distinct topology. However, to measure the kinematics of the top quark or the W boson, the kinematics of the two neutrinos need to be properly assigned. This can be done in $t\bar{t}$ production but not tW production, as the latter is an under-constrained system.¹⁸ The next cleanest channel is the single-lepton channel, with relatively high branching fraction (comparable to the all-hadron channel). The all-hadron channel is usually considered last for analysis because it is challenging to separate signal from background. This thesis will focus on the dilepton decay channel of tW production.

2.5 Interference between tW and $t\bar{t}$ productions

The theoretical calculation of the tWb final state is non-trivial. The tW production process, known as singly resonant, interferes with $t\bar{t}$ production, so-called doubly resonant, see Fig. 2.9. The total amplitude, $\mathcal{A}$, is composed of the contribution from all singly resonant diagrams, $\mathcal{A}_{tW}$, and the contribution from all doubly resonant diagrams, $\mathcal{A}_{t\bar{t}}$:

$$\mathcal{A} = \mathcal{A}_{tW} + \mathcal{A}_{t\bar{t}},$$

The calculation of the cross-section involves the modulus square of the total amplitude:

$$\begin{aligned} \sigma_{tWb} &\propto |\mathcal{A}|^2 = |\mathcal{A}_{tW}|^2 + 2\Re(\mathcal{A}_{tW}\mathcal{A}_{t\bar{t}}^\dagger) + |\mathcal{A}_{t\bar{t}}|^2 \\ &= \mathcal{S} + \mathcal{I} + \mathcal{D}. \end{aligned}$$

In order to remove the $t\bar{t}$ contribution from the calculation, two treatments were developed [75]. The first one, called diagram removal (DR), removes the $t\bar{t}$ contribution at the amplitude level, leaving only the tW contribution $\mathcal{S}$. However, this treatment violates gauge invariance as it removes contributions all over the phase space. The second one, called diagram subtraction (DS), subtracts a gauge-invariant

¹⁷ In experiment, since τ^- can decay leptonically or hadronically, the branching fraction of the dilepton final status is further suppressed to be about $\mathcal{B} = 6.4\%$.

¹⁸ Both of them have the same number of degrees of freedom. However, an additional top-quark mass constraint in $t\bar{t}$ production makes the system iso-constrained.

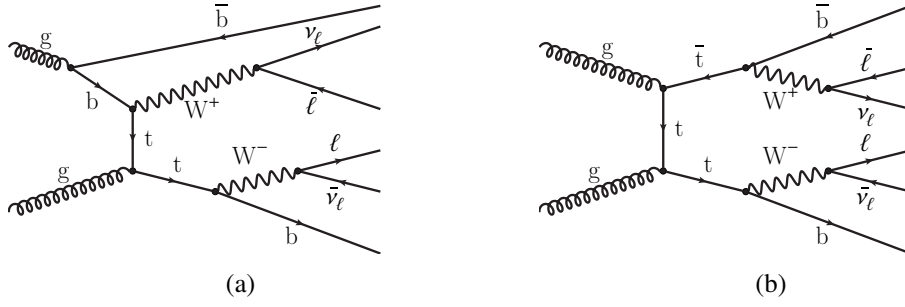


Figure 2.9: Representative Feynman diagrams of tW channel (a) and $t\bar{t}$ production with top-quark decay (b) with the same final states.

term $\mathcal{D}'$, such that the difference $\delta = \mathcal{D} - \mathcal{D}'$ is as close to zero as possible.¹⁹ So the DS scheme ends up with $\mathcal{S} + \mathcal{I} + \delta$.

As the $t\bar{t}$ production has a much larger cross-section than the tW channel, the interference is non-negligible; it turns out that the interference term amounts to about 10 % of the total cross-section [76]. The question arises how well the tW production can be treated as a separate process. This has been discussed in the literature [76–79]. One general idea is to access tW in some particular phase space where the interference is small, the strategy this thesis uses. For instance, if the invariant mass of the final state Wb is far from the top-quark mass, the $t\bar{t}$ process is suppressed. In this sense, the tW channel can be regarded as a well-defined channel, given that proper selections are imposed.

A recent study proposes another scheme, called DR2, which removes the $t\bar{t}$ contribution $\mathcal{D}$ at the cross-section level, leading to $\mathcal{S} + \mathcal{I}$ [80]. Although DR2 is also gauge dependent, the effects of the gauge dependence have been found to be negligible. So it provides an alternative approach to this problem. There is also an inclusive calculation of the cross-section for the $W^+W^-b\bar{b}$ final states at NLO [81]. In this treatment, instead of trying to separate tW from $t\bar{t}$, one can identify the final states regardless of which process they are. However, this is not as trivial as it seems to be. Lots of care has to be taken to avoid double counting from the NLO corrections and PS (non-)emissions from the intermediate top quarks. Moreover, the non-resonant process, bottom quark associated $\ell\bar{\ell}\nu\bar{\nu}$ production, has to be included in the calculation, too.

¹⁹ The difference is non-zero if the gauge invariance is preserved.

Particle detection and reconstruction

This chapter discusses the technologies used to conduct the particle physics research in this thesis. The collider and detector used to collect the data analysed are presented in Section 3.1 and Section 3.2, respectively. Reconstruction of the stable particles detected can be found in Section 3.3.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [42] is the highest-energy particle collider ever built. It is a circular superconducting accelerator housed in the former LEP [82] tunnel at CERN¹ [83] near to the city of Geneva, Switzerland.

3.1.1 Accelerator complex

The accelerator complex is a succession of machines accelerating the beam energy progressively. The LHC is the last element of this chain. It mainly collides proton beams but also collides heavy ions.² This thesis will use the data produced in proton–proton collisions.

The protons are obtained by stripping electrons from hydrogen atoms. Then they are successively injected into the Linac 2, the Booster, the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), to reach an energy of 450 GeV. Finally proton beams enter the LHC and are accelerated to 6.5 TeV after 20 min. The beams are kept circulating and colliding for about 24 h under normal operating conditions. More details can be found in Fig. 3.1.

3.1.2 Main ingredients

The LHC has a circumference of 27 km and is installed roughly 100 m underground. Two adjacent parallel beam pipes are installed in the tunnel, and the beams travel in opposite directions. Dipole magnets are used to keep the beams in their nearly circular orbits. Quadrupole magnets focus the beam and bring them to collide at four different points. Magnets of higher multipole orders (such as sextupoles, octupoles, decapoles, etc.) correct smaller imperfections in the field geometry. About

¹ The European Organization for Nuclear Research; name derived from Conseil Européen pour la Recherche Nucléaire.

² For a few weeks, proton–lead collisions were performed in 2013 and 2016; lead–lead collisions took place in 2010, 2011, and 2015. A short run of xenon–xenon collisions took place in 2017 [84–87].

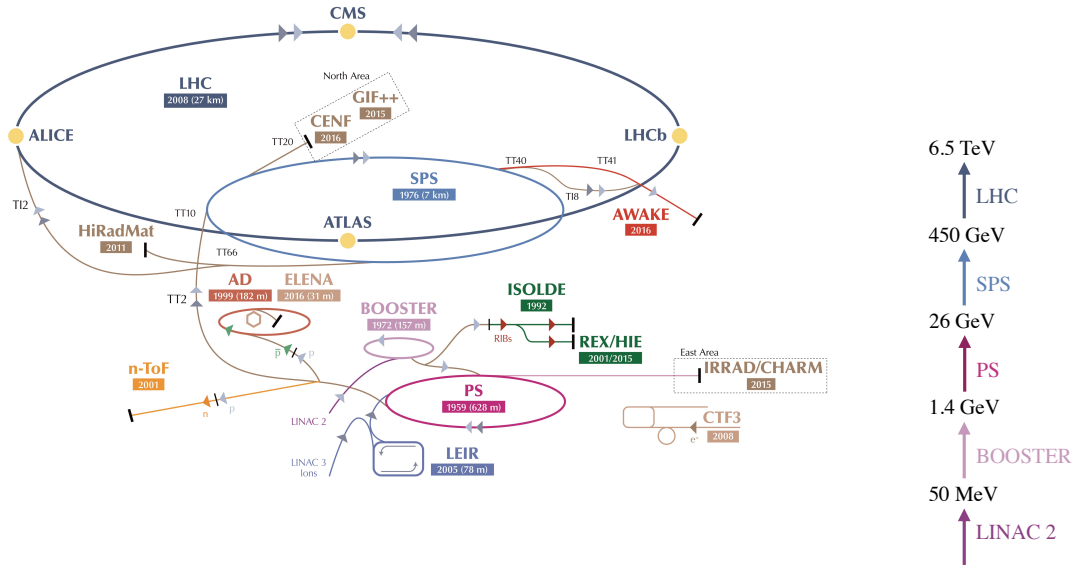


Figure 3.1: Schematic of the CERN accelerator complex including the LHC (left) and the acceleration chain (right). Protons are injected from Linac 2 to Booster. Then PS and SPS accelerate them sequentially before they enter the LHC. For each accelerator, the year of start and the circumference/size are shown. Figure is adapted from [88].

9 600 magnets in total are used. To maximise the number of collisions, i.e. to ensure high luminosity, the LHC achieves 2 808 bunches per beam. They are separated by 25 ns, and each contains about 10^{11} protons.

3.1.3 Experiments at the LHC

Detectors are installed at the four collision points of the LHC ring, capturing the particles generated in the collisions. The two biggest experiments, A Toroidal LHC ApparatuS (ATLAS) [89] and CMS (Compact Muon Solenoid) [90], are general-purpose detectors. They adopt forward-backward symmetric cylindrical geometry to cover nearly the full solid angle. Having two independent experiments allow scientists to cross-confirm any discovery at the LHC. The other two detectors are for specific purposes. LHCb (LHC beauty) [91] aims at studying the CP violation and searching for rare B meson decays. It adopts an asymmetric geometry, as B mesons are emitted mostly at small angles to the beams due to their relatively small mass compared with the high collision energy. ALICE (A Large Ion Collider Experiment) [92] studies the quark–gluon plasma, a state of matter where quarks and gluons are free, which is essential to understand the mechanism of colour confinement and hadron formation.

In addition, three minor experiments are installed for different physics goals. The TOTEM (TOTAl Elastic and diffractive cross section Measurement) experiment [93], aiming at the total cross-section, elastic scattering, and diffractive processes of the proton–proton collisions. The LHCf (LHC forward) experiment [94], studies particles generated in the forward region of collisions. With its unique

location, particles at nearly zero degrees to the proton beam pipe direction can be captured and measured. The MoEDAL (the Monopole and Exotics Detector at the LHC) experiment [95] searches directly for the magnetic monopole and other exotic particles.

3.1.4 Operation energies and integrated luminosities

At the LHC, physics operation started in November 2009 and achieved 1.18 TeV per beam [96], beating the record of 0.98 TeV per beam at the Tevatron. In 2010, the LHC was able to reach 3.5 TeV per beam [97], producing a centre-of-mass energy, $\sqrt{s}$, of 7 TeV. It delivered 6.1 fb^{-1} at $\sqrt{s} = 7 \text{ TeV}$ in 2011 and 23.1 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ in 2012 [98]. The LHC was shut down in February 2013 for its two-year technical stop [99], preparing to double the collision energy of Run 1. On 3 June 2015, the LHC restarted delivering physics data with $\sqrt{s} = 13 \text{ TeV}$ [100], known as Run 2. In 2016 and 2017, the machine focused on increasing the luminosity and reached its design value of $1.0 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ on 29 June 2016 [101]. The maximal instantaneous luminosity is $2.05 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, which was reached on 2 November 2017 [102]. Run 2 was stopped on 10 December 2018 [103], and will be followed by a second two-year technical stop in 2019–2020 [86].

Thanks to the successful and stable operation of the LHC, the ATLAS experiment, like all the other experiments, has accumulated a vast amount of data. Figure 3.2 depicts the data that LHC delivered and ATLAS recorded in the past years.

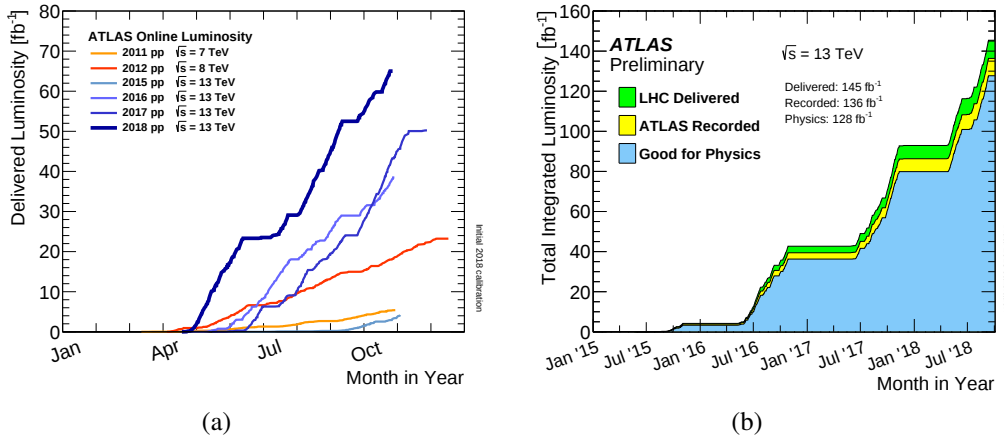


Figure 3.2: Integrated luminosity in ATLAS during stable beams for different data-taking periods (2011–2018) (a). Integrated luminosity delivered by the LHC (green), recorded by ATLAS (yellow), and certified quality data for physics analytics (blue) at $\sqrt{s} = 13 \text{ TeV}$ from 2015 to 2018 (b). The plateaus correspond to so-called technical stops of the LHC. Plots are taken from ATLAS public website [104].

3.2 The ATLAS experiment

The ATLAS experiment [89, 105] is the largest detector ever built for a particle collider. It has a cylindrical shape, 46 m long, 25 m in diameter, and a weight of 7 000 t. It is a general-purpose particle physics experiment, designed to capture a huge range of physics opportunities that can happen at the LHC. For this purpose, ATLAS consists of several subdetectors stacked from inside to outside, to

record the trajectory, momentum, and energy of particles. Such information together allows particle reconstruction and identification. A huge magnet system bends the trajectories of the charged particles so that their momenta can be measured with a high precision. Particles emerging from the collisions transverse to the beam directions are detected in the barrel part while those close to the beam are measured in so-called endcaps. The four major components of the ATLAS detector are the tracker, the calorimeter, the muon spectrometer, and the magnet system. Figure 3.3 depicts a schematic view of the ATLAS detector and its components.

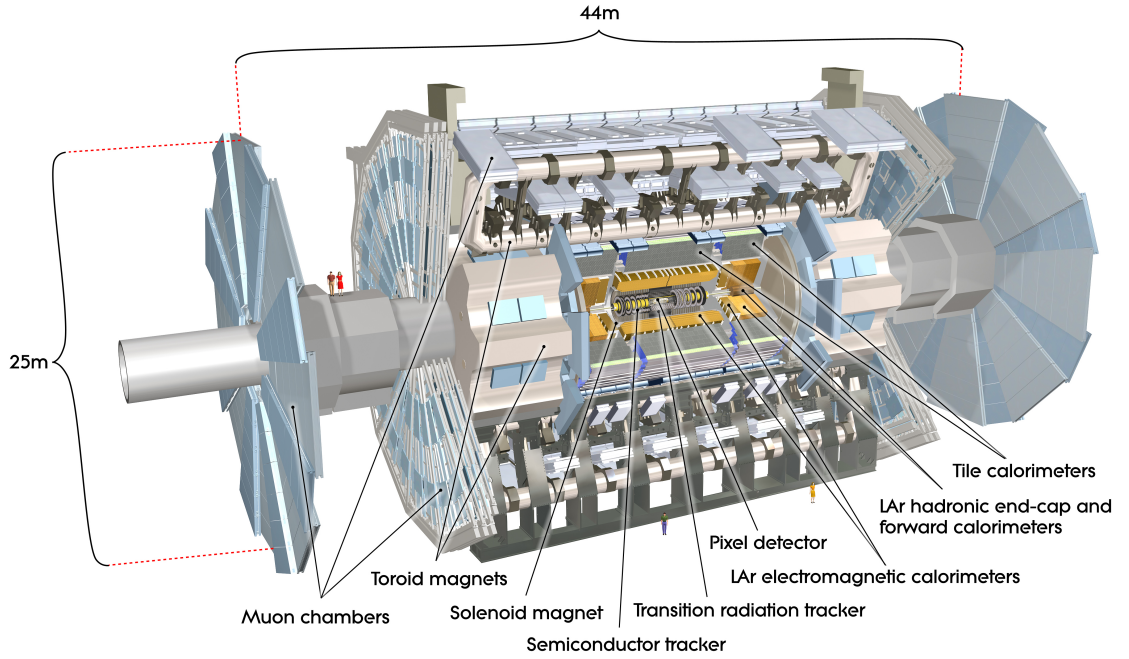


Figure 3.3: Schematic of the ATLAS detector. All subdetectors are included. From inside to outside they are the inner tracker, calorimeter, solenoid magnets, and muon spectrometer in an onion-like structure. A few human models are placed for size comparison. Figure is taken from [106].

Coordinate system

A right-handed coordinate system is used in ATLAS. The origin is at the nominal interaction point (IP) in the centre of the detector and the z -axis along the beam pipe. The x -axis points from the IP to the centre of the LHC ring, and the y -axis points upward. Cylindrical coordinates (r, ϕ) are also used in the transverse plane. The azimuthal angle, ϕ , is measured around the z axis, and the polar angle, θ , is the angle to the z axis.

3.2.1 The inner detector

The inner detector (ID) is the innermost part of the ATLAS detector used to reconstruct the primary and secondary vertices, and charged³ particle tracks with high efficiencies over the pseudorapidity range of $|\eta| < 2.5$. It consists of three sub-components all present in a magnetic field parallel to the beam axis: pixel detector (Pixel), semiconductor tracker (SCT), and transition radiation tracker (TRT), each of which is comprised of a barrel and two endcaps (see Fig. 3.4). The main parameters of the subcomponents are summarised in Table 3.1.

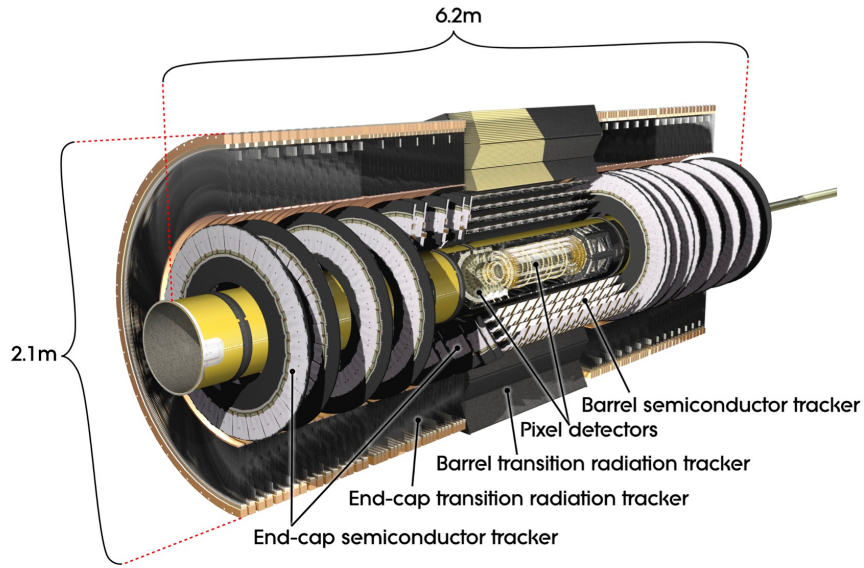


Figure 3.4: Cutaway view of the ATLAS ID. From inside to outside it consists of the Pixel, the SCT, and the TRT. Figure is taken from [107].

Sub-component	Technique	Sensor size [$\mu\text{m} \times \mu\text{m}$]	Resolution [$\mu\text{m} \times \mu\text{m}$]	$ \eta $	r [mm]	z [mm]
IBL	Silicon pixel	50×250	8×40	<2.5	31–242	$<3\,092$
Pixel (no IBL)	Silicon pixel	50×400	10×115	<2.5		
SCT	Silicon microstrip	80	17×580	<2.5	251–610	$<2\,797$
TRT	Gaseous straw tube	4 000	130	<2.0	554–1 166	$<2\,744$

Table 3.1: Summary of the main parameters of the ID subdetectors [108]. The resolution of the IBL, the Pixel, and the SCT is reported along ϕ and z (r if endcaps) and for TRT along ϕ (z if endcaps). For SCT and TRT the sensor sizes refer to the spacing of the readout strips and the diameter of the straw tubes, respectively.

³ Hereinafter “charge” implies electric charge and colour charge will be mentioned explicitly.

The Pixel detector

The Pixel [105, 109] consists of four layers of concentric cylinders (one layer from the Insertable B-Layer (IBL)⁴ [110]) enclosing the beam pipe in the barrel section and three layers of disks in endcap. It covers a pseudorapidity region of $|\eta| < 2.5$. Silicon pixel sensors are used with the size of $50 \times 250 \mu\text{m}^2$ for the IBL and $50 \times 400 \mu\text{m}^2$ for the other parts. The intrinsic spatial resolution in ϕ - z (ϕ - r if endcaps) is $8 \times 40 \mu\text{m}^2$ for the IBL and $10 \times 115 \mu\text{m}^2$ for the rest. Overall, it occupies the space of $31 \text{ mm} < r < 242 \text{ mm}$ and $|z| < 3092 \text{ mm}$. The exact sizes and positions of all layers and disks without IBL can be found in Fig. 3.5.

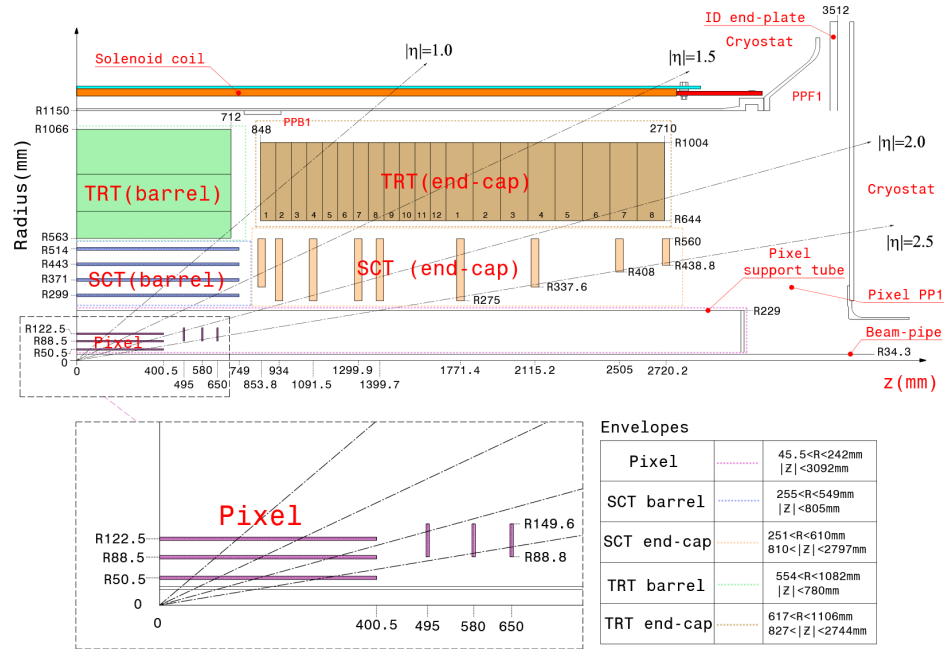


Figure 3.5: A quarter-section of the ATLAS ID. Sizes and positions for each barrel layer and endcap disc of each sub-component can be read off. Bottom left shows more details on Pixel, yet the IBL is not displayed. The envelopes of each sub-component are summarised in the table on the bottom right. Figure is taken from [111].

The SCT

The SCT [105, 109] is a silicon microstrip tracker consisting of four cylindrical barrel layers and nine planar discs in each endcap. The sizes and positions of the components have been optimised so that any track typically crosses four SCT layers, independently of its pseudorapidity. Both the barrel and the endcaps are build by modules, which are formed mostly by four microstrip sensors. Two microstrip sensors with each a size of $6.36 \times 6.40 \text{ cm}^2$ are wire-bonded together to form a 12.8 cm long strip, and two such strips are glued together back-to-back at a 40 mrad stereo angle, in order to provide a required z (barrel) or r (endcaps) measurements. The SCT achieves an intrinsic spatial resolution in ϕ - z (ϕ - r if endcaps) of $17 \times 580 \mu\text{m}^2$. The envelope of the SCT is $251 \text{ mm} < r < 610 \text{ mm}$ and $|z| < 2797 \text{ mm}$.

⁴ The IBL was installed during the ATLAS upgrade from 2013 to 2015, to improve the tracking and flavour tagging performance.

The TRT

The TRT [105, 109] is a gaseous detector, made of 4 mm diameter straw tubes, each acting as a single drift chamber. Radiators are placed between the straws to increase the transition radiation of particles crossing the boundary between two materials. The barrel contains about 50 000 straws grouped into three rings. In the endcaps, straws are oriented radially and arranged in 20 wheels on each side, with eight layers in each wheel. The straws are filled with a gas mixture of 70 % base component, Xe, 27 % neutral quencher, CO₂, and 3 % electronegative gas, O₂. The TRT not only improves the momentum resolution of tracks but also provides a particle identification due to the transition radiation. In total, a track crosses approximately 30 straws, with a maximum of 36, over a distance of about half a meter. The intrinsic resolution is 130 μm in the r direction within $|\eta| < 2$. It encloses the space of $554 \text{ mm} < r < 1\,106 \text{ mm}$ and $|z| < 2\,744 \text{ mm}$.

3.2.2 The calorimeter

The calorimeter system absorbs the energy of incident charged and neutral particles and measures the energy deposit over a large pseudorapidity range of $|\eta| < 4.9$. Different types of calorimeters complement each other, focusing on different particle-matter interactions. Showers are produced by a cascade of secondary particles, giving an energy measurement and particle identification. The electromagnetic calorimeter (EMC) precisely measures the energy of particles that interact primarily via the electromagnetic force, such as electrons and photons. The hadron calorimeter (HCal) measures the energy of hadrons that interact strongly. Besides the EMC and the HCal, ATLAS calorimetry also includes the forward calorimeter (FCal) to cover the forward region. The total weight of the calorimeter system is about 4 000 t, more than half of the ATLAS weight. An overview of the calorimeter system can be found in Fig. 3.6.

The EMC

The EMC is a liquid-argon (LAr) sampling detector using lead as an absorber material. With an accordion geometry, it provides a full azimuthal coverage until $|\eta| < 3.2$ without any cracks. It consists of a barrel part ($|\eta| < 1.475$) and two endcaps ($1.375 < |\eta| < 3.2$) with some overlap. The barrel is further divided into two halves at $z = 0$ for necessary electronics and cabling. Each endcap calorimeter is mechanically composed of two coaxial wheels. In the region $|\eta| < 1.8$, a presampler is installed to correct for the energy loss in the material upstream of the calorimeter. Three (two if without presampler) samplings layers are designed to have different granularities. Finer granularity is in the inner layer to have excellent resolution while coarser in outer layer to detect leakage. The total thickness of the EMC is above 24 radiation lengths (X_0) in the barrel and above 26 X_0 in the endcaps.

The HCal

The HCal uses a plastic-scintillator-tiles technique with iron absorber in the barrel ($|\eta| < 1.0$) and two moveable extended barrel parts ($0.8 < |\eta| < 1.7$). The endcaps are LAr calorimeters with copper as absorber, covering $1.5 < |\eta| < 3.2$. Some overlaps are left with both the tile and the FCal. The three-layered FCal is placed radially inside the HCal endcaps in the range of $3.1 < |\eta| < 4.9$. The first layer that uses copper absorber is optimised for electromagnetic shower measurements, while the other two layers using tungsten focus on hadronic shower measurements. The total thickness of the HCal

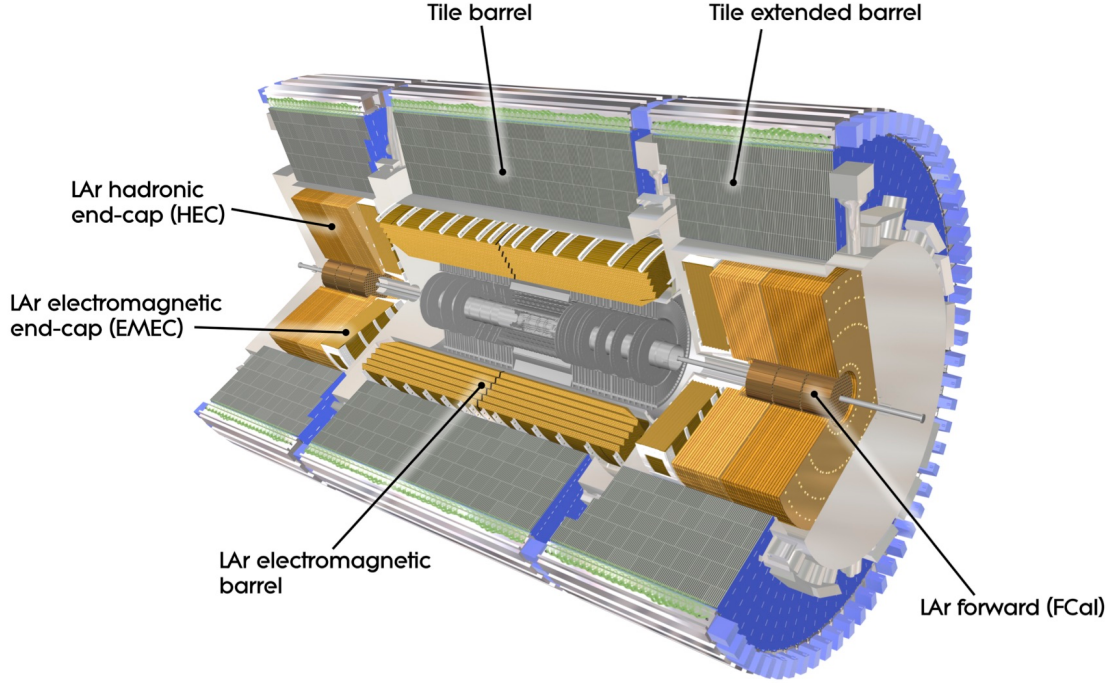


Figure 3.6: Cutaway view of the ATLAS calorimeter system. From inside to outside it consists of the EMC and the HCal. The FCal is in the very forward and backward regions close to the beam pipe. Figure is taken from [112].

is 11 absorption length⁵ at $\eta = 0$, sufficient to reduce the punch-through for the muon spectrometer. Figure 3.7 shows the amount of material in the calorimeter as a function of η .

A comparison of the different parts of the ATLAS calorimeter system is made in Table 3.2.

3.2.3 The muon spectrometer

The outer part of the ATLAS detector is the muon spectrometer (MS) [114]. It is designed to identify muons that pass through the detector and measure their momenta. It comprises trigger chambers for a bunch-crossing identification and high-precision chambers for a momentum measurement. The trigger chambers use fast time response: resistive plate chambers (RPCs) in the barrel and thin gap chambers (TGCs) in the endcaps, covering $|\eta| < 2.4$. The high-precision chambers employ technologies with good spatial resolution: monitored drift tubes (MDTs) for the majority and resistive plate chambers (CSCs) for the innermost part of the innermost endcap disks, covering $|\eta| < 2.7$. Figure 3.8 shows a sketch of the MS with the four different sets of chambers.

⁵ The distance after which $1/e$ of the incident particles survive.

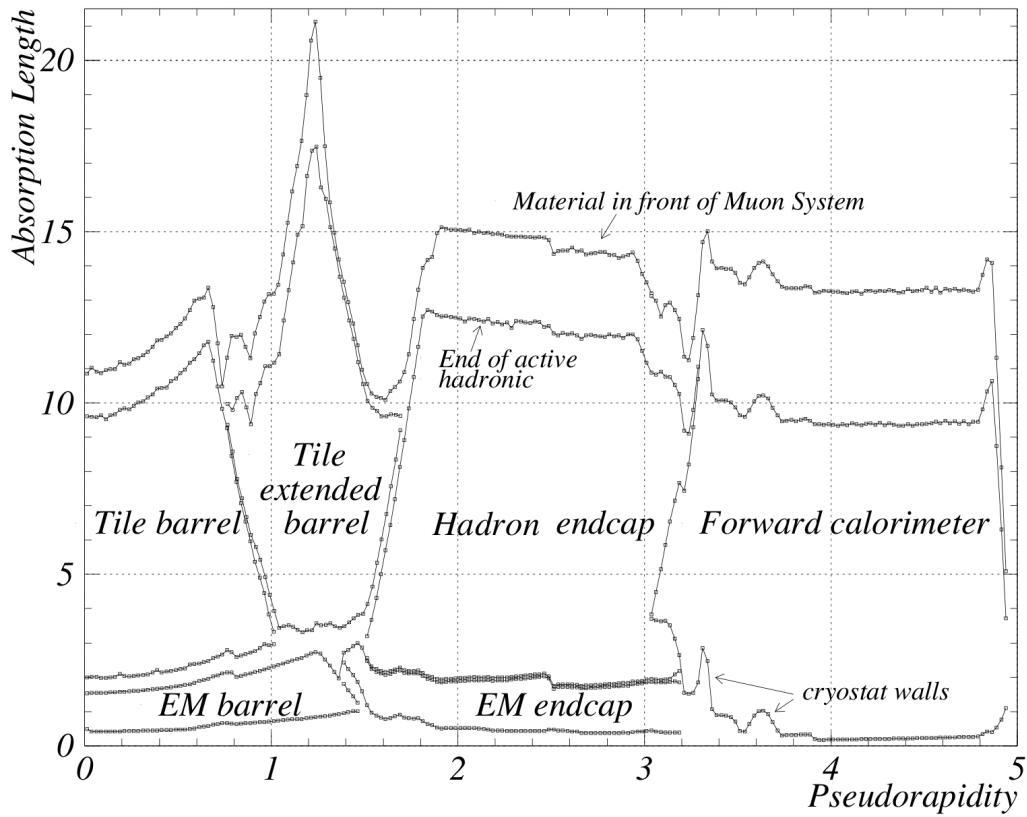


Figure 3.7: Stacked distribution of the material budget for individual parts of the calorimeter system in the unit of absorption length as a function of pseudorapidity η . Figure is taken from [113].

Calorimeter system	Active medium	Absorber	$ \eta $	Number of layers
EMC barrel	LAr	Lead	<1.475	3
EMC endcaps			1.375–1.8	3
			1.8–3.2	2
Tile barrel	Plastic scintillator	Iron	<1.0	3
Tile extended barrels			0.8–1.7	3
HCal endcaps	LAr	Copper	1.5–3.2	3
FCal		Copper	3.0–4.9	1
		Tungsten	3.1–4.9	2

Table 3.2: The active and absorber material used in different calorimeter parts, as well as some geometrical parameters.

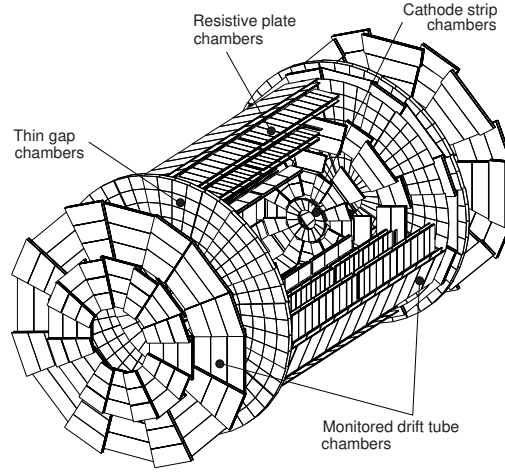


Figure 3.8: A sketch of the ATLAS MS. Four technologies are employed. Figure is taken from [114].

The high-precision chambers

The MDT, a gaseous tube-style detector, is used for precision tracking of muons throughout all of the MS. The barrel ($|\eta| < 1.3$) consists of three concentric layers at radii of approximately 5, 7.5, and 10 m, each of which embraces 16 small or large sectors overlapping each other to avoid holes in acceptance. The endcaps ($1.3 < |\eta| < 2.4$) have three wheels on each side, placed at $z = 7.5, 14$, and 22.5 m, and sectors overlap to achieve a full coverage. The single tube resolution is about $80 \mu\text{m}$, leading to a full tracking resolution of about $35 \mu\text{m}$.

The CSCs, multi-wire proportional chambers, are deployed in the place near the beam pipe of the endcap innermost layer, taking advantage of its higher rate capability and lower neutron sensitivity. They can provide not only a similar ($\sim 40 \mu\text{m}$) resolution as the MDT, but also a very coarse measurement in ϕ . In addition, they have a decent time response ($\sim 7 \text{ ns}$) as the sensitivity for neutrons is low and the drift times are small, although the response is still slower than the trigger chambers.

The trigger chambers

The trigger chambers are optimised to have a good timing response, including contributions from signal propagation and backend electronics. The timing information contributes to the muon level 1 trigger. Moreover, the trigger chambers measure the coordinate both in the bending and non-bending planes, which is also used for the final muon reconstruction. Three layers of the RPCs, wireless gaseous detectors, are installed in the barrel ($|\eta| < 1.05$), achieving a temporal resolution of about 1.5 ns. In the endcaps ($1.05 < |\eta| < 2.4$), multi-wire proportional chambers, TGCs, are placed to provide a temporal resolution of 4 ns.

The parameters of the four technologies in the MS are summarised in Table 3.3.

Technology	Function	Resolution			$ \eta $	r	z
		z/r	ϕ [mm]	t [ns]			
MDT	Tracking	35 μm	—	—	<2.4	5, 7.5, 10	7.5, 14, 22.5
CSC	Tracking	40 μm	5	7	2.0–2.7	—	5
RPC	Trigger	10 mm	10	1.5	<1.05	8, 8, 10	—
TGC	Trigger	2–6 mm	3–7	4	1.05–2.4	—	~ 14

Table 3.3: Summary of the main parameters of the MS subdetectors. Resolution shows the track level resolution. The z/r resolution shows the resolution in z in the barrel and in r in the endcaps. Distinct values in “ r ” and “ z ” columns represent the positions of different layers. “—” indicates unavailable items.

3.2.4 The magnet system

The ATLAS detector is equipped with a superconducting magnet system [115] that bends the charged tracks to measure the momentum with great precision and determine the charge sign. It consists of a central solenoid [116] surrounding the ID and three large toroids (one barrel and two endcaps) [117, 118] as a part of the MS (see Fig. 3.9).

The solenoid has the shape of a cylinder with a diameter of 2.4 m and a length of 5.3 m. A homogeneous magnetic field of 2 T is generated longitudinally, bending the charged particles in the transverse plane. The solenoid coil and its supporting structure must be as transparent as possible to particles. In fact, the solenoid shares the same cryostat with the LAr barrel calorimeter to minimise the material. The energy loss due to the solenoid and other material is corrected by the presampler as mentioned before.

The toroidal magnet is composed of a barrel and two endcaps, providing an azimuthal 4 T magnetic field, perpendicular to the solenoid field. The barrel toroid is 25.3 m long with a 20.1 m outer diameter, while each of the endcap toroids has an axial length of 5.0 m and a outer diameter of 10.7 m. Each of the three toroids consists of eight coils symmetrically arranged around the beam pipe.

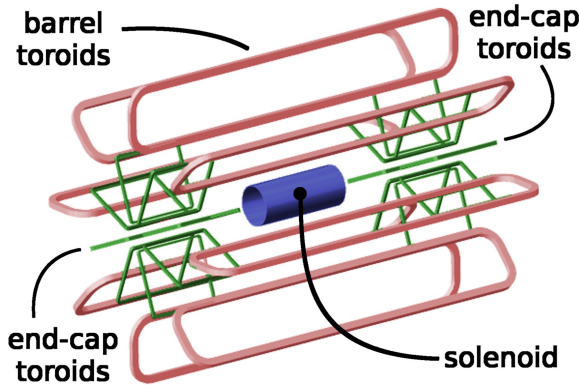


Figure 3.9: A sketch of the ATLAS magnet system. It comprises a central solenoid [119] providing a 2 T magnetic field for the ID, and three toroids generating a 4 T magnetic field for the MS.

3.2.5 The trigger system

With a 25 ns bunch spacing of the LHC, the bunch-crossing rate in the ATLAS detector is around 40 MHz. At this rate, it is impossible to record every event. Therefore, a trigger system [120] is needed to select and record useful events. A substantial change in the trigger system has been made for Run 2 to deal with the higher rate stemming from the higher centre-of-mass energy and higher luminosity. The Run 2 trigger [121] consists of two steps: the level-1 (L1) trigger and the high-level trigger (HLT). The L1 trigger is hardware based, determining Regions-of-Interest in the detector based on coarse calorimeter and muon information with a latency of less than 2.5 μ s, while the HLT is software based, executing fast reconstruction algorithms in either the Regions-of-Interest or the whole event. The total event rate is reduced from 40 MHz to 100 kHz by the L1 trigger, and further down to 1 kHz by the HLT.

The L1 trigger consists of the L1 calorimeter trigger system (L1Calo), the L1 muon trigger system (L1Muon), the L1 topological trigger modules (L1Topo), and the central trigger processors (CTP). As the names suggested, the L1Calo processes information from the EMC and HCal and the L1Muon exploits the responses from the MS trigger chambers. The L1Topo then takes the information from the both and quickly calculates event topological quantities, for instance, the angular distance, the invariant mass, to increase the probabilities of hard collision identification. The CTP makes the final L1 trigger decision based on the information from the three. If the L1 trigger accepts the event, the CTP informs the subdetector readout systems and the fragment of the event passes to the HLT with an L1 trigger summary.

The HLT runs fast reconstruction algorithms to identify electrons, muons, tauons, jets, missing transverse momentum or other objects that are interested in physics analyses, and makes some basic requirements on their properties. Events passing the HLT trigger are written to permanent storage to be further revisited.

3.3 Object reconstruction

Particles interact with the detector in different ways. Each particle has a specific signature, illustrated in Fig. 3.10. The offline reconstruction algorithms make use of all subdetector information to reconstruct and identify these objects.

3.3.1 Charged tracks and vertices

Tracks are reconstructed trajectories of the charged particles that pass through the ID or MS. They can be assumed as helices, as they are bent by uniform, axial magnetic fields, and parametrised in terms of perigee parameters at the closest approach of the reconstructed track to a reference point in the transverse plane:

- d_0 , transverse impact parameter, signed distance to the z axis,
- z_0 , longitudinal impact parameter, z -component,
- ϕ_0 , azimuthal angle,
- θ , polar angle,
- Q/p_T , inverse transverse momentum with charge-sign.

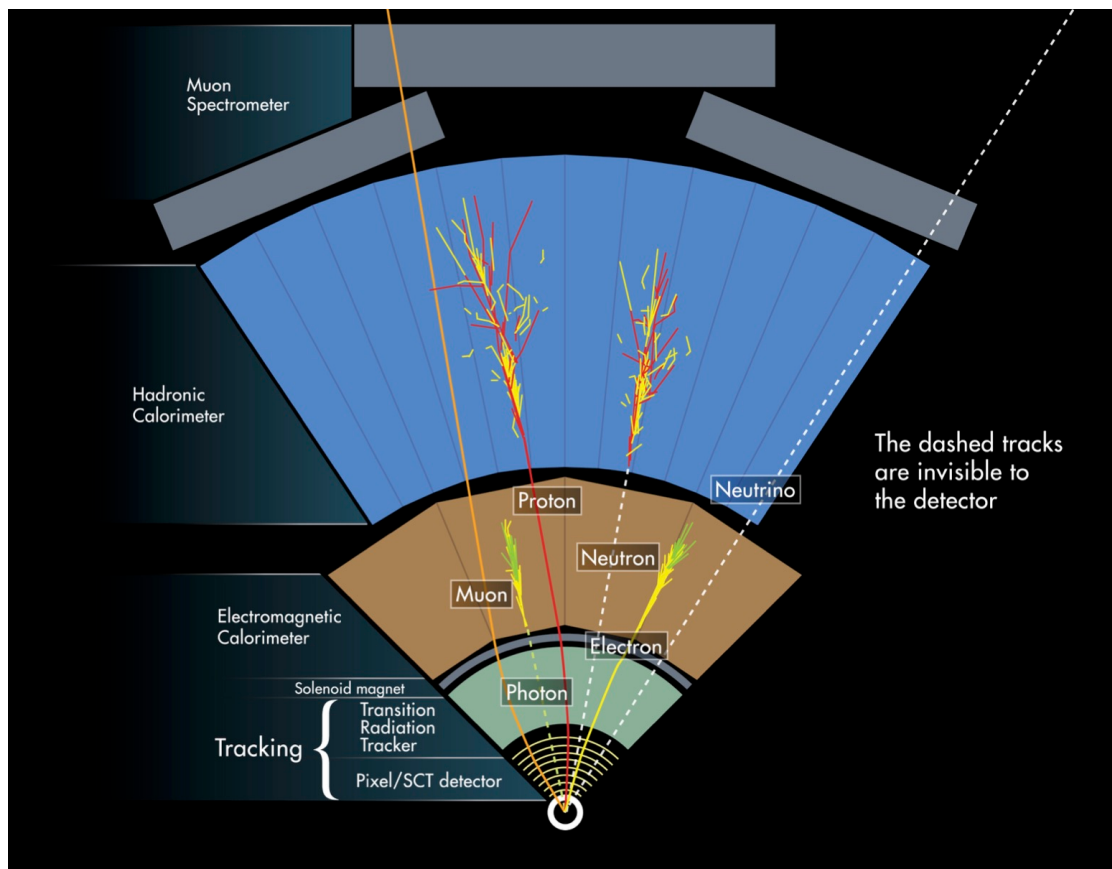


Figure 3.10: Signatures of various particles in the ATLAS detector. Each of them interacts differently with the matter, leaving signals in various subdetectors. Figure is taken from [122].

Tracking

The reconstruction of tracks, or tracking, consists of two stages [123, 124].⁶ The first stage is inside-out. Information in the Pixel and the SCT subdetectors are converted to three dimensional space points, called *hits*. Starting with a seed formed by three hits, hits in other layers are iteratively added to the track candidate. This procedure is done by a combinatorial Kalman filter algorithm, which uses the hits information (positions and covariances) to predict the track representation on the next layer. A massive number of track candidates are created in this way. The candidates are ranked by their *score* evaluated from the number of hits, χ^2 of the track fit, etc. Sharing hits between tracks is allowed if dedicated requirements are satisfied. The remaining tracks passing the requirements are prolonged to the TRT. With all the hits including the TRT, a refit is made on the track and the score is updated. If the newer score is lower, the silicon track is kept, and the TRT hits are flagged as outliers. This completes the inside-out sequence of reconstruction. To further recover ambiguities and reconstruct tracks from secondary vertices or photon conversions, which may not leave a seed in the inner silicon detectors, a second stage, called outside-in sequence, is applied. In this stage, segments are reconstructed by the

⁶ Only ID tracks are discussed here as these tracks are common for all charged particles. MS tracks will be described in Section 3.3.4.

unused TRT drift circles⁷ and then silicon hits are added to complete the reconstruction.

Vertexing

Closely related to tracking is the reconstruction of vertices, points from which a set of tracks originate. The vertices where a collision happened are called *primary* vertices. They are reconstructed using a vertex finding algorithm [124, 125] that iteratively gathers tracks for a vertex candidate. During the iteration, the vertex position is updated by weighting down the less compatible tracks. After the last iteration, a vertex candidate is created based on the final weights. A valid primary vertex candidate must have at least two associated tracks. When no associated tracks are left, or no further vertex can be reconstructed from the remaining tracks, the procedure stops. Among all valid primary vertices, the one associated with the hard collision is the hard-scatter vertex, to be distinguished from the PU vertices.

Secondary vertices, where a particle has presumably decayed into two or more charged particles, are reconstructed as well. They can be used for identifying heavy flavour jets [126]. Starting with all two-track vertices, uninteresting vertices from photon conversions, detector material interactions, etc., are removed. Then the remained vertices are merged, and only a single one associated with the most tracks is kept per jet.

3.3.2 Calorimeter clusters

A calorimeter measures the energy deposited by the particles flying into it. The calorimeter cells first need to be grouped to clusters, which are typically assumed to correspond to individual particles that hit the calorimeter. Calorimeter clusters are further calibrated and used to reconstruct electrons, photons, jets, and missing transverse momenta.

Clustering

Two types of clustering algorithms are used for different purposes: the sliding window algorithm and the topological algorithm [127, 128].

The sliding window algorithm defines a rectangular window with a fixed-size, and all cells inside the window form the cluster. The window is placed such that the total transverse energy it contains is a local maximum. This algorithm is mainly used for electron and photon reconstruction. The fact that the cluster size is fixed allows for a precise cluster energy calibration.

The topological algorithm starts with a seed cell with a large signal to noise (electronic and PU related) ratio ($s/b > t_{\text{seed}}$) and iteratively adds its neighbouring cells⁸, if their signal to noise ratio exceed a threshold, $t_{\text{neighbour}}$. Then, the neighbours of the newly included cells are checked with their signal to noise ratio and appended if the ratio surpasses t_{edge} . Seeds are not allowed to be in the presampler layer of the calorimeter, in order to suppress low-energy PU showers. In most physics analyses, the “420” topological cluster is used, where $t_{\text{seed}} = 4$, $t_{\text{neighbour}} = 2$, and $t_{\text{edge}} = 0$. Higher thresholds for seeds and neighbours effectively suppress the noise, while lower threshold at the perimeter ensures tails of showers to be included. When two showers overlap, they might

⁷ The TRT records the time at which the ionisation signal first exceeds the threshold. Later, this time can be translated into a drift circle radius, making use of the relation between drift time and drift distance.

⁸ Neighbouring cells can come from either the same calorimeter layer, or adjacent layers and subsystems.

be constructed as one cluster with two local maxima. If so, they are further resolved by a splitting algorithm. The topological algorithm is appropriate for clusters with a large number of cells and is used for jet and missing transverse momentum reconstruction.

3.3.3 Electrons

Electrons leave charged tracks in the ID and create clusters in the EMC, so the electron reconstruction combines the information of track and cluster. Electron identification and isolation criteria determine whether a candidate is accepted.

Reconstruction and identification

Electron reconstruction starts from the sliding window calorimeter clusters in the region of $|\eta| < 2.47$. These clusters match ID tracks originating from the hard-scatter vertex, based on the angular distance between the cluster barycentre and extrapolated track position to the calorimeter. An algorithm resolves ambiguity matching using the cluster-track distance under different momentum hypotheses and some criteria on hits information. Afterwards, the cluster is enlarged a bit to account for the energies deposited outside of the sliding window. Once the matching is determined, energy and momentum of the electron are computed using the final calibrated cluster energy and the matched track, respectively [129]. The direction of the momentum is obtained from the track direction extrapolated to the hard-scatter vertex.

The electron identification algorithm uses a multivariate technique to identify electrons and to suppress fake electrons⁹. A likelihood value is built from the information of tracks and clusters, including hits in the IBL and the Pixel, the radiation-energy measurement in the TRT, the track quality, shower shape, cluster-track matching, etc. Based on the thresholds of the likelihood, “Loose”, “Medium”, or “Tight” electrons are defined, where the latter is a subset of the former.

The overall efficiency of electron reconstruction and identification is depicted in Fig. 3.11. High efficiencies are achieved at higher E_T due to better performance with reasonable energy deposits. The efficiency drops in the calorimeter transition region between the barrel and the endcaps ($1.37 < |\eta| < 1.52$), where a large amount of material is placed. A drop is also seen in the crack region of the EMC barrel ($z = 0$).

Isolation

Electron isolation further purifies prompt electron candidates by spatially isolating them from other reconstructed objects. Two types of discriminating variables are designed to build the requirements: $E_{T, \text{coneXY}}$ and $p_{T, \text{coneXY}}$. The $E_{T, \text{coneXY}}$ is defined as the sum of E_T of additional topological clusters, within the cone of $\Delta R = 0.XY$ around the candidate electron cluster, while the $p_{T, \text{coneXY}}$ is the sum of p_T of additional tracks originating from the hard-scatter vertex, within a cone of $\Delta R = \min(0.XY, 10 \text{ GeV}/p_T)$ around the candidate electron track. The typical value of XY is 20 for electrons. By posing requirements on $E_{T, \text{coneXY}}/p_T$ and $p_{T, \text{coneXY}}/p_T$, different working points are defined. The criteria vary in electron p_T ranges such that the isolation efficiencies can be constant or as a function of p_T . Table 3.4 summarises the isolation efficiencies for different working points.

⁹ Fake electrons arise from other physical objects misidentification (mostly light-flavour jets), photon conversions, heavy-flavour hadron semileptonic decays, etc.

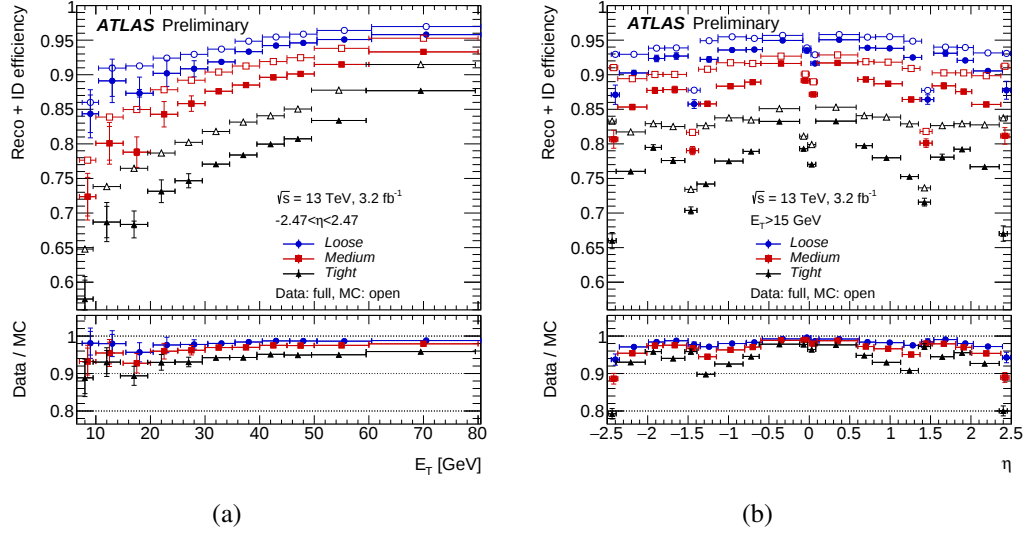


Figure 3.11: Combined electron reconstruction and identification efficiencies with respect to E_T (a) and η (b) from a simulated $Z \rightarrow ee$ sample. Correction factors are applied to the simulated events to compare with data. Data efficiencies are derived from a tag-and-probe method. The uncertainties are obtained with pseudo-experiments. Figures are taken from [130].

Working point	Calorimeter isolation	Efficiency [%]	
		Track isolation	Combined isolation
LooseTrackOnly	—	99	99
Loose	99	99	99
Gradient	$92.14 + 0.1143 \times p_T$	$92.14 + 0.1143 \times p_T$	90 (99) @ 25(60) GeV
GradientLoose	$95.57 + 0.057 \times p_T$	$95.57 + 0.057 \times p_T$	95 (99) @ 25(60) GeV

Table 3.4: Efficiencies for different lepton isolation working points. p_T is in a unit of GeV. The table holds for both electrons and muons.

Calibration

The calibration of the electron energy corrects the electron energy response from the cluster energy [129, 131]. Firstly, the relative energy scales of the different layers of the EMC are corrected. Then a multivariate regression algorithm is used to take into account the effect from the material in front of the calorimeter. The non-uniformity of the detector response arising from non-uniform magnetic fields or non-nominal high-voltage settings in some regions of the calorimeter is also corrected. Afterwards, an in-situ procedure using data $Z \rightarrow ee$ events is applied to correct remaining residual disagreement in the energy scale and resolution.

3.3.4 Muons

Muons are reconstructed from objects that leave tracks in both the ID and MS [132]. The reconstruction algorithms run independently in the two subdetectors, and the information is combined to form the

muon tracks. Identification and isolation algorithms determine whether a candidate is accepted.

Reconstruction and identification

Muon reconstruction in the ID is similar to the electron reconstruction described above. In the MS, Hough transforms are adopted to search for hit patterns to build segments in each muon chamber. Reconstruction starts from the segments in the middle layer, the layer that usually contains more hits. Track candidates are built by extending the segments to the inner and outer layer segments. An MS track must have at least two segments, except in the MS barrel–endcap transition region, where a single high-quality segment is sufficient. Two tracks can share a segment under certain conditions. All hits associated with the candidate are fitted, and the track is accepted if it passes some high-quality checks.

Then a combined muon reconstruction takes into account the information from the ID, MS, and calorimeters. Four different situations are possible [133]. “Combined” muon uses tracks that are independently built in the ID and the MS, and perform a global fit to extract the track parameters. In the “Segment-tagged” scenario, an ID track is extrapolated to the MS and it must match at least one segment there. “Calorimeter-tagged” muon extrapolates an ID track to the calorimeter and accepts it when it matches a minimum-ionising shower. In the “Stand-alone” scenario, the muon trajectory is reconstructed only in the MS and extrapolated to the interaction point. It is mainly used to extend the acceptance to the range $2.5 < |\eta| < 2.7$, which is not covered by the ID. Afterwards, overlaps between different categories are resolved before finalising the muon candidate collection.

The muon identification algorithm identifies prompt muons and suppresses fake muons¹⁰ by requiring particular criteria on the quality of tracks, such as the number of hits, from a global fit of the ID and MS tracks. Some benchmark selections (“Medium”, “Loose”, “Tight”, and “High- p_T ”) are studied in ATLAS and provided for different physics analyses.

The total efficiency of the muon reconstruction and identification is illustrated in Fig. 3.12. High efficiencies, close to 99 %, are observed over the full η range with an average fake rate of 0.2 %, except for the central region ($|\eta| < 0.1$), where the “Medium” muon efficiency is reduced to 60 %.

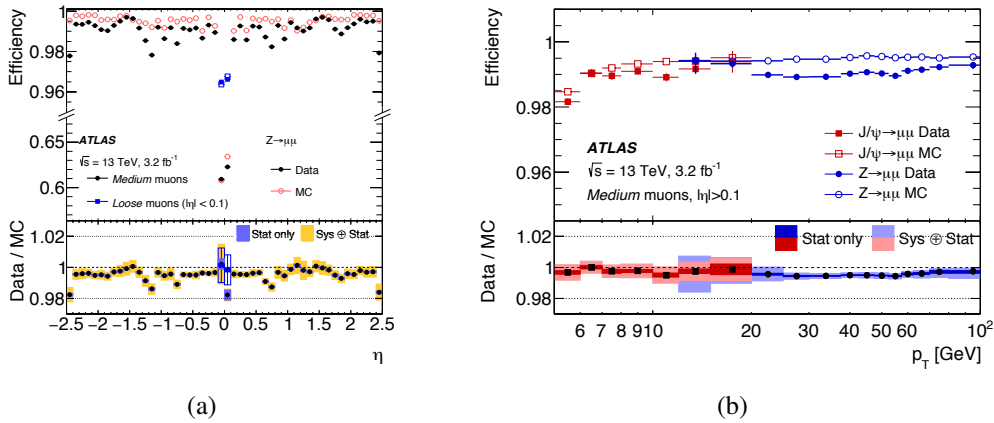


Figure 3.12: Combined muon reconstruction and identification efficiency as a function of η (a) with $p_T > 10$ GeV, and as a function of p_T (b) for the “Medium” muon (and “Loose” on the left) in the region of $0.1 < |\eta| < 2.5$. Figures are taken from [132].

¹⁰ Fake muons arise from falsely reconstructed charged hadrons.

Isolation

Muon isolation distinguishes prompt muons, which are often produced isolated, from those stemming from jet constituents decay. Like electron isolation, a track-based isolation quantity $p_{T, \text{coneXY}}$ and a calorimeter-based isolation quantity $E_{T, \text{coneXY}}$ are available. The values of XY are 20 for $E_{T, \text{coneXY}}$ and 30 for $p_{T, \text{coneXY}}$. Different working points are defined by posing requirements on the quantities. The working points are harmonised to electron isolation working points (Table 3.4).

Calibration

The transverse momenta of muons are calibrated due to the energy loss in the calorimeters. The procedure is done in different η - ϕ bins that are homogeneous regarding detector technology and performance [134]. The calibration constants are extracted by comparing the invariant mass distributions for $J/\psi \rightarrow \mu\mu$ and $Z \rightarrow \mu\mu$ candidates in data and Monte Carlo (MC) simulation.

3.3.5 Jets

Jets are reconstructed as a bunch of particles originating from the hadronisation of a parton. The particles, predominantly $\pi^\pm$, photons from the decays of π^0 , $K^\pm$, deposit large amounts of energy in the calorimeters. Before running the jet algorithm, calorimeter clusters are reconstructed, each of which presents approximately such an object. The jet reconstruction algorithm then groups collimated objects to form a jet and deduces the initial properties of the parton.

Reconstruction

The anti- k_t algorithm [135] is the most common jet algorithm used in ATLAS. The distance between two objects i and j , d_{ij} , is defined as

$$d_{ij} = \min \left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2} \right) \frac{\Delta R_y^2}{R^2},$$

where p_T is the transverse momentum of each object and ΔR_y is the angular distance calculated from $\Delta R_y^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$. R is a tunable radius parameter and is set to 0.4 for the small- R jets used in the analysis. The algorithm searches for the smallest distance between any pair of objects and merges them. This process continues until no object remains or the smallest distance d_{ij} is greater than the distance to the beam line. Then the group of merged objects form a jet. The whole procedure repeats to reconstruct multiple jets in an event, and stops when no object is left.

As a small p_T leads to a considerable distance, the anti- k_t algorithm is insensitive to the emission of real and virtual soft gluons that are both individually infrared divergent but cancel each other. Because of this, the algorithm produces very smooth cone-shaped jets with an area of $A_{\text{jet}} = \pi R^2$. In other words, the anti- k_t algorithm achieves both theoretically infrared safety and phenomenologically sensible behaviour.

Identification

Several discriminants are built to separate hard-scatter jets from PU jets. The jet-vertex-fraction (JVF) [136] is defined by the p_T of the tracks that are associated with a certain jet:

$$\text{JVF} = \frac{\sum_i p_T(\text{trk}_i)}{\sum_i p_T(\text{trk}_i) + \sum_j p_T(\text{trk}_j)},$$

where i runs over all hard-scatter tracks and j runs over the rest (presumably PU) tracks associated with the jet. A more significant value of JVF implies that the jet comes more likely from the hard-scatter vertex while a smaller fraction indicates the jet is more likely to be a PU jet. However, JVF is vertex number dependent. A modified quantity, corrJVF, corrects for the number of vertices dependence from the scale sum p_T of PU tracks:

$$\text{corrJVF} = \frac{\sum_i p_T(\text{trk}_i)}{\sum_i p_T(\text{trk}_i) + \frac{\sum_j p_T(\text{trk}_j)}{k \cdot N_{\text{trk}}(\text{PU})}},$$

where $k = 0.01$. It is validated, though, that the performance is insensitive to the choice of k [137]. Another variable is R_{pT} , defined as the scalar sum of the hard-scatter track p_T in the jet over the calibrated jet p_T :

$$R_{pT} = \frac{\sum_i p_T(\text{trk}_i)}{p_T(\text{jet})}.$$

R_{pT} peaks at 0 for PU jets while is spread largely for hard-scatter jets. Since it uses only hard-scatter tracks, it is insensitive to the number of PU tracks, thus provides supplementary information to corrJVF.

A combined discriminant called jet-vertex-tagger (JVT) [137] is built as a two-dimensional corrJVF- R_{pT} likelihood, based on a k -nearest neighbour algorithm. $k = 100$ is a user-defined constant. The algorithm looks for the nearest k neighbour points of the unlabelled query point in the two-dimensional space and assigns the probability for the jet to be hard-scattered as the fraction of hard-scatter jets in the k neighbours. The distance in the two-dimensional plane is defined by a Euclidean metric. The training sample contains hard-scatter jets and PU jets with $20 \text{ GeV} < p_T < 50 \text{ GeV}$ and $|\eta| < 2.4$. Figure 3.13 shows PU fake rate versus efficiency curves for JVT, as well as corrJVF and R_{pT} . The JVT achieves the lowest fake rate given a certain efficiency. Its performance is driven by corrJVF in the high efficiency region and R_{pT} in the high rejection range. Efficiencies of 80 %, 90 %, 92 % and 95 % are achieved for fake rates of respectively 0.4 %, 1 %, 2 % and 4 %.

Jet energy scale calibration

Calibration of the jet energy scale (JES) restores the truth energy of the jet, accounting for non-compensated calorimeter, dead material effects, energy leakage, energy deposits below thresholds, etc. The calibration is done in several steps [138]. Firstly, a jet is forced to point back to the identified hard-scatter vertex. Then, the contamination from PU tracks is estimated and subtracted using $\rho \times A$, where ρ is the PU energy density and A is the jet area in the η - ϕ plane. A PU residual term depending on the number of primary vertices, N_{PV} , and the number of bunch crossings, μ , is also deducted. Thirdly, the energy responses ($p_T^{\text{reco}}/p_T^{\text{truth}}$) in p_T and η regions are corrected. The residual dependencies on the responses from jets with different longitudinal and transverse features (most

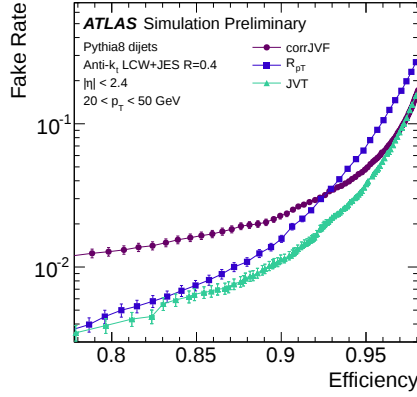


Figure 3.13: Fake rate vs efficiency for corrJVF, R_{pT} , and JVT. Results are shown for $20 \text{ GeV} < p_T < 50 \text{ GeV}$ and $|\eta| < 2.4$ in simulated dijet events. Figure is taken from [137].

notably between quark- and gluon-initiated jets), such as energy distribution within the jet, particle composition and shower shape, therefore different levels of energy leakage, are also taken into account by the global sequential calibration (GSC). It uses five variables to access the response dependence on several sources. The fractions of deposits in the first layer of the Tile calorimeter and in the third (last) layer of the EMC give a hint to the energy leakage. The number of tracks in the jet and their p_T -weighted width¹¹ access the information of jet flavours. The number of muon segments in the jet helps for high- p_T jets. The last step is the in-situ calibration, applied only to data measurements. This step accounts for the differences in the jet response stemming from the imperfect simulation of detector material and response, approximation in the hard scattering, UE, PS simulation, inaccurate particle-detector interaction, etc. The relative in-situ calibration (η -intercalibration) and absolute in-situ calibrations (Z/γ +jet balance and multijet balance) are developed, and both together cover a large kinematic phase space,

The jet energy resolution (JER) quantifies the amount of fluctuations in the measurement of a reconstructed energy at a fixed true energy. It is measured with two in-situ techniques [139]: dijet balance and bisector method in both data and MC events. Overall, the JER in data is well modelled by the MC.

The systematic uncertainties are also studied thoroughly in [138]. 88 JES systematic uncertainty terms, propagated from the individual calibrations, are associated with the final calibration. 75 in-situ calibration related terms can be further grouped to 8 for easier and faster implementation in physics analyses. In total, 21 sources of JES (including JER) uncertainties are used in this thesis. Figure 3.14 shows the relative size of even further categorised JES uncertainties as a function of p_T at $\eta = 0$ and as a function of η at $p_T = 80 \text{ GeV}$. The uncertainty is notable at low p_T , reaches a minimum value around a few hundred GeV, then rise sharply after 2 TeV. The significant uncertainties mainly come from insufficient statistical information in in-situ calibration. The flavour composition and response uncertainties from the assumption of a quark-gluon ratio of one-to-one have significant contributions. This is further reduced in the analysis by using the proper fractions in $t\bar{t}$ and tW events. The uncertainties over η are relatively flat. Slight raise presents in the forward regions and humps in $2.0 < |\eta| < 2.6$ due to the non-closure uncertainty of the η -intercalibration.

¹¹ The track width is defined as the p_T -weighted mean of the angular distances between the jet axis and all the associated tracks.

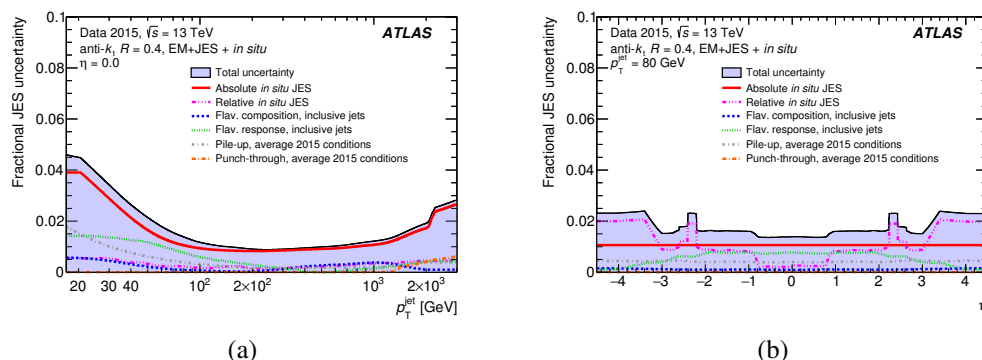


Figure 3.14: Relative uncertainty in the JES of fully calibrated jets as a function of jet p_T at $\eta = 0$ (a) and η at $p_T = 80$ GeV (b). Figures are taken from [138].

3.3.6 Tagged jets

The identification of jets that stem from the hadronisation of a b quark, or b -jets, makes use of the distinctive features of these quarks or hadrons. Usually, the b hadron carries about 70 % of the original b -quark momentum. Due to its large mass, decay products of a b hadron may have large transverse momenta and large opening angle. Last but not least, b hadrons have a relatively long lifetime, of the order of a picosecond, and can fly a few millimetre before they decay. Therefore, there is a high probability to create a secondary vertex. The procedure of identifying b -jets, called b -tagging, is done using a multivariate analysis that combines all the information.

b -tagging algorithm

Three basic algorithms are used in Run 2 [140, 141] to exploit the b -jet properties. The impact parameter based algorithm is based on large impact parameter of tracks from the b -hadron decay. It builds two likelihood ratios between b -jet and other flavour jets. One such ratio, IP3D, uses both transverse and longitudinal impact parameter and their correlation, while the other one, IP2D, takes into account only the transverse impact parameter. The secondary vertex finding algorithm searches for a secondary vertex and checks its properties such as the invariant mass of objects associated to it, the number of tracks, the transverse decay length, etc. The multi-vertex fit algorithm attempts to reconstruct the full b -hadron to c -hadron decay chain using a Kalman filter technique. This approach, when resolution allows, can resolve b - and c -hadron vertices. In total, 24 variables are derived from the three basic algorithms and fed into a boosted decision tree algorithm (MV2¹²). The name of the algorithm with the pattern MV2cXY suggests that the training is done on signal b -jets against a mixture background of about $(1 - XY\%)$ light-flavour jets and $XY\%$ c -jets¹³. The tagger used in the thesis is MV2c10, which gives a good trade-off between light-jets and c -jet rejections. The efficiencies of different jet-flavours are depicted in Fig. 3.15.

¹² MV1 was used during Run 1, based on a neural network approach rather than a boosted decision tree algorithm.

¹³ Given that the majority of physics analyses are more limited by the c -jet rejection, the fraction of c -jets used is 7 % (15 %) for MV2c10 (MV2c20).

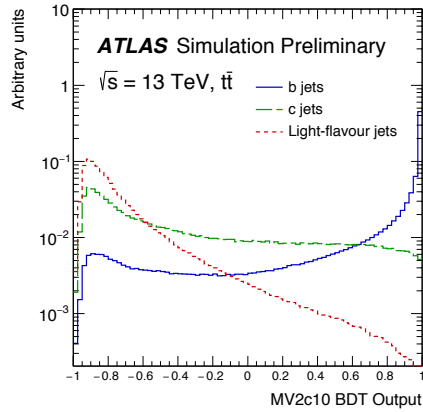


Figure 3.15: The MV2c10 output for b -jets (blue), c -jets (green) and light-flavour jets (red) in simulated $t\bar{t}$ events. Figure is taken from [141].

Working points

For harmonisation purpose and centralisation of the systematic uncertainty estimation efforts, several b -tagging working points are provided and recommended for all ATLAS analyses. They are defined by a single cut on the MV2c10 output (Fig. 3.15) so that a particular b -jet efficiency is satisfied. Table 3.5 summarises the efficiencies for different working points. Rejection is calculated such that one out of this number of jets is misidentified as a b -jet. The numbers are averaged over jet p_T and η . In this thesis, 77 % efficiency working point is chosen to balance b -tagging efficiency and rejection power.

Efficiency [%]	Cut value	Purity	c -jet rejection	τ -jet rejection	Light-jet rejection
60	0.9349	99.0	34	184	1 538
70	0.8244	97.5	12	55	381
77	0.6459	95.2	6	22	134
85	0.1758	89.7	3.1	8.2	33

Table 3.5: Efficiency and rejections rates for each MV2c10 b -tagging algorithm working points. Rejection is calculated such that one out of its value number of jets is misidentified as a b -jet. Values are estimated using a simulated $t\bar{t}$ sample. Numbers are taken from [141].

3.3.7 Missing transverse momentum

The missing momentum in the transverse plane is computed to cover the undetectable particles, such as neutrinos. It relies heavily on track and cluster measurements, as well as jet measurement, therefore is also a measure of overall event reconstruction performance.

The x and y components of the missing transverse momentum, using calibrated hard-scatter objects, are defined as [142]

$$E_{x,y}^{\text{miss}} = E_{x,y}^{\text{miss},e} + E_{x,y}^{\text{miss},\gamma} + E_{x,y}^{\text{miss},\tau} + E_{x,y}^{\text{miss},\text{jets}} + E_{x,y}^{\text{miss},\mu} + E_{x,y}^{\text{miss},\text{soft}},$$

where each object term is the *negative* vectorial sum of the momenta of the respective calibrated objects in the event. The final term, *soft* term, comes from contributions of low momenta objects that

originate from hard-scatter vertex but could not be reconstructed as any of the other identified objects. It can be computed using low- p_T tracks, resulting in a track-based soft term, or using low energy topological clusters, resulting in a calorimeter-based soft term. As the tracker is insensitive to the PU tracks and is able to catch particles that cannot reach the calorimeters, the track-based soft term is preferred. In this way, E_T^{miss} is robust against PU effects but is sensitive to soft neutral particles.

From the x and y components, the magnitude can be calculated as

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2},$$

and the azimuthal angle is

$$\phi^{\text{miss}} = \arctan(E_y^{\text{miss}} / E_x^{\text{miss}}).$$

Another important quantity is the scalar sum of transverse momenta of those objects, $\sum E_T$, defined as

$$\sum E_T = \sum p_T^e + \sum p_T^\gamma + \sum p_T^\tau + \sum p_T^{\text{jets}} + \sum p_T^\mu + \sum p_T^{\text{soft}}.$$

Analysis setup

This chapter presents the setup of the analyses. Section 4.1 details the dataset delivered by the LHC and collected by ATLAS. Simulated samples that are used to check the sensitivity of the analysis and to understand the backgrounds are summarised in Section 4.2. To enrich the tW events in data, a series of selections is designed and explained in Section 4.3. After the event selection, Section 4.4 documents the fake-lepton background estimation. Section 4.5 presents the performance of the full event selection.

4.1 Collision data

The dataset used in this thesis corresponds to an integrated luminosity of 36.1 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$. It was collected by the ATLAS detector at the LHC in 2015 and 2016, operating with a bunch spacing of 25 ns and in-time PU $\langle\mu\rangle \sim 24$ (see Fig. 4.1). The dataset is labelled as “good for physics” when all ATLAS detector systems are operating properly.

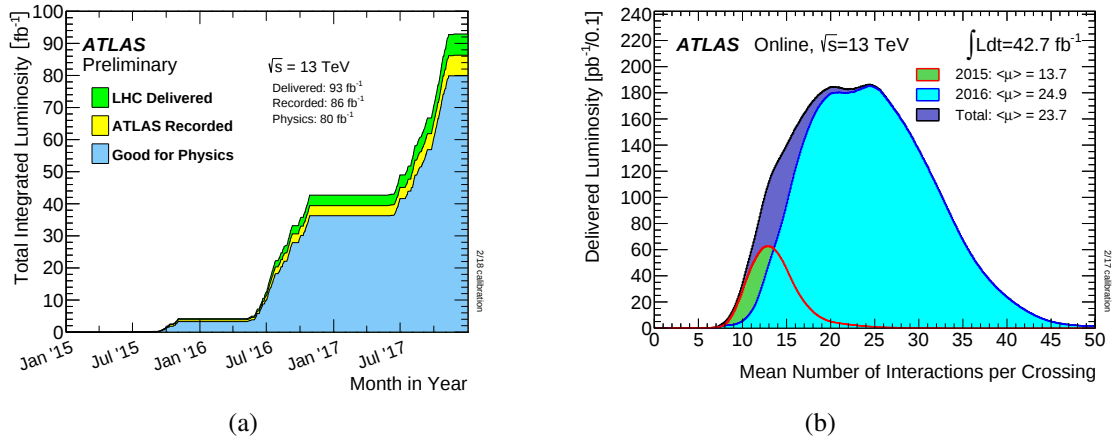


Figure 4.1: The integrated luminosity up to 2017 data taking period (a) and in-time PU condition in 2015–2016 (b). The thesis uses only 2015–2016 data. Plots are taken from ATLAS public website [104].

4.2 Monte Carlo (MC) simulation

MC simulation includes event generation and detector simulation. The generator programs determine which partons collide using the PDFs information, perform the hard collisions according to the ME, generate initial- and final-state radiation (ISR/FSR), develop parton showering and hadronisation, overlay UE and PU, choose the decay modes of each unstable particle, and compute their angles and momenta. In addition, the detector simulation models the geometry of the detector from the beam pipe to the end of the cavern. The simulation traces the interactions between particles and matter in the entire detector. Millions of events are simulated to cover all interesting phase space. For data analysis, all physical processes that are relevant are simulated, added up, and compared to real collision data.

4.2.1 Detector simulation

GEANT 4 [143] is used in ATLAS to simulate the geometry and material of the subdetectors in great detail [144]. Alternatively, two programs are also developed to accelerate the simulation speed. FASTCALOSIM (Fast Calorimeter Simulation) package [145] provides a parametrised simulation of the particle energy response and the energy distribution in the calorimeters and hence reduces the time of calorimeter simulation. The parametrisation is based on the GEANT 4 simulations of a single photon, electron and charged pion in a fine grid of simulated particle energies and directions. The overall simulation time is reduced by approximately one order of magnitude compared to the full GEANT 4 simulation. FATRAS (Fast ATLAS TRacking Simulation) package [146] simplifies the geometry used in tracking and uses parametrised models for the particle-matter interaction. It achieves a time reduction of one order of magnitude in the trackers [147]. By independently turning on and off both programs, it ends up with four different configurations with varying degrees of accuracy and simulation speed. Table 4.1 summarises the four strategies in ascending order of speed (descending order of accuracy).

Strategy	Nickname	Description	Used in the thesis?
Full GEANT 4 Simulation	FULLSIM	Detailed detector simulation	✓
Fast GEANT 4 Simulation	—	Pre-simulated showers	✗
Atlfast-II	AF2	FASTCALOSIM	✓
Atlfast-IIF	AF2F	FASTCALOSIM and FATRAS	✗

Table 4.1: Different detector simulation strategies used in ATLAS. “FULLSIM” and “AF2” are used in the thesis.

4.2.2 Signal tW channel

The nominal tW event samples [148] (DR scheme) are produced using the POWHEG-Box 1 [149] event generator with the CT10 PDF set [150] in the ME calculation. The PS, hadronisation, and UE are simulated using PYTHIA 6.428 [151] with the CTEQ6L1 PDF set [152] and the corresponding Perugia 2012 (P2012) tune [153]. The EVTGEN v1.2.0 program [154] is used to simulate properties of the bottom and charmed hadron decays and FULLSIM is used for the detector simulation. PU effects are included by overlaying collisions with the soft QCD processes from PYTHIA 8.186 [155] using a set of tuned parameters called the A2 tune [156] and the MSTW2008LO PDF set [157]. The five-flavour

scheme is used for the PDFs. The top-quark mass, m_t , is set to 172.5 GeV and the $W \rightarrow \ell \nu$ branching ratio is set to 0.108 per lepton flavour. The renormalisation and factorisation scales are set to m_t . The tW cross-section at $\sqrt{s} = 13$ TeV is scaled by a K -factor and set to the NLO value with NNLL soft-gluon corrections: $\sigma_{\text{theory}} = 71.7 \pm 1.8$ (scale) ± 3.4 (PDF) pb. The first uncertainty accounts for the renormalisation and factorisation scale variations (from 0.5 to 2 times m_t), while the second uncertainty originates from uncertainties in the MSTW2008 NLO PDF sets.

Additional tW samples are generated to estimate systematic uncertainties in the modelling of the signal process. An alternative tW sample is generated using the DS scheme instead of DR. A tW sample generated with MADGRAPH5_aMC@NLO 2.2.2 [158] (instead of the POWHEG-Box) interfaced with HERWIG++ 2.7.1 [159] is used to estimate uncertainties associated with the modelling of the ME event generator. A sample generated with POWHEG-Box interfaced with HERWIG++ 2.7.1 (instead of PYTHIA 6.428) is used to estimate uncertainties associated with the PS, hadronisation, and UE models. This sample is also compared with the previously mentioned MADGRAPH5_aMC@NLO sample to estimate an ME event generator uncertainty with a consistent PS event generator. In both cases, the UE-EE-5 tune [160] is used for the UE and the Af2 fast simulation is used for the detector simulation. Finally, in order to estimate uncertainties arising from additional ISR/FSR in the tW events, a pair of samples are generated with POWHEG-Box interfaced with PYTHIA 6.428 using Af2 and the P2012 tune with higher and lower radiations relative to the nominal set, together with various renormalisation and factorisation scales. To avoid comparing two different detector response models when estimating systematic uncertainties, another version of the nominal POWHEG-Box with PYTHIA 6.428 sample is also produced with Af2. The parameters such as the top-quark mass, branching ratios, cross-section, renormalisation and factorisation scales are set to be the same as the nominal sample settings in all these alternative samples. For all the additional samples, EVTGEN v1.2.0 is used to simulate properties of the bottom; charmed hadron decays and PU effects are simulated in the same manner as the nominal sample. Table 4.2 summarises simulated tW samples used in the thesis.

4.2.3 Major background

The major background process comes from the top-quark pair production, $t\bar{t}$. The nominal $t\bar{t}$ event sample [148] is produced using the POWHEG-Box 2 [161] event generator with the CT10 PDF set [150] in the ME calculations. The PS, hadronisation, UE and PU effects are simulated using the same setup as the tW nominal samples. Similarly, the EVTGEN v1.2.0 program is used to simulate properties of the bottom and charmed hadron decays. FULLSIM is used for the detector simulation. The parameters such as the top-quark mass, branching ratio, and cross-section are set to be the same as well, except for the renormalisation and factorisation scales being $\sqrt{m_t^2 + p_T(t)^2}$. The h_{damp} resummation damping factor is set to equal the mass of the top quark. The $t\bar{t}$ cross-section is set to be $\sigma_{t\bar{t}} = 831.8^{+19.8}_{-29.2}$ (scale) ± 35.1 (PDF + α_s) pb as calculated with the TOP++ 2.0 program to NNLO, including soft-gluon resummation to NNLL [162]. The first uncertainty comes from the independent variation of the factorisation and renormalisation scales, while the second one is associated with variations in the PDF and strong interaction coupling, α_s , following the PDF4LHC prescription with the MSTW2008 68% CL NNLO, CT10 NNLO and NNPDF2.3 5f FFN PDF sets [163–166].

Additional $t\bar{t}$ samples are generated to estimate systematic uncertainties. These are used to estimate the uncertainties associated with the ME event generator (a sample produced using Af2 fast simulation with MADGRAPH5_aMC@NLO 2.2.2 interfaced with HERWIG++ 2.7.1), PS and hadronisation models

(a sample produced using AF2 with POWHEG-Box interfaced with HERWIG++ 2.7.1) and additional ISR/FSR. For uncertainty estimation of the ISR/FSR, a pair of samples is produced using FULLSIM with the varied sets of P2012 parameters for higher and lower radiations, as well as with varied renormalisation and factorisation scales. In these samples, the resummation damping factor h_{damp} and scale factors are doubled in the sample for higher radiation while they are halved in the sample for lower radiation. Table 4.3 summarises the simulated $t\bar{t}$ samples used in the thesis.

4.2.4 Other (small) backgrounds

Samples used to model the Z + jets background [169] are simulated with SHERPA 2.2.1 [170]. In these samples, the ME is calculated for up to two partons at NLO and four partons at LO using COMIX [171] and OPENLOOPS [172], and merged with the SHERPA PS [173] using the ME+PS@NLO prescription [174]. The NNPDF3.0 NNLO PDF set [175] is used in conjunction with SHERPA PS tuning, with a generator-level cut-off on the dilepton invariant mass of $m_{\ell\ell} > 40$ GeV applied. FULLSIM is used for the detector simulation. The Z + jets events are normalised using NNLO cross-sections computed with FEWZ [176]. They are simulated in p_T slices for each leptonic Z decay channel and jet flavour (see Table 4.4).

Diboson processes with four charged leptons, three charged leptons plus one neutrino, or two charged leptons plus two neutrinos [178] are simulated using the SHERPA 2.1.1 event generator. The ME contains all diagrams with four electroweak vertices. NLO calculations are used for the purely leptonic final states as well as for final states with two or four charged leptons plus one additional parton. For other final states with up to three additional partons, the LO calculations of COMIX and OPENLOOPS are used. Their outputs are combined with the SHERPA PS using the ME+PS@NLO prescription. The CT10 PDF set with dedicated PS tuning is used. FULLSIM is used for the detector simulation. The cross-sections provided by the event generator (which are already at NLO) are used for diboson processes. A scaling of 0.91 is applied to the predicted cross-section from the SHERPA diboson samples to account for the choice of QED coupling parameter and other electroweak parameters chosen in the sample generation. Table 4.5 lists the generators and cross-sections used for simulated diboson processes.

Other processes such as W + jets, $\ell\ell$ + jets with $m_{\ell\ell} < 40$ GeV are found to be negligible. Therefore they will not be mentioned explicitly, although they are included in fake-lepton event estimates (Section 4.4). The samples used and their corresponding cross-sections are documented in Appendix A. Example Feynman diagrams for these processes are illustrated in Fig. 4.2.

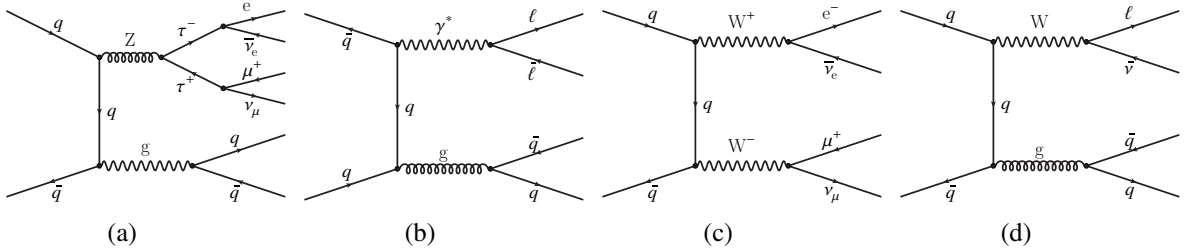


Figure 4.2: Example LO Feynman diagrams of small backgrounds Z + jets (a), $\ell\ell$ + jets (b), diboson (c), W + jets (d). $\ell\ell$ + jets and W + jets contribute via a fake lepton.

tW sample	ME	+	PDF	PS	+	PDF	Tune	PU + PDF + tune	b/c decay	Detector
Nominal (DR)	POWHEG-Box 1			PYTHIA 6.428			P2012			FULLSIM
Af2 nominal (DR)	POWHEG-Box 1			PYTHIA 6.428			P2012			
DS	POWHEG-Box 1			PYTHIA 6.428			P2012	PYTHIA 8.186 + MSTW2008LO		
ME (DR)	AMC@NLO 2.2.2	CT10		HERWIG++ 2.7.1	CTEQ6L1		UE-EE-5	+ A2	EvtGEN v1.2.0	Af2
PS (DR)	POWHEG-Box 1			HERWIG++ 2.7.1			UE-EE-5			
ISR/FSR (DR)	POWHEG-Box 1			PYTHIA 6.428			P2012radHi			
ISR/FSR (DR)	POWHEG-Box 1			PYTHIA 6.428			P2012radLo			

Table 4.2: Summary of generators used in simulated tW events. “Nominal” sample is used as baseline while alternatives samples are used to estimate the systematic uncertainties. More details can be found [167].

$t\bar{t}$ sample	ME	+	PDF	PS	+	PDF	Tune	PU + PDF + tune	b/c decay	Detector
Nominal	POWHEG-Box 2			PYTHIA 6.428			P2012			FULLSIM
Af2 nominal	POWHEG-Box 2			PYTHIA 6.428			P2012			
ME	AMC@NLO 2.2.2	CT10		HERWIG++ 2.7.1	CTEQ6L1		UE-EE-5	PYTHIA 8.186 + MSTW2008LO + A2	EvtGEN v1.2.0	Af2
PS	POWHEG-Box 2			HERWIG++ 2.7.1			UE-EE-5			
ISR/FSR	POWHEG-Box 2			PYTHIA 6.428			P2012radHi			FULLSIM
ISR/FSR	POWHEG-Box 2			PYTHIA 6.428			P2012radLo			

Table 4.3: Summary of generators used in simulated $t\bar{t}$ events. “Nominal” sample is used as baseline while alternatives samples are used to estimate the systematic uncertainties. More details can be found [168].

Jet filter	Z decay	Cross-section [pb]					
		Jet p_T [GeV]					
		0–70	70–140	140–280	280–500	500–1 000	> 1 000
Light-jet	ee	1 590	74.9	24.4	4.80		
	$\mu\mu$	1 590	73.4	23.4	4.66		
	$\tau\tau$	1 610	74.7	24.6	4.76		
c -jet	ee	220	20.3	9.24	2.25		
	$\mu\mu$	220	20.9	9.15	2.21		
	$\tau\tau$	212	20.5	9.30	2.24		
b -jet	ee	127	12.7	6.08	1.49		
	$\mu\mu$	127	12.5	6.08	1.47		
	$\tau\tau$	127	12.0	6.17	1.49		
—	ee					1.76	0.145
	$\mu\mu$					1.74	0.144
	$\tau\tau$					1.76	0.145

Table 4.4: The cross-sections for Z + jets as a function of jet p_T slices and lepton-jet flavours from Z decay. They are normalised to NLO (LO for more than 2 jets) theoretical calculation. Jet flavours are simulated inclusively when jet $p_T > 500$ GeV. Numbers are extracted from [177].

Sample	Generator	PDF	Detector	σ [pb]	Comment
$ZZ \rightarrow \ell\bar{\ell} \ell^{(\prime)}\ell^{(\prime)}$	SHERPA 2.2.1	CT10	FULLSIM	11.7	$m_{\ell\ell} > 2 \times m_\ell + 250$ MeV, $p_T(\ell) > 5$ GeV
$ZW^- \rightarrow \ell\bar{\ell} \ell^- \nu$				1.68	
$ZW^- \rightarrow \ell\bar{\ell} \ell'^- \nu$				3.30	
$ZW^+ \rightarrow \ell\bar{\ell} \ell^+ \bar{\nu}$				2.33	
$ZW^+ \rightarrow \ell\bar{\ell} \ell'^+ \bar{\nu}$				4.57	
$ZZ (WW) \rightarrow \ell\bar{\ell}^{(\prime)} \nu\bar{\nu}^{(\prime)}$				12.7	
$W^+ W^- \rightarrow \ell^+ \bar{\nu} q\bar{q}'$				23.7	
$W^+ W^- \rightarrow q\bar{q}' \ell^- \nu$				23.6	
$W^+ Z \rightarrow \ell^+ \bar{\nu} q\bar{q}$				11.4	
$W^+ Z \rightarrow q\bar{q}' \ell\bar{\ell}$				3.42	
$ZZ \rightarrow q\bar{q} \ell\bar{\ell}$				2.15	

Table 4.5: Generator information and the input cross-sections of the diboson processes for each decay channel. The cross-sections are calculated at NLO (LO for more than 1 jet) and the numbers shown are before applying the scale factor 0.91 (see text). Numbers are extracted from [179].

4.3 Event selection

Event selection is applied to filter out the signal events, at the same time to minimise the background contribution. Event selection is done according to the topology of the final state.¹

4.3.1 Preselection

Events are triggered by a single-electron or single-muon trigger. Single-lepton triggers are designed to select events containing a well-identified charged lepton with a high transverse momentum [121]. They require a p_T of at least 20 GeV (26 GeV) for muons and 24 GeV (26 GeV) for electrons for the 2015 (2016) dataset, and also have requirements on the lepton quality and isolation. These triggers are complemented by triggers with higher p_T thresholds and relaxed isolation and identification requirements to ensure maximum efficiency at higher lepton p_T .

Events must contain exactly one electron and one muon with opposite charges to minimise the Z + jets contamination. The electron is required to be in the $|\eta| < 2.47$ region, with the exclusion of the transition region between the barrel and endcap of the EMC ($1.37 < |\eta| < 1.52$). It is required to have a transverse momentum of $p_T > 27$ GeV to suppress fake electrons from the jets. The electron must pass the Tight criterion for the identification and Gradient for the isolation. Electron tracks are also required to be consistent with the beam spot, requiring the impact parameter to satisfy $|d_0|/\sigma_{d_0} < 5$ and $|\Delta z_0 \sin \theta| < 0.5$ mm, with σ_{d_0} being the uncertainty in d_0 , Δz_0 being the longitudinal distance from the primary vertex along the beam pipe, and θ being the angle of the track to the beam pipe. The muon must have $p_T > 27$ GeV as well as $|\eta| < 2.5$. The muon candidate must pass the Medium criterion for the identification and Gradient for the isolation. For muons tracks the impact parameter is required to satisfy $|d_0|/\sigma_{d_0} < 3$ and $|\Delta z_0 \sin \theta| < 0.5$ mm, tighter than those for electrons. Events with a third charged lepton with $p_T > 20$ GeV are rejected. At least one of the selected electrons (muons) must be matched within a ΔR cone of size 0.07 (0.1) to the electron (muon) selected online by the corresponding trigger.

A track-based E_T^{miss} is required to have $E_T^{\text{miss}} > 20$ GeV to count for neutrinos.

4.3.2 Jet and tag selection

Events must contain at least one interaction vertex and at least one jet. The jet must have $p_T > 25$ GeV and $|\eta| < 2.5$. The JVT has to be larger than 0.59 for the jets with $p_T < 50$ GeV and $|\eta| < 2.4$, to suppress PU jets with a 92 % efficiency and a 2 % mistagging rate. Events are required to contain at least one b -jet with the 77 % working point to have a fair balance between tagging efficiency and background jet rejection. At LO, the signal process results in a final state with one b -jet arising from the top-quark decay, and no additional jets, while the $t\bar{t}$ process results in two b -jets from the top-quark decays. However, events with additional jets are also studied since the UE, higher order QCD, and other effects may produce an additional jet in signal events. Therefore, up to two jets are required, and the sample of selected events is further divided into regions based on the jet and b -jet multiplicities.

¹ It is similar to the selection used in [180]

4.3.3 Overlap removal

To avoid cases where a single physical object is reconstructed as two separate final-state objects, the overlap removal is implemented. It works as follows. First, identified muons that deposit energy in the calorimeter and share a track with an electron are removed, followed by the removal of any remaining electrons sharing a track with a muon. This step is designed to avoid cases where a muon mimics an electron through radiation of a hard photon. Next, the jet closest to each electron within an angular distance of $\Delta R_y \equiv \sqrt{(\Delta y)^2 + (\Delta \phi)^2} = 0.2$ is removed to reduce the number of electrons being reconstructed as jets. Electrons with a distance $\Delta R_y < 0.4$ from any of the remaining jets are removed to reduce backgrounds from non-prompt, non-isolated electrons originating from heavy-flavour hadron decays. Jets with fewer than three tracks and distance $\Delta R_y < 0.2$ from a muon are then removed to reduce the number of jet fakes from muons depositing energy in the calorimeters. Finally, muons with a distance $\Delta R_y < 0.4$ from any of the surviving jets are removed to avoid contamination due to non-prompt muons from heavy-flavour hadron decays.

4.3.4 Cut flow

The resulting numbers of events after the event selection of each step are displayed in Fig. 4.3. It depicts the yields of real data and each simulated process, as well as the sum of all simulations, with event weight correction. The initial simulation events do not necessarily match the initial data as unrelated processes are not included in the simulation and therefore trigger efficiencies can be different. When requiring two opposite charged leptons though, data and MC events reach the same level. Opposite lepton-flavour requirement helps to suppress the $Z + \text{jets}$ and $W + \text{jets}$ background. E_T^{miss} further reduces the $Z + \text{jets}$ background without losing too much signal. After the jet and tag selections, the number of simulated events matches quite well to the number of data events.

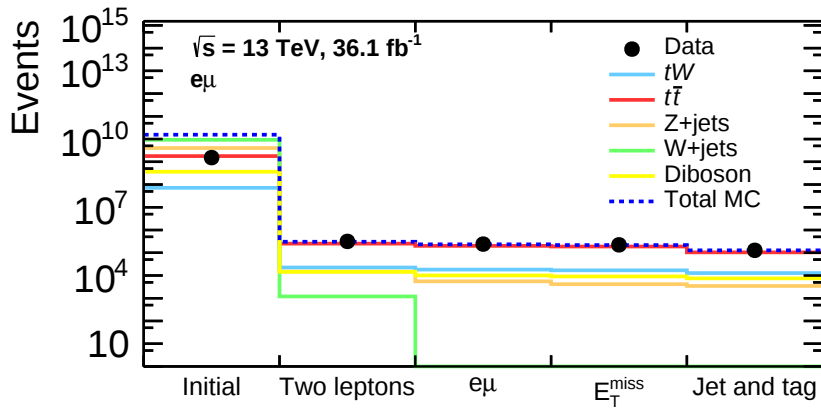


Figure 4.3: Measured and expected event yields in log-scale after each requirement. Data are shown in black dots and the sum of all simulated samples is shown as the dashed blue line. Fake lepton events are not included.

4.4 Fake-lepton estimation

In simulated events, information recorded by the event generator is used to identify events in which any selected lepton does not originate from the hard-scatter process. These non-prompt or fake leptons arise from processes such as the decay of a heavy-flavour hadron, photon conversion, or hadron misidentification, and are identified when the electron or muon does not originate from the decay of a W or Z boson (or a τ lepton itself originating from a W or Z). Events with a selected lepton which is non-prompt or fake are labelled as fake (from tW or other sources), and are treated as a contribution to the background. Events with a fake lepton are estimated from tW , $t\bar{t}$, Z + jets, W + jets, diboson, and $\ell\ell$ + jets with $m_{\ell\ell} < 40$ GeV MC samples.² The contribution of each process to the events with a fake lepton are given in Fig. 4.4.

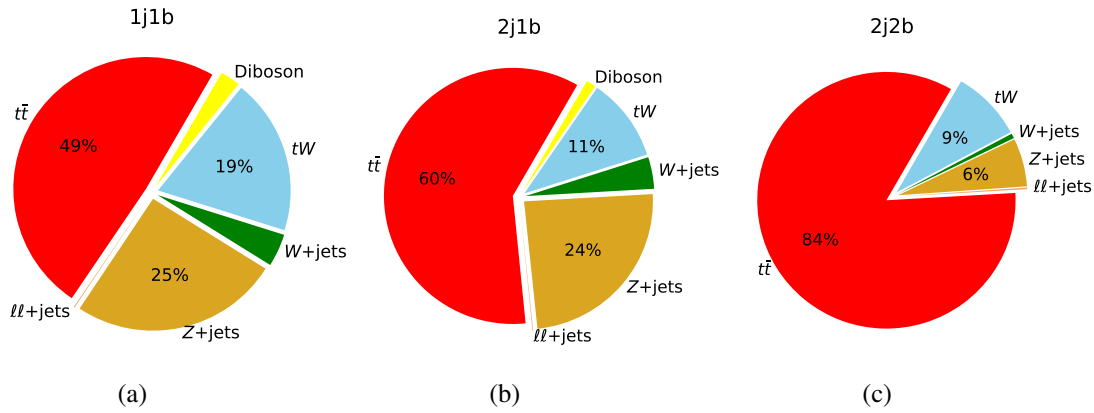


Figure 4.4: Relative contributions to the events with a fake lepton from each process in the 1j1b (a), 2j1b (b), 2j2b (c) regions. No uncertainties are included.

4.5 Selection performance

4.5.1 Analysis regions and event yields

Two signal regions are defined by the presence of exactly one b -jet and either zero (denoted 1j1b) or one (denoted 2j1b) additional jets. A $t\bar{t}$ -enriched control region is defined by the presence of exactly two jets, which are both b -tagged (denoted 2j2b). This control region is used to constrain the $t\bar{t}$ background normalisation. These three regions—1j1b, 2j1b and 2j2b—are called fit regions, as they will be used in the simultaneous fit for the inclusive cross-section measurement (see Chapter 6). The 1j1b is further used to measure the differential cross-sections (see Chapter 7). Two additional regions, in which events are required to contain one (denoted 1j0b) or two (denoted 2j0b) jets without b -jets, are used to validate the simulation. A schematic view of the regions definition can be found in Fig. 4.5(a). The event yields for signal and backgrounds with their total systematic uncertainties (see Chapter 6) and the number of observed events in data are shown in Fig. 4.5(b).

² Most people argue that this has to be done using a data-driven estimate. For this analysis, due to its small contribution, an estimating from the simulation is sufficient.

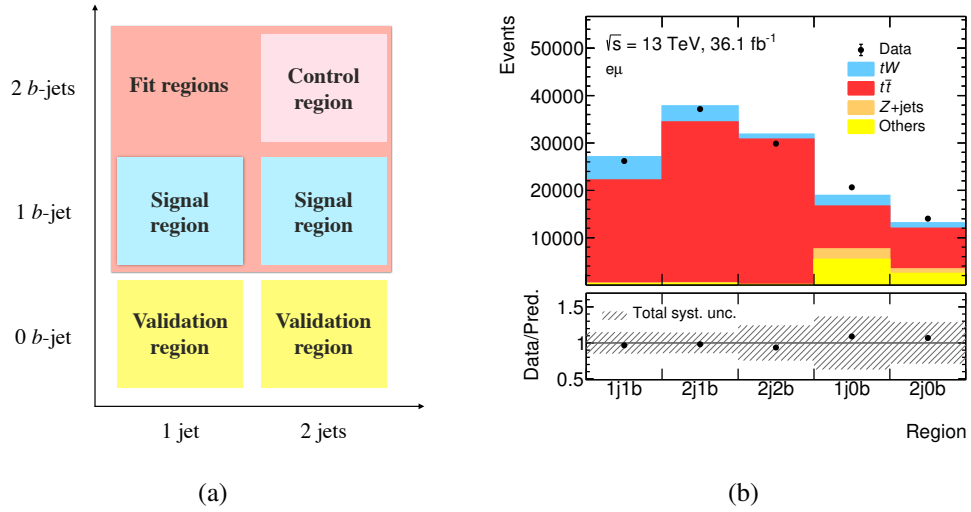


Figure 4.5: A schematic view of different regions (a) and event yields in each region (b). The signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties.

As expected, the 1j1b region has the highest signal purity (signal over background ratio), followed by 2j1b. The 2j2b region contains more than 96% $t\bar{t}$ events. $t\bar{t}$ is also the dominant background in both signal regions. Additional background sources include the Z + jets process, mostly from $Z \rightarrow \tau\tau$ where one τ decays to an e and the other to a μ , and the diboson process, where both bosons (W) decay leptonically. These processes contribute to around 1% of the total background in the three regions together.

4.5.2 Control plots

Distributions of kinematic variables reveal underlying physics that might not be visible in the yields. Separation of signal and background in shape helps to determine the number of signal events in data more precisely. The shapes are also crucial to verify the theoretical models and the detector simulation.

Figures 4.6 to 4.8 present kinematic distributions of the final objects, and of some variables that use multiple objects, for each region. Good agreement between data and the MC predictions can be seen. Shape disagreements in some variables are covered by the total systematic uncertainties.

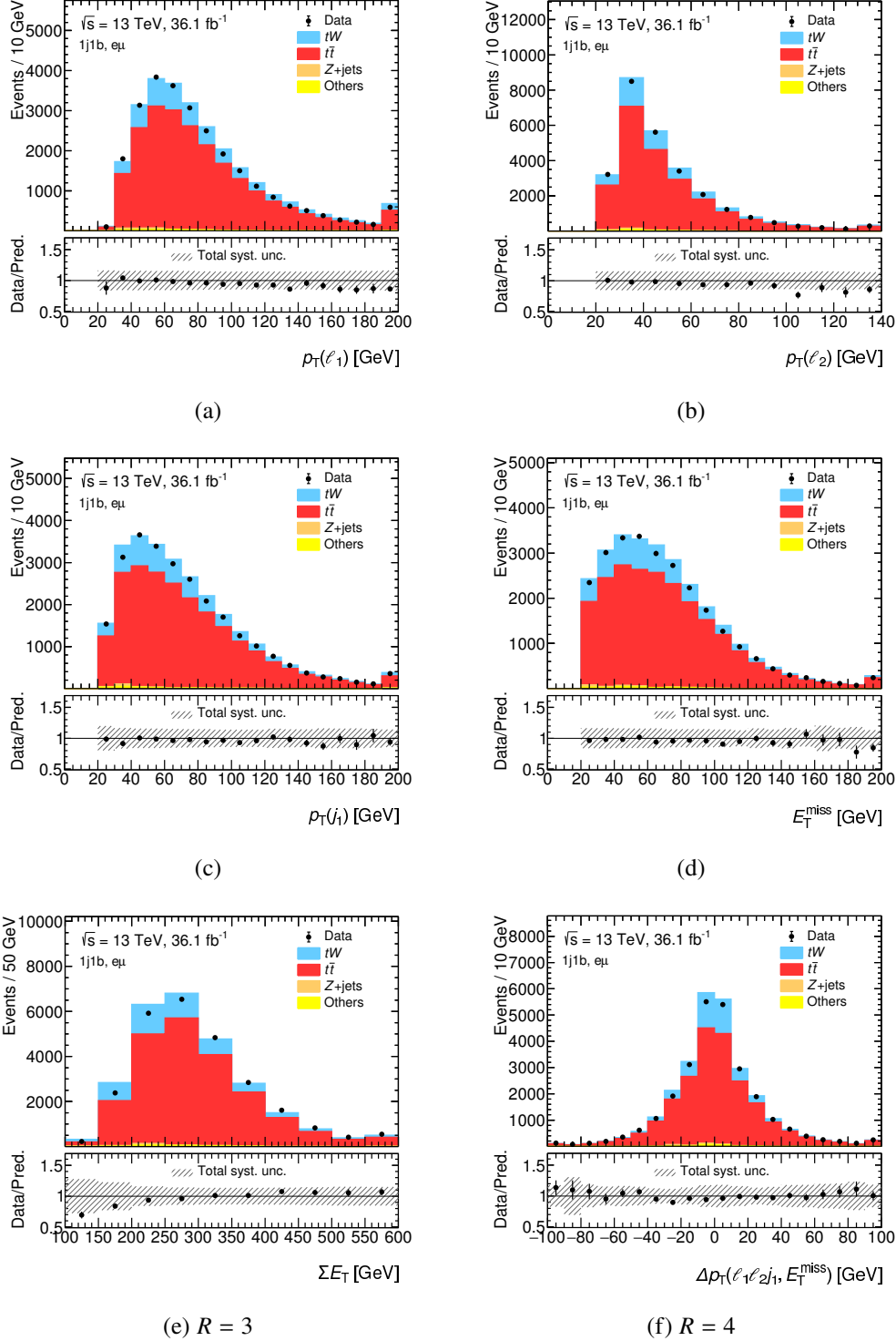


Figure 4.6: Comparison of data and MC predictions of some kinematic variables in the 1j1b region. Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panels give the yields in the number of events per bin, while the lower panels give the ratios of the numbers of observed events to the total prediction in each bin.

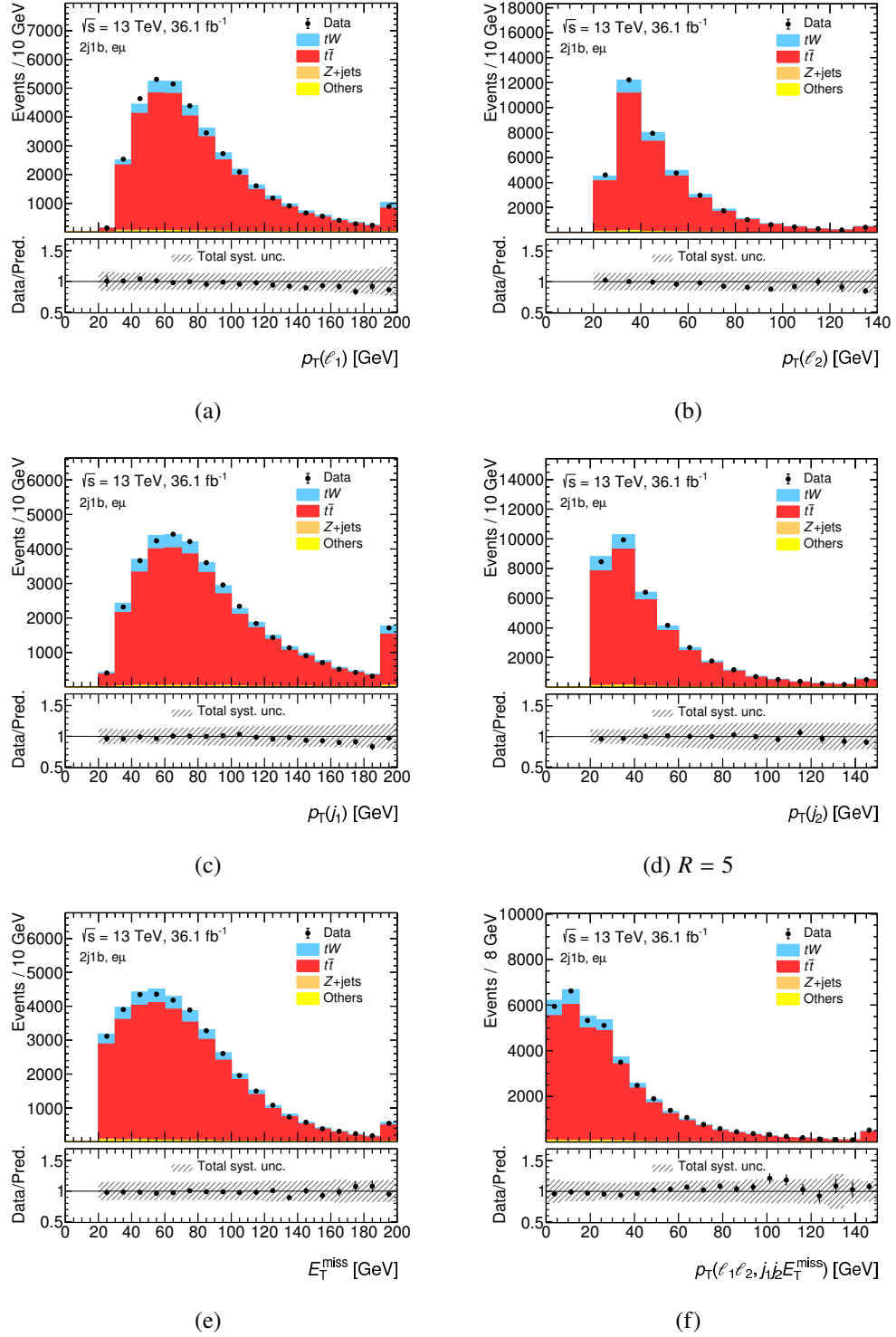


Figure 4.7: Comparison of data and MC predictions of some kinematic variables in the 2j1b region. Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panels give the yields in the number of events per bin, while the lower panels give the ratios of the numbers of observed events to the total prediction in each bin.

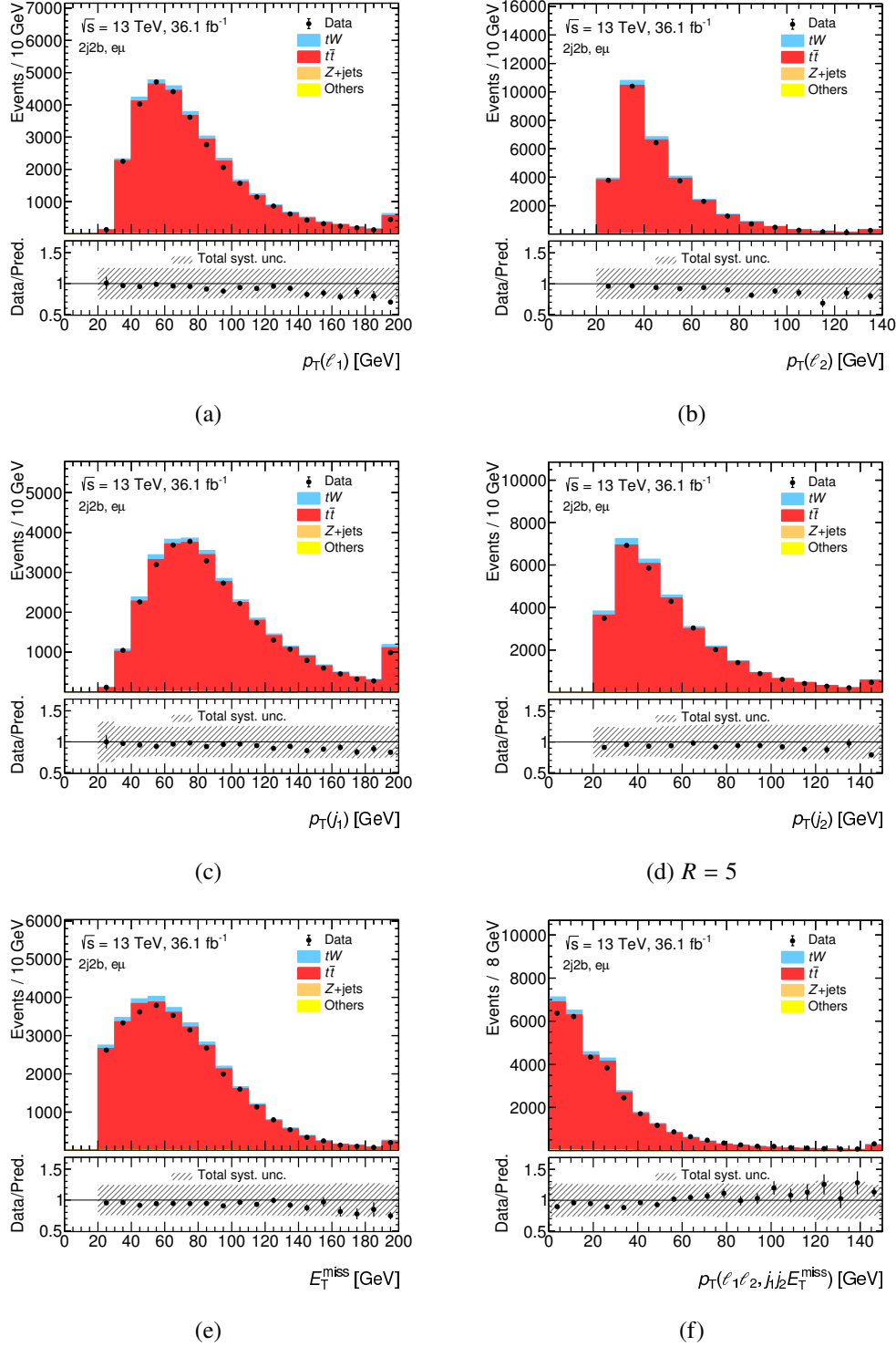


Figure 4.8: Comparison of data and MC predictions of some kinematic variables in the 2j2b region. Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panels give the yields in the number of events per bin, while the lower panels give the ratios of the numbers of observed events to the total prediction in each bin.

Multivariate analysis

After the event selection, the background consists primarily of $t\bar{t}$ events (see Fig. 4.5(b)). The similarity of signal and background event topologies excludes the possibility of extracting signal yields using a single observable, nor from a cut and count analysis. Therefore, a model that combines several observables into one powerful discriminant is necessary. This chapter explores the use of the multivariate analysis to separate signal from vast backgrounds. Section 5.1 briefly introduces the boosted decision trees technique. Then its application in the analysis is presented in Section 5.2.

5.1 Boosted decision trees technique

The boosted decision trees (BDTs) [181] uses the decision trees idea and ensembles them by a boosting algorithm [182].

5.1.1 Decision tree

A decision tree is a tree-like structure, where a predictive rule is represented in each path from the root to a leaf, to predict the value of a target variable, based on multiple input variables. A binary tree is sufficient, since any multi-way tree could degenerate to a binary tree without losing information. An upside down view of a decision tree is shown in Fig. 5.1. The root node or each internal node represents a condition that contains a single variable at a time. Two branches that connect down to the node represent the *pass* and *fail* of that condition. A stop node, also known as a leaf node, together with conjunct branches and internal nodes back to the root node, define a predictive rule. A decision tree can be used for either classification or regression. In the *classification* trees, the target are discrete values and leaves represent class labels (**S**ignal or **B**ackground) depending on the majority of training events that end up in the node. In the *regression* trees, the target takes continuous real values and leaves represent prediction score estimated by a pre-defined regression function.

5.1.2 Training a decision tree

Training a decision tree is a procedure that each node learns the splitting criteria from a collection of training events with known targets (so-called *training* sample). Starting with the root node, a split is determined by looking for a variable and a corresponding cut value that provide the best

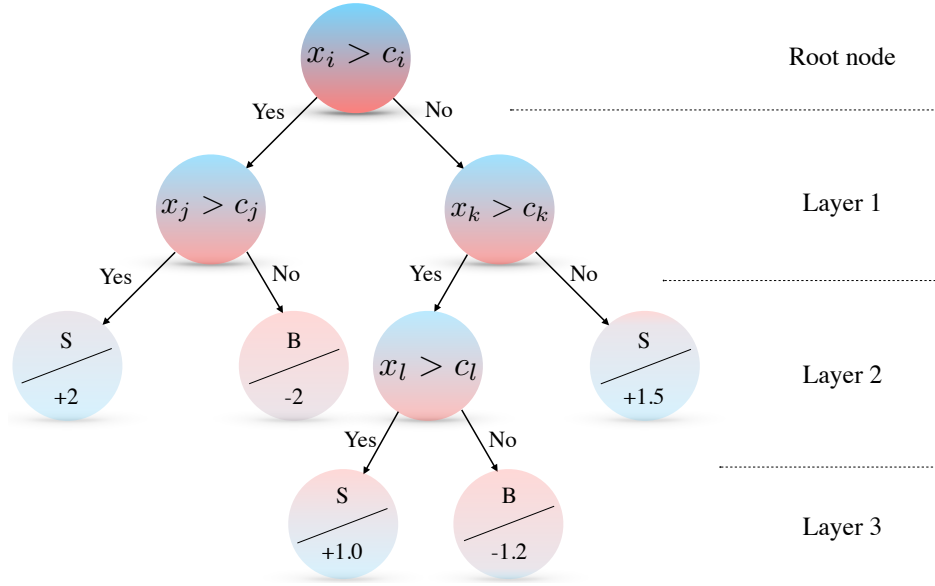


Figure 5.1: Schematic view of an upside down decision tree. Starting from root node indicated by the circle on the top, a sequence of binary splits using the discriminating variables $x_{i,j,k,\dots}$ is applied to the data. The variable used at a node gives the best separation between signal and background when being cut on the value c , which is also optimised. The same variable may be used at several nodes. The leaf nodes at the end of the tree either produce the label “S” for signal or “B” for background in a classification tree, or return a real value as a score in a regression tree; both pieces of information are labelled on the leaf nodes. The higher opacity of a node indicates a higher number of events remained in that node.

signal-background separation for the training sample. The split leads to two subsets of the training events. Each subset goes through the same algorithm to determine the next splitting criterion. The procedure is repeated at each node until one of the stopping criteria is fulfilled. The stopping criteria can be related to the minimum number of events in leaf nodes, the minimum percentage of events in leaf nodes, the maximum number of layers, etc.

The splitting criterion can be configured using a variety of indices, and they are found to have no significant difference. The Gini index is used in the thesis. At each node, the Gini index is defined as

$$G = \sum_{i \in S, B} w_i \cdot p \cdot (1 - p),$$

where w_i is the event weight of event i , and p is the probability to find a signal event in the total events in this node:

$$p = \frac{\sum_{s \in S} w_s}{\sum_{i \in S, B} w_i}.$$

Small G indicates imbalance of signal and background contributions, and thus is preferred.

5.1.3 Boosting

Boosting is an ensemble technique that combines sequentially trained decision trees (each tree is called a learner). Early learners fit a simple model to the training sample and identify the events that

are hard to fit the model. Later learners focus primarily on those examples and try to get them right. The boosting of a decision tree [182] creates a tree iteratively and combines them to a forest. Each tree generates an output $f(\vec{x})$ associated with a weight W , which is related to the accuracy. The ensemble response is the weighted sum of the single tree output:

$$\hat{y}(\vec{x}) = \sum_t W_t \cdot f_t(\vec{x}),$$

where t runs over all trees. Boosting stabilises the decision trees response concerning fluctuations in the training sample, as well as enhancing the performance of prediction to a single tree.

Different techniques vary in how to improve the early learners with the failed events. One idea is to increase the importance of those events, so that the later learners pay more attention to them. This method is known as adaptive boosting [183]. Another idea is to reformulate the problem as a minimisation of a loss function, given by the inaccuracy of the prediction to the truth. Then it can be generalised by using other loss functions. This method is known as gradient boosting [184].

Gradient boosting

The performance of the prediction of an imperfect decision tree $f_t(\vec{x})$ is improved by adding a correction,

$$f_{t+1}(\vec{x}) = f_t(\vec{x}) + \beta_t h_t(\vec{x}; \alpha_t),$$

such that $f_{t+1}(\vec{x})$ is closer to the truth $y(\vec{x})$. Here, $h_t(\vec{x}; \alpha_t)$ is a function of the inputs $\vec{x}$ and unknown parameters α_t , while β_t is another parameter of the model that needs to be determined. A loss function $L(f_{t+1}(\vec{x}); y(\vec{x}))$ that measures the deviation of the prediction and the true value is minimised:

$$(\beta_t; \alpha_t) = \arg \min_{\beta_t, \alpha_t} \sum_{\text{data points}} L(f_t(\vec{x}) + \beta_t h_t(\vec{x}; \alpha_t); y(\vec{x})).$$

The above equation is hard to solve if $h_t(\vec{x})$ is unknown. However, if one takes a step back and considers the following minimisation problem:

$$f_{t+1}(\vec{x}) = \arg \min_{f_{t+1}(\vec{x})} \sum_{\text{events}} L(f_{t+1}(\vec{x}); y(\vec{x})),$$

it formally fits the gradient descent algorithm,¹ and can be given in the closed-form solution,

$$f_{t+1}(\vec{x}) = f_t(\vec{x}) - \gamma \left. \frac{\partial L(f(\vec{x}); y(\vec{x}))}{\partial f(\vec{x})} \right|_{f(\vec{x})=f_t(\vec{x})},$$

where γ is the step size of the iteration. Comparing the two iterative formulas above, one gets

$$\beta_t = \gamma,$$

$$h_t(\vec{x}; \alpha_t) = - \left. \frac{\partial L(f(\vec{x}); y(\vec{x}))}{\partial f(\vec{x})} \right|_{f(\vec{x})=f_t(\vec{x})}.$$

¹ It is worth pointing out that the gradient descent algorithm is usually performed in parameter space, whereas $f(\vec{x})$ is a function. However, for this particular case, the algorithm still works (see the discussion in [184]).

β_t is renamed as *shrinkage* in gradient boosting and it tunes the learning speed of the algorithm. A small shrinkage demands more trees to be grown but it can significantly improve the robustness and accuracy of the prediction in challenging settings. The boosting procedure is repeated until the number of trees reaches the designed value. The final prediction of the BDT is the averaged output from all decision trees.

Different types of loss functions can be chosen, for instance, the absolute loss function, $L(f; y) = |f - y|$, or the squared loss function, $L(f; y) = \frac{1}{2}(f - y)^2$. To maintain the high efficiency of squared loss function while benefitting from the robustness of absolute loss for the outliers, the Huber loss function [185] is defined in the following way:

$$L_{\delta}(f; y) = \begin{cases} \frac{1}{2}(f - y)^2 & |f - y| \leq \delta, \\ \delta|f - y| - \frac{1}{2}\delta^2 & |f - y| > \delta, \end{cases}$$

where δ is a parameter to tune the transition boundary and the coefficients are chosen such that the function is continuous. It is suitable for the regression problem. For classification problems, however, the logistic loss function below is more capable [186]:

$$L(f; y) = \ln(1 + e^{-2f \cdot y}).$$

5.1.4 Bagging

Bootstrap aggregating, also called *bagging*, is a resampling technique where a model is repeatedly trained using resampled training events such that the combined model represents an average of the individual models. Resampling with replacement from the training data could represent the population and suppress noise. It effectively improves the performance and stabilises the response. Technically, resampling is achieved by applying a random mean-one Poisson weight to each training event.

5.1.5 Figure of merit

To verify how well a model performs, one can inspect the signal efficiency, the fraction of correctly identified signal events, and the background rejection, the fraction of correctly rejected background events, given a threshold of the predicted value. The efficiency and rejection quantities are combined in the receiver operating characteristic (ROC) curve, which is a common figure of merit. Various interpretations of the ROC curve are available, one of which is the area under the ROC curve (AUC). Larger AUC translates to better performance of the model as it keeps higher background rejection while maintaining higher signal efficiency.

5.1.6 Overtraining and undertraining

A critical aspect in the BDT, or in general in any multivariate analysis technique, is how to avoid overtraining/overfitting. An overtraining occurs when the model contains more parameters than the ones that can be determined by the data. An overtrained model is too dependent on the training data and does not generalise well to unseen data. In order to avoid this, a statistically independent dataset, the *test* sample is prepared. Overtraining can be recognised and its impact can be measured by comparing the relative performance of training and test samples.

On the contrary, undertraining/underfitting happens when a model cannot identify properties that are owned by the data. This can also be measured with the test sample. Increasing the complexity of the model, such as increasing the number of nodes or the number of trees, gives more chance to capture the underlying structure of the data.

5.2 Application of BDT

In this thesis, the BDT, provided by the TMVA package [187] using the gradient boosting algorithm, is implemented to solve the classification problem. Separate BDTs are constructed for the signal regions (1j1b, 2j1b) and the control region (2j2b). MC simulated tW nominal and DS samples are used as a signal sample to increase its statistics, and the $t\bar{t}$ nominal sample is used as a background sample (see Tables 4.2 to 4.3). Other backgrounds are not included in the training due to their low number of events. The three individual samples are randomly split to half-half. One half is used as a training sample while the other as a test sample. The input list of variables is optimised, followed by the hyperparameter optimisation of the BDT models. The optimisations are done iteratively until a convergence is reached.

5.2.1 Variables selection

Variables used in the BDT are chosen to maximise the signal to background separation power of the BDT response. The separation power S of a variable x is defined by

$$S = \frac{1}{2} \int \frac{(Y_s(x) - Y_b(x))^2}{(Y_s(x) + Y_b(x))} dx,$$

where $Y_s(x)$ and $Y_b(x)$ are the signal and background probability distribution functions of variable x , respectively.

A few hundreds of variables are constructed from combinations of final objects: for a set of objects $o_1 \dots o_n$,

- $p_T(o_1 \dots o_n)$ is the vector sum of transverse momenta;
- $H_T(o_1 \dots o_n)$ is the scalar sum of transverse momenta of the visible objects;
- $\sigma(p_T)(o_1 \dots o_n)$ is the ratio of $p_T(o_1 \dots o_n)$ to $\sqrt{H_T(o_1 \dots o_n) + \sum E_T}$ with a unit of $\sqrt{\text{GeV}}$;
- $m(o_1 \dots o_n)$ is the invariant mass of the system;
- $m_T(o_1 \dots o_n)$ is the transverse mass of the system;
- $E/m(o_1 \dots o_n)$ is the ratio of the total energy to the invariant mass;
- $C(o_1 \dots o_n)$, the centrality, is the ratio of H_T to $\sum E_T$.

For two systems of objects s_1 and s_2 (using a comma in between):

- $\Delta R(s_1, s_2)$ is the angular distance;
- $\Delta p_T(s_1, s_2)$ is the p_T difference;
- $\Delta \phi(s_1, s_2)$ is the ϕ difference.

The selection of the variables for the BDT training is done as follows:

1. Rank the variables by descending separation power S .
2. Keep only one variable with the highest S if several variables are highly correlated (linear correlation above 70 %). An exception is accepted when the correlation shows a large difference

- (> 10 %) between signal and background samples. After this step, the number of variables are largely reduced to a few dozen in each region.
3. Append variables one by one from the top of the list and repeat training. If the S of the newly trained BDT response increases more than 5 % (10 %) compared with previous response, or it exceeds more than 1 % (2 %) of the maximum S ever emerged in the 1j (2j) region, the new variable is kept.
 4. Finalise the collection with fewer variables if different collections give similar S (difference smaller than 1 %).

5.2.2 Hyperparameters optimisation

After the variable selection, a set of BDT parameters are optimised in a sequential procedure. These parameters include:

- *NTrees*, the number of trees used in the boosting;
- *Shrinkage*, the learning rate in the boosting;
- *MaxDepth*, the maximal layers of a decision tree;
- *nCuts*, the number of grid points in the variable range scanned when finding the optimal cut in the node splitting;
- *MinNodeSize*, the minimum percentage of training events in a leaf node;
- *BaggedSampleFraction*, relative size of bagged event sample to the original size of the training sample.

A certain range is scanned for each of the hyperparameters, and the S of the training response is examined. The value that gives the highest S is kept. An accepted value should also pass the overtraining check, which is done by computing the Kolmogorov–Smirnov (KS) [188, 189] probabilities for the training and test distributions. High KS probabilities suggest that two distributions share the similar underlying distributions, while low values imply that they differ. Both the signal KS and background KS probabilities are required to be higher than 10 % to avoid overtraining.

After the optimisation of the hyperparameters, the variables might not be the optimal any longer. Therefore, variable optimisation and hyperparameters optimisation are iteratively reprocessed until the result does not change.

5.2.3 BDT performance

The final hyperparameter values used are summarised in Table 5.1. *NTrees* are tested from 20 to 400. Due to its larger signal portion (~ 18 %), the 1j 1b region uses higher number of trees to capture the characteristics encoded in the input variables. As the signal sample has only 3 % of its events in the 2j2b region, fewer trees are ensembled to reduce the risk of overtraining. The *MinNodeSize*, together with the *BaggedSampleFraction*, gives a minimum of 22 events per leaf node in the extreme case², which is statistically sufficient. *Shrinkages* are found to be small (< 0.3) such that they provide good performance in general [187]. Varying *nCuts* values do not noticeably modify the performance; therefore, smaller values are preferred to reduce the computing. *MaxDepth* is scanned among 2, 3 and 4; such small trees tend to not easily be overtrained and thus no further regularisation techniques, for instance pruning, are needed.

² About 7 500 tW events in the 2j2b region before applying event weights multiplies the *MinNodeSize* and the *BaggedSampleFraction*.

Region	$NTrees$	$MinNodeSize$ [%]	$BaggedSample$ Fraction	$Shrinkage$	$nCuts$	$MaxDepth$	No. of variables
Range	20–400	0.5–10	0.05–1	0.01–0.2	5–100	{2,3,4}	
Step	20	0.5	0.05	0.01	5	—	
1j1b	300	0.5	0.40	0.03	20	4	14
2j1b	280	2.0	0.15	0.14	5	3	11
2j2b	100	0.5	0.60	0.19	5	4	11

Table 5.1: The final settings of the hyperparameters of the BDTs in different regions. Initial ranges and scanning steps are also included.

Variable lists

The selected variable lists for each region are presented in Table 5.2. They are ranked by their separation power. In the 1j1b region, the combination of ℓ_1 , ℓ_2 , j_1 , and E_T^{miss} reveals significant separations, appearing till the seventh variable (except for the third one). The most powerful variable is $p_T(\ell_1 \ell_2 j_1 E_T^{\text{miss}})$, which is expected to be zero for tW events but non-zero for $t\bar{t}$ events due to the second/missing jet. In the two-jet regions, the sub-leading jet j_2 in tW is either from loop contributions or PU; therefore, it is softer compared with the one from a top-quark decay in $t\bar{t}$. Due to the distinctive feature, j_2 shows up in all variables in both two-jet regions. Moreover, at least one of $m(\ell_1 j_2)$ and $m(\ell_2 j_2)$ in $t\bar{t}$ events has a constraint due to the top-quark mass if the lepton-jet pair comes from the same top-quark decay, while such a constraint does not persist in tW events. The W from $t\bar{t}$ also has a top-quark mass threshold, resulting in a softer lepton production. This information is encoded in $\Delta p_T(\ell_1 \ell_2, j_2)$, which is kept in both regions.

The MC modelling of the variables used in the BDTs and their correlations are carefully checked to avoid any poorly described variables. Figure 5.3 shows the two most important variables in each region. The rest of the variables are attached in Appendix B. The MC modelling is compatible with data, increasing the confidence in the BDT training results.

Variable correlations

Figure 5.2 shows the correlation of each variable pair in signal (left) or background (right) sample. Bin numbers correspond to the the rank of the variables, which can be read off in Table 5.2, and which, in the following discussions, are denoted by underlined numbers in the parentheses.

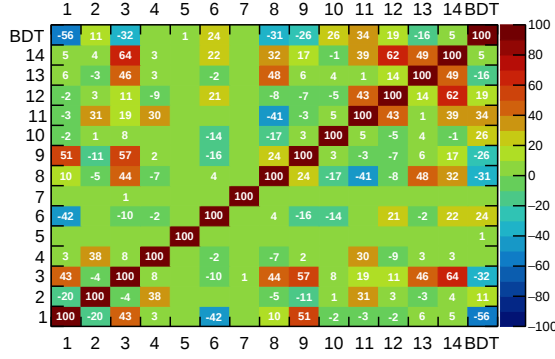
High correlations persist among the same type of variables. For example, in the 1j1b region, p_T (1, 9) are highly correlated, as are Δp_T (2, 4, 11) and m (8, 13, 14). $\sum E_T$ (3) has high correlations with p_T variables such as (1, 9). In the 2j1b region, (2, 4, 11), (3, 7), and (5, 6, 9) are highly correlated among each set, as they represent ΔR , m , and p_T collections, respectively. $\sum E_T$ (10) again shows large correlation with p_T (5, 6, 9). In the 2j2b region, four mass variables (1, 5, 9, 10) are highly correlated.

Variables calculated using the same objects also show a certain degree of correlation. For instance, a high correlation between (12, 14) in the 1j1b region comes from the fact that they are constructed by l_1 and j_1 . The same argument holds for (3, 11) in the 2j1b region where both variables are derived from l_1 and j_2 ; or for (8, 10) in the 2j2b region.

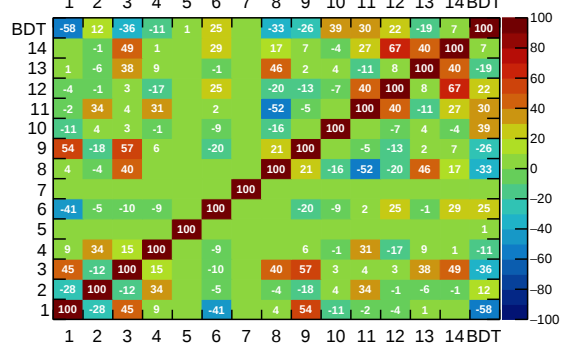
1j1b		S	Figure
R	Variable	$[10^{-2}]$	reference
1	$p_T(\ell_1 \ell_2 j_1 E_T^{\text{miss}})$	4.8	Fig. 5.3(a)
2	$\Delta p_T(\ell_1 \ell_2, j_1 E_T^{\text{miss}})$	2.8	Fig. 5.3(b)
3	$\sum E_T$	2.8	Fig. 4.6(e)
4	$\Delta p_T(\ell_1 \ell_2 j_1, E_T^{\text{miss}})$	2.7	Fig. 4.6(f)
5	$\Delta \phi(\ell_1 \ell_2 j_1, E_T^{\text{miss}})$	1.7	Fig. B.1(a)
6	$\Delta R(\ell_1 \ell_2, j_1 E_T^{\text{miss}})$	1.6	Fig. B.1(b)
7	$\eta(\ell_1 \ell_2 j_1 E_T^{\text{miss}})$	1.4	Fig. B.1(c)
8	$m(\ell_2 j_1 E_T^{\text{miss}})$	1.2	Fig. B.1(d)
9	$p_T(\ell_1 j_1 E_T^{\text{miss}})$	1.2	Fig. B.1(e)
10	$C(\ell_1 \ell_2)$	1.1	Fig. B.1(f)
11	$\Delta p_T(\ell_1 \ell_2, E_T^{\text{miss}})$	1.0	Fig. B.2(a)
12	$\Delta R(\ell_1, j_1)$	0.6	Fig. B.2(b)
13	$m(j_1)$	0.5	Fig. B.2(c)
14	$m(\ell_1 j_1)$	0.1	Fig. B.2(d)
BDT response		8.6	Fig. 5.6(a)

2j1b			S	Figure	2j2b			S	Figure
R	Variable		$[10^{-2}]$	reference	R	Variable		$[10^{-2}]$	reference
1	$\Delta p_T(\ell_1 \ell_2, j_2)$		1.9	Fig. 5.3(c)	1	$m(\ell_1 j_2)$		4.0	Fig. 5.3(e)
2	$\Delta R(\ell_1 \ell_2, j_1 j_2 E_T^{\text{miss}})$		1.7	Fig. 5.3(d)	2	$\Delta p_T(\ell_1 \ell_2, j_2)$		3.4	Fig. 5.3(f)
3	$m(\ell_1 j_2)$		1.4	Fig. B.3(a)	3	$p_T(j_2)$		2.9	Fig. B.4(a)
4	$\Delta R(\ell_1 \ell_2, j_1 j_2)$		1.3	Fig. B.3(b)	4	$p_T(j_1 j_2)$		2.9	Fig. B.4(b)
5	$p_T(j_2)$		1.0	Fig. 4.7(d)	5	$m(\ell_1 \ell_2 j_1 j_2)$		2.4	Fig. 4.8(d)
6	$p_T(j_1 j_2)$		0.9	Fig. B.3(c)	6	$\Delta R(\ell_1 \ell_2, j_1 j_2)$		2.2	Fig. B.4(c)
7	$m(\ell_1 j_1 j_2)$		0.9	Fig. B.3(d)	7	$p_T(\ell_2 j_1 j_2 E_T^{\text{miss}})$		1.2	Fig. B.4(d)
8	$\sigma(p_T)(\ell_1 \ell_2 j_1 j_2 E_T^{\text{miss}})$		0.9	Fig. B.3(e)	8	$\Delta R(\ell_2, j_2)$		1.1	Fig. B.4(e)
9	$p_T(\ell_2 j_1 j_2 E_T^{\text{miss}})$		0.8	Fig. B.3(f)	9	$m(\ell_2 j_2 E_T^{\text{miss}})$		0.8	Fig. B.4(f)
10	$\sum E_T(j_2 E_T^{\text{miss}})$		0.6	Fig. B.3(g)	10	$m(\ell_2 j_2)$		0.7	Fig. B.4(g)
11	$\Delta R(\ell_1, j_2)$		0.5	Fig. B.3(h)	11	$E/m(\ell_1 \ell_2 j_1 j_2)$		0.6	Fig. B.4(h)
BDT response			7.3	Fig. 5.6(b)	BDT response			16.2	Fig. 5.6(c)

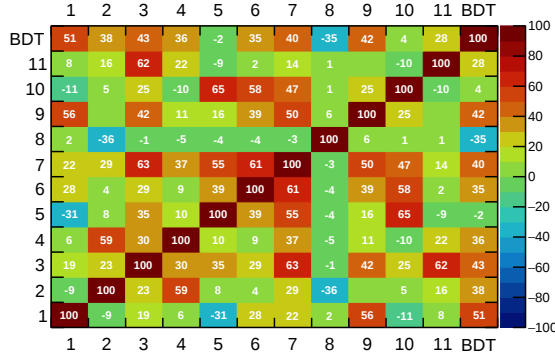
Table 5.2: The variables used in the 1j1b region (top), in the 2j1b region (bottom left), and in the 2j2b region (bottom right), and their rank R , separation power S , and figure reference. The variables are derived from the four-momenta of the leading lepton (ℓ_1), sub-leading lepton (ℓ_2), the leading jet (j_1), sub-leading jet (j_2), and E_T^{miss} . The last row gives the separation power of the BDT response.

Correlation Matrix (tW) 1j1b

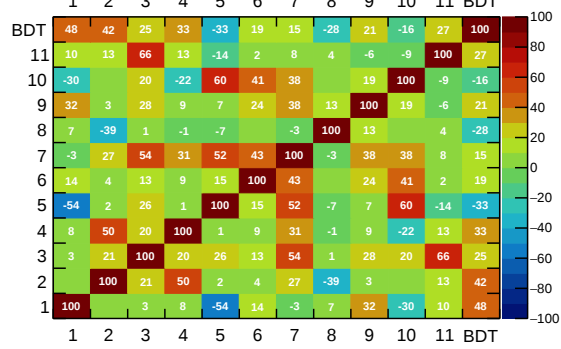
(a)

Correlation Matrix ($t\bar{t}$) 1j1b

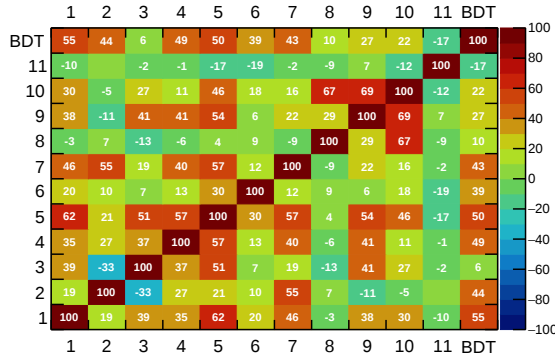
(b)

Correlation Matrix (tW) 2j1b

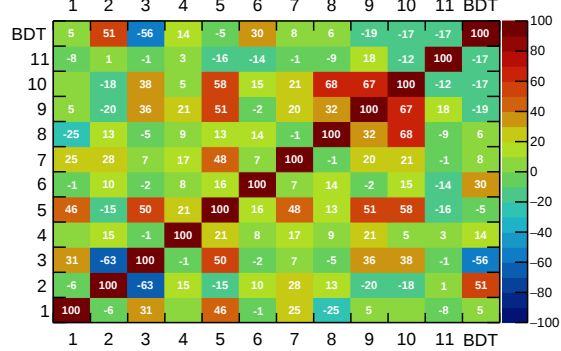
(c)

Correlation Matrix ($t\bar{t}$) 2j1b

(d)

Correlation Matrix (tW) 2j2b

(e)

Correlation Matrix ($t\bar{t}$) 2j2b

(f)

Figure 5.2: Correlation matrices of BDT input variables in the tW sample (left column), and in the $t\bar{t}$ sample (right column) for the three regions. Numbers are shown in percent. Numerical variable label stands for the R of the variable, which can be read off from Table 5.2.

In the 2j1b and 2j2b regions, the information of l_1 and j_1 (especially their p_T) plays an important role. Although the p_T variables of the single object are removed by the linear correlation criteria, the information is preserved in the combination of other variables. In the 2j1b region, for instance, (5, 6) give an access to the $p_T(j_1)$ information while (9) is complementary to $p_T(l_1)$ (as it is built from all other objects). From the correlation matrices in the 2j1b region (Figs. 5.2(c) to 5.2(d)), the tW BDT response correlates differently with (5, 6) than that of $t\bar{t}$ sample. The difference also appears with (9): it has 42 % correlation in tW sample with BDT response, while 21 % in $t\bar{t}$ sample. Similar phenomena emerge in the 2j2b region. $p_T(j_1)$ is accessed via (3, 4), while (7) is the complement of $p_T(l_1)$. Information of l_1 and j_1 is not only presented in the above variables but also in the rest of variables. This can be seen from their correlation with the BDT response, e.g. in the tW samples. In the 2j1b region, the BDT response shows quite low correlation with (5, 10) (Fig. 5.2(c)), which contain neither l_1 nor j_1 , while the rest have much stronger correlations. In the 2j2b region, the BDT response shows similar behaviour such as with (3) in tW sample (Fig. 5.2(e)). Different correlations in tW and $t\bar{t}$ samples help BDTs to better separate them.

BDT response

The ROC curves of each region are given in Fig. 5.4 and the AUC values are around 0.65 to 0.73 in different regions. The substantially larger values of AUC in the 2j2b are due to the characteristic of the j_2 as just mentioned,³ which is also reflected by the separation power S of the BDT response in Table 5.2.

Overtraining is checked in Fig. 5.5. KS test probabilities are printed on the plots, which are required to be larger than 10 % as aforementioned. Signal tW is shifted to large values of the BDT response, while background $t\bar{t}$ tends to sit on the other side. The shapes of test samples are close to the training shapes, implying no obvious overtraining. The stacked BDT responses are illustrated in Fig. 5.6, with the same range and binning as the one used in the final fit (Section 6.4). Overall, the prediction agrees very well with data, and the discrepancy is covered by the systematic uncertainties.

³ The additional b -jet requirement enhances the separable behaviour in that phase space, although the tagging information is not directly used in the BDT.

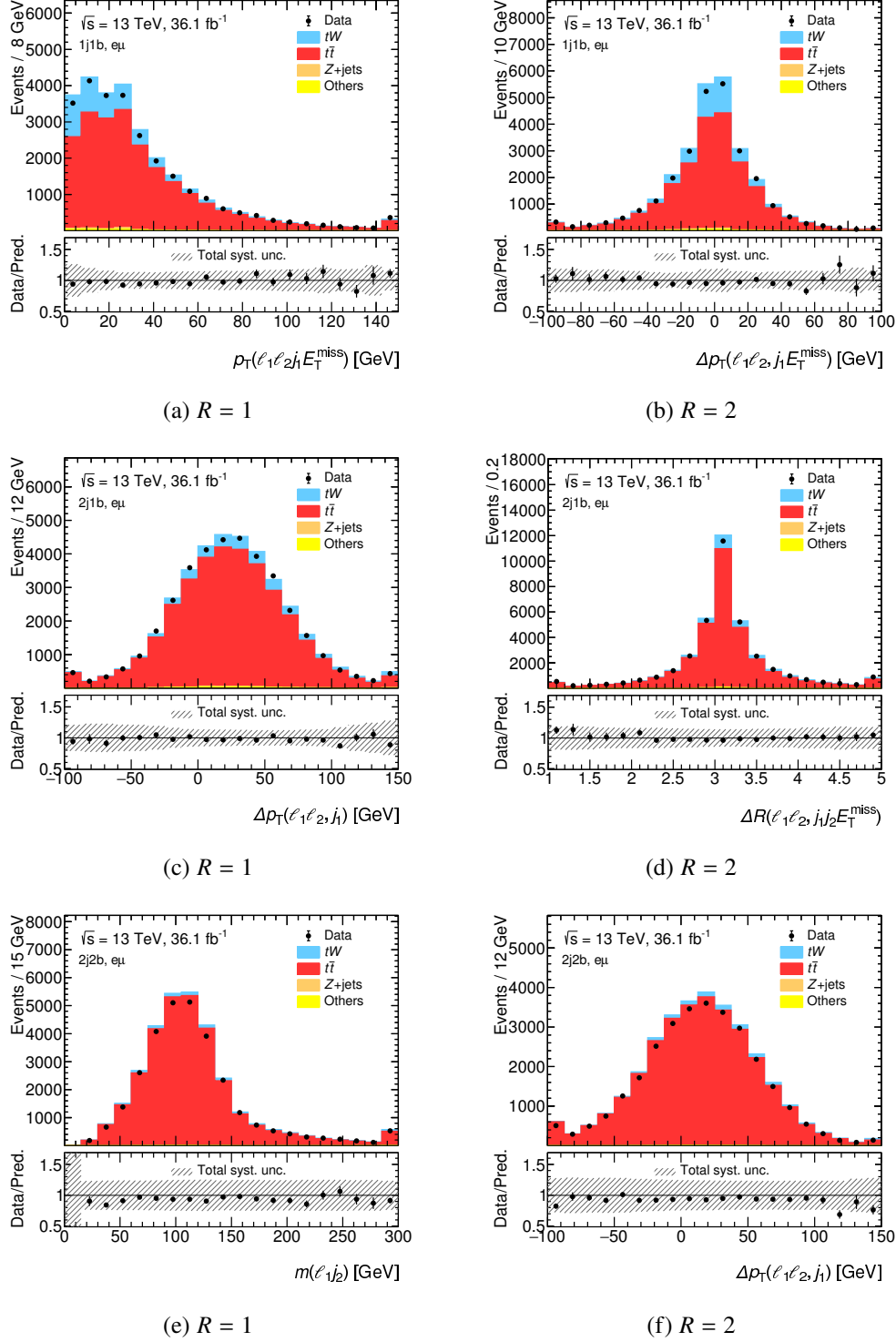


Figure 5.3: Top two BDT input variables used in the 1j1b region (top row), 2j1b region (middle row) and 2j2b region (bottom row). Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panel gives the yields in the number of events per bin, while the lower panel gives the ratios of the numbers of observed events to the total prediction in each bin. The sub-caption gives the rank R defined in Table 5.2.

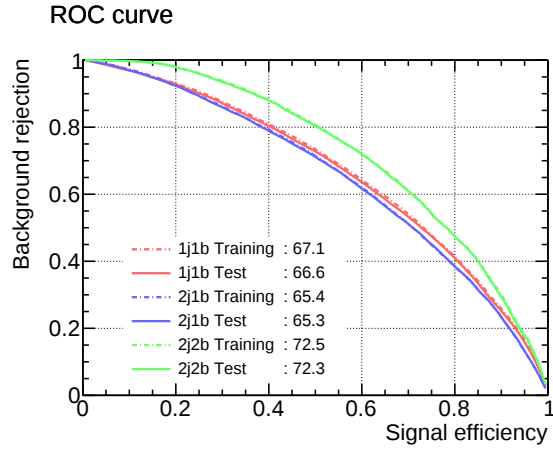
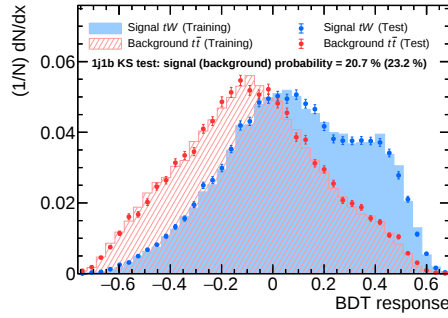


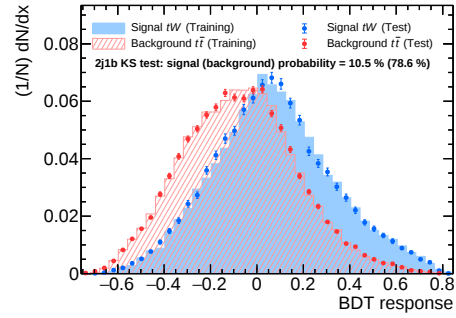
Figure 5.4: ROC curves of 1j1b in red, 2j1b in blue, and 2j2b in green. Dashed lines are from training samples while solid lines are from test samples. The AUC values are given in percent in the legend.

Overtraining check



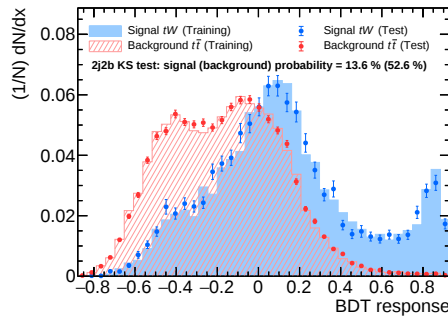
(a)

Overtraining check



(b)

Overtraining check



(c)

Figure 5.5: Comparison of training and test samples in the 1j1b (a), 2j1b (b), and 2j2b (c). Statistical uncertainties are added for the test shapes. All the shapes are normalised to the unit. A KS test is performed on each pair of training and test shapes.

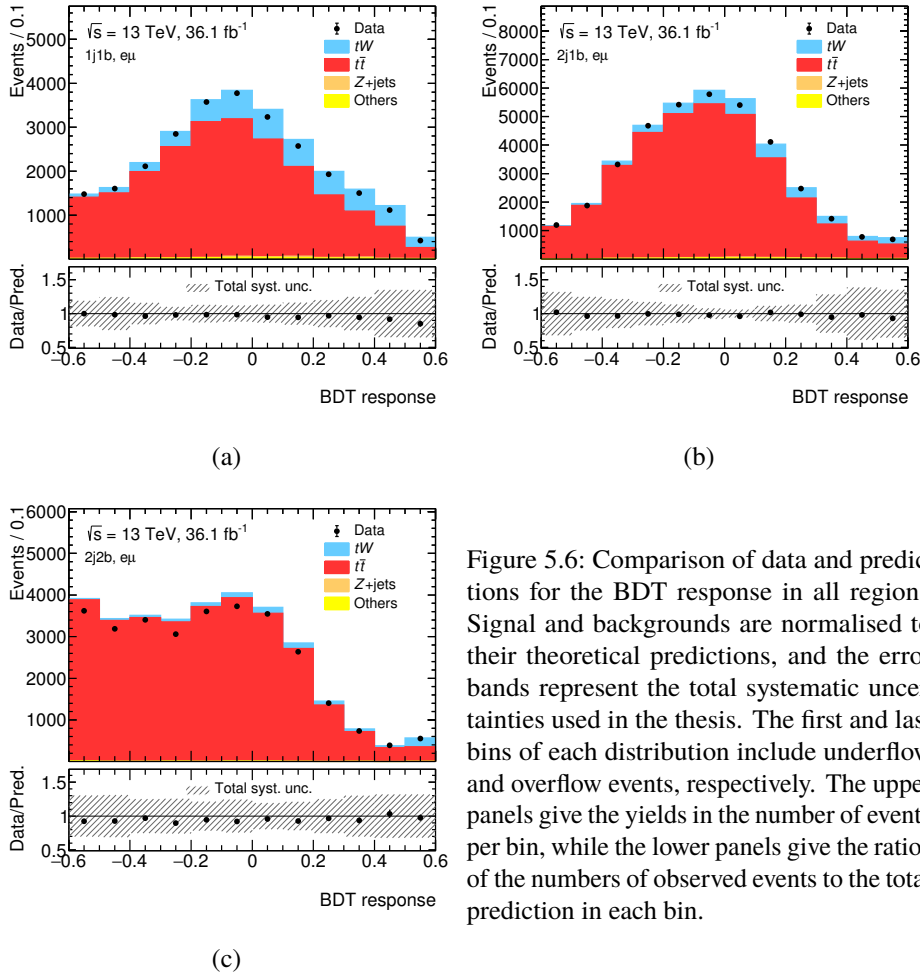


Figure 5.6: Comparison of data and predictions for the BDT response in all regions. Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panels give the yields in the number of events per bin, while the lower panels give the ratios of the numbers of observed events to the total prediction in each bin.

Uncertainty estimation and statistical analysis

This chapter describes the uncertainty estimation of the measurement and the statistical analysis. Section 6.1 introduces some basic ideas on the measurement errors. The maximum likelihood method is used to extract signal yields, which is introduced in Section 6.2. Section 6.3 details sources of uncertainty used in the analysis while Section 6.4 presents the results of the inclusive cross-section measurement.

6.1 General concepts

Measurements are accompanied by uncertainties, or *errors*, which indicate the accuracy of measurements. The uncertainties are conventionally divided into statistical and systematic uncertainties.

Statistical uncertainty describes the stochastic fluctuations of a finite sample size used in the measurement. They are uncorrelated among independent samples. Enlarging the sample size will reduce the statistical uncertainty.

Systematic uncertainty stems from implicit or explicit assumptions/approximations in the measurement, or from the theoretical models. It includes contributions from many different sources. Each source of uncertainty is correlated among independent samples if these samples share the same assumption, or if they are generated by the same model. They come along with each step of the measurement and propagate downstream. The propagation must satisfy the following coverage probability:

$$P(\hat{y} - \sigma_y < y < \hat{y} + \sigma_y) = P(\hat{x} - \sigma_x < x < \hat{x} + \sigma_x).$$

$\hat{y}$ and σ_y denote the final measured observation and its standard deviation, respectively; $\hat{x}$ and σ_x denote the source quantity and its standard deviation, respectively.

If the outcome of the measurement is a continuous distribution, the uncertainties can be split into two components: the rate uncertainty that is an error of the total number of events, and the shape uncertainty that distorts the distribution.

6.2 Maximum likelihood method

The goal of the measurement is to extract the signal cross-section from the observed events. The task can be encoded into a statistical model where the signal cross-section is represented by a parameter. A

maximum likelihood fit of the model to data is used to obtain the parameter estimate.

6.2.1 Multi-region and multi-bin likelihood function

Binned likelihood model

Consider the Poisson distribution that describes event yields n in an experiment given the true unknown yields, λ :

$$\mathcal{P}(n|\lambda) = \lambda^n \frac{e^{-\lambda}}{n!},$$

If an observable x follows a distribution $f(x)$, the likelihood of the parameter λ can be expressed as

$$\mathcal{L}(\lambda) = \mathcal{P}(n|\lambda) \prod_{\text{event}}^n f(x).$$

When several disjoint regions are available, such as multiple signal and control regions, the likelihood is the multiplication of the likelihood in each region,

$$\mathcal{L}(\lambda_r) = \prod_{r \in \text{regions}} \left[\mathcal{P}(n_r|\lambda_r) \prod_{\text{event}}^n f(x) \right].$$

The probability density function $f(x)$ is usually represented by histograms where the observable x is *binned* to discrete values. The event yields in each bin of each region are Poisson distributed, making the likelihood a product of multiple Poisson distributions. Regions and bins are commutable and are treated uniformly. This is sometimes referred to as the *binned likelihood*:

$$\mathcal{L}(\lambda_{r,b}) = \prod_{r \in \text{regions}} \prod_{b \in \text{bins}} \mathcal{P}(n_{r,b}|\lambda_{r,b}).$$

$n_{r,b}$ is the observed event yields in region r bin b . $\lambda_{r,b}$ is the expectation in that region and bin, and can be factorised as $\lambda_{r,b} = \mu \lambda_{\text{sig}}^{\text{pred.}} + \lambda_{\text{bkg}}^{\text{pred.}}$. Here $\lambda_{\text{sig}}^{\text{pred.}}$ and $\lambda_{\text{bkg}}^{\text{pred.}}$ denote the predictions for signal and background yields from the simulation, respectively; $\mu = \sigma/\sigma_{\text{SM}}$ is the scale factor on the signal cross-section to be determined with respect to the SM prediction σ_{SM} , called the *signal strength*. The measurement of the signal cross-section is translated to a measure of μ .

Parameter estimation and its error

The values of parameters that maximise the likelihood function are the maximum likelihood estimates, denoted with a hat (^) on top of the parameter. The natural logarithm of the likelihood function (sometimes multiplied with a non-unit constant factor), known as the log-likelihood function, is more popular in practice as it converts the multiplication into a summation: (expressed in terms of μ)

$$\ln \mathcal{L}(\mu) = \sum_{r \in \text{regions}} \sum_{b \in \text{bins}} \ln \mathcal{P}(n_{r,b}|\mu \lambda_{\text{sig}}^{\text{pred.}} + \lambda_{\text{bkg}}^{\text{pred.}}).$$

Since the logarithm is a monotonic transformation, the argument that maximises the likelihood also maximises the log-likelihood function. With a negative sign, the whole problem becomes a

minimisation problem, which can make use of various minimisation software programs. Moreover, the negative sign makes it an up-opening shape, emphasising that it is a likelihood not a probability density function.

The estimate $\hat{\mu}$ can be solved via

$$\left. \frac{\partial \ln \mathcal{L}(\mu)}{\partial \mu} \right|_{\mu=\hat{\mu}} = 0.$$

To access to its uncertainty, one can expand the log-likelihood to a Taylor series around $\hat{\mu}$, and ignore higher than second order terms:

$$\begin{aligned} \ln \mathcal{L}(\mu) &= \ln \mathcal{L}(\hat{\mu}) + \left. \frac{\partial \ln \mathcal{L}(\mu)}{\partial \mu} \right|_{\mu=\hat{\mu}} (\mu - \hat{\mu}) + \frac{1}{2} \left. \frac{\partial^2 \ln \mathcal{L}(\mu)}{\partial \mu^2} \right|_{\mu=\hat{\mu}} (\mu - \hat{\mu})^2 + \dots \\ &\simeq \ln \mathcal{L}(\hat{\mu}) + \frac{1}{2} \left. \frac{\partial^2 \ln \mathcal{L}(\mu)}{\partial \mu^2} \right|_{\mu=\hat{\mu}} (\mu - \hat{\mu})^2. \end{aligned}$$

The first-derivative term is also eliminated due to the definition of $\hat{\mu}$. Therefore,

$$\mathcal{L}(\mu) = \mathcal{L}(\hat{\mu}) \cdot \exp \left[\frac{1}{2} \left. \frac{\partial^2 \ln \mathcal{L}(\mu)}{\partial \mu^2} \right|_{\mu=\hat{\mu}} (\mu - \hat{\mu})^2 \right].$$

Since a likelihood is asymptotically Gaussian distributed, the uncertainty of the estimates σ can be extracted from

$$\frac{1}{\sigma^2} = \left. \frac{\partial^2 \ln \mathcal{L}(\mu)}{\partial \mu^2} \right|_{\mu=\hat{\mu}}.$$

Thus¹,

$$\ln \mathcal{L}(\hat{\mu} \pm Z\sigma) \simeq \ln \mathcal{L}(\hat{\mu}) + \frac{1}{2} Z^2.$$

Let $Z = 1$; σ is then accessed by shifting the maximum log-likelihood value down by one half and the intersections correspond to $\hat{\mu} \pm \sigma$.

6.2.2 Incorporating systematic uncertainties

Generalised statistical model

Information from systematic estimation adds constraint terms to the likelihood function:

$$\mathcal{L}(\mu, \vec{\theta}) = \prod_{r \in \text{regions}} \prod_{b \in \text{bins}} \mathcal{P}(n_{r,b} | \lambda_{r,b}(\hat{\mu}, \hat{\vec{\theta}})) \cdot \prod_{s \in \text{systematics}} \mathcal{G}(\hat{\theta}_s | \theta_s).$$

The signal strength μ is called the parameter of interest (POI), and sometimes can be many. θ_s that parametrised the systematic effect from the source s is called the nuisance parameter (NP), and $\vec{\theta}$ is the assembly of all NPs. $\vec{\theta}$ also alter the Poisson expectation $\lambda_{r,b}(\hat{\mu}, \hat{\vec{\theta}})$ and need to be determined but

¹ The conclusion is in fact stronger. The 68.3 % central confidence interval for a non-Gaussian shaped likelihood can be constructed in the same way. See detailed discussions in [190].

are of minor interest. This likelihood function gives the most general statistical model used in the data analyses. The constraint terms chosen in the thesis are Gaussian [191, 192].

Interpolation and extrapolation

Systematic variations are measured at one standard derivation from the nominal, corresponding to $\theta_s = \pm 1$, respectively. For instance, for the luminosity uncertainty, $\theta_s = +1$ increases the event yields by 2.1 %, while $\theta_s = -1$ reduces the yields by 2.1 %. Interpolation and extrapolation of θ_s are necessary for the likelihood evaluation; a 6th-order polynomial interpolation and a linear extrapolation are used in the thesis [191, 192]:

$$\lambda(\theta_s) = \begin{cases} \left(\frac{\lambda(1)}{\lambda(0)}\right)^{\theta_s} & \theta_s > +1, \\ 1 + \sum_{p=1}^6 a_p \cdot (\theta_s)^p & |\theta_s| \leq 1, \\ \left(\frac{\lambda(-1)}{\lambda(0)}\right)^{-\theta_s} & \theta_s < -1. \end{cases}$$

$\lambda(1)$, $\lambda(-1)$, and $\lambda(0)$ are bin-wise yields for variation up, down, and nominal, respectively. Coefficients a_p are fixed by continuous conditions of the function itself and its first and second derivatives.

6.2.3 Hypothesis testing

p-value and discovery significance

Physics discovery is translated to a hypothesis test of whether the background-only hypothesis (null hypothesis $H_0(\mu = 0)$) is rejected, or equivalently, whether the signal-plus-background hypothesis (alternative hypothesis $H_1(\mu = 1)$) is accepted. The probability that a measurement D occurs under the null hypothesis is defined as the *p*-value.² If the *p*-value is less than a chosen threshold, we shall say the null hypothesis is rejected at the chosen level; if it is not the case, the measurement is insufficient to reject the null hypothesis. The *p*-value is calculated with a random variable called the *test statistic*. According to the Neyman–Pearson lemma [193], the likelihood ratio of the null and alternative hypotheses is a good test statistic:

$$\Lambda = \frac{\mathcal{L}(H_0|D)}{\mathcal{L}(H_1|D)}.$$

Furthermore, the *p*-value can be converted to the number of the Gaussian standard deviations, beyond which the cumulated area in the upper tail is equal to the *p*-value. A three sigma significance, corresponding to a *p*-value of 1.35×10^{-3} , conventionally reveals evidence, while a five sigma significance, corresponding to a *p*-value of 2.87×10^{-7} , claims a discovery. To exclude a hypothesis, a threshold of 0.05 (i.e. 95 % confidence level) is often used, corresponding to 1.64σ .

² As a Bayesian you might use the posteriori test and the Bayes factors instead. However, the debate of the frequentist and Bayesian is beyond the scope of the thesis.

Profile likelihood ratio

If the test statistic would rely on the NPs, the hypothesis testing should take into account p -values for all values of the NPs and find out the maximum, which is practically impossible. Therefore, a NP-free test statistic is used. From model $\mathcal{L}(\mu, \vec{\theta})$, a *profile* likelihood ratio $\Lambda(\mu)$ is defined as

$$\Lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\hat{\theta}})} = \frac{\mathcal{L}(\hat{\hat{\theta}}(\mu)|\mu, D)}{\mathcal{L}(\hat{\mu}, \hat{\hat{\theta}}|D)},$$

where $\hat{\hat{\theta}}(\mu)$ denotes the maximum likelihood estimates of the NPs given μ . $\Lambda(\mu)$ explicitly depends only on μ , so the dependence on the NPs is removed via *profiling*.

$\Lambda(\mu)$ is between 0 and 1. To stretch the range to non-negative values, a negative logarithm is taken as the test statistic t_μ , with a factor of 2 that cancels a factor of half coming from the logarithm:

$$t_\mu = -2 \ln \Lambda(\mu).$$

If the distribution of t_μ is known, the p -value is the integration from the observed value to positive infinity.

Toy MC approach and asymptotic approach

The distribution of t_μ can be obtained from two alternative approaches. In the toy MC approach, the histograms generated by the measurement are randomised according to Poisson distributions, and corrected by the systematics contributions converted from Gaussian distributed NPs. Then with the pseudo-experiment, the likelihood function is re-maximised and t_μ is re-calculated. After doing this hundreds of thousands of times, a distribution of t_μ is constructed and the integration can be evaluated.

The toy MC approach is computational expensive. According to the Wilks' theorem [194–196], the distribution of t_μ asymptotically converges to a chi-square distribution for one degree of freedom, with a sufficiently large sample size. Thus, the integration no longer requires generating a massive number of pseudo-experiments, and can be computed much faster and easier.

Pull and impact

The *pull* ρ of a NP θ is defined as [197]

$$\rho = \frac{\hat{\theta} - \langle \theta \rangle}{\sigma_\theta},$$

where $\langle \theta \rangle$ is the expectation and σ_θ is the standard deviation. The pull obeys a standard Gaussian distribution and thus can clarify whether the fit value is pulled from expectation and the uncertainty is under-/over-estimated.

The *impact* of a NP θ on the POI μ is defined as the shift of the POI estimates due to fixing the NP to its up/down variation [198]

$$\Delta\mu^\pm = \hat{\mu}_{\hat{\theta} \pm \sigma_\theta} - \hat{\mu}.$$

The other NPs, in the meantime, can be either floated or fixed to their maximum likelihood estimates.

The floated version usually gives a smaller impact since the other NPs can absorb the impact by a little pull due to correlations. Therefore, the difference in impacts of two schemes unmasks the degree of correlations.

6.2.4 Validation of the statistical model

Before fitting the statistical model to the real collision data, the model is validated on an *Asimov* dataset [196] to obtain the measurement sensitivity. An Asimov dataset is artificially constructed with known POI and NPs, for instance all NPs are set to zero and the POI to one for an SM process or zero for a new physical process. In other words, the Asimov data is the stacked nominal distributions of the SM processes. The fit performed on such a dataset should reflect the input values and reveal the expected uncertainties. It helps to check the behaviour of the fit in terms of constraints, pulls and impacts, and find potential problems or wrong assumptions.

6.3 Systematic uncertainty estimation

Systematic uncertainty estimation needs to be consistent both with the statistical uncertainty and among its different sources. In the ATLAS experiment, every step from event generation to the final object reconstruction is inspected and a systematic uncertainty is assigned.

6.3.1 Theory uncertainties

In simulation, the choices of generators and their parameters are empirical. Different choices lead to different observable spectra, which will affect the measurement. A general procedure to estimate the impact is to replace a generator by another one or vary the values of the parameters, re-simulate the full sample, and then evaluate the effect on the final measurement.

Matrix Element (ME)

Computation of the ME at NLO involves subtraction of one-loop virtual contributions and the emission of additional soft real partons, to cancel divergences. Two frameworks are available: POWHEG and AMC@NLO. For both the tW and $t\bar{t}$ processes, the POWHEG+HERWIG++ and AMC@NLO+HERWIG++ with AF2 simulation are compared (the ME sample vs the PS sample in Tables 4.2 to 4.3). Event weights are derived from the difference on BDT responses between the two samples and are applied to the nominal sample. The difference is taken as a one-sided uncertainty, and then symmetrised to the other side. This uncertainty is uncorrelated between tW and $t\bar{t}$.

Parton Showering (PS)

The uncertainty on PS and hadronisation is evaluated by comparing the cluster model implemented in HERWIG, and the Lund string model implemented in PYTHIA. For both the $t\bar{t}$ and tW processes, the POWHEG+PYTHIA6 and POWHEG+HERWIG++ with AF2 simulation samples are compared (AF2 nominal vs the PS sample in Tables 4.2 to 4.3). The variation is symmetrised. This uncertainty is correlated between tW and $t\bar{t}$.

Initial- and final-state radiation (ISR/FSR)

POWHEG-BOX+PYTHIA samples using a dedicated P2012 tune with two different radiation levels are generated in order to evaluate the effect of different radiation emission. The high radiation sample uses the P2012radHi tune with factorisation scale 2.0 and renormalisation scale 2.0. The low radiation sample with P2012radLo tune sets these factors to 0.5. They are generated for both the tW and $t\bar{t}$ processes. Event weights are derived from the difference of BDT responses between the two samples and applied to the nominal sample. Half of the difference is taken as the uncertainty and symmetrised. AF2 simulation is used for the tW process while FULLSIM simulation is used for $t\bar{t}$ (see Tables 4.2 to 4.3). This uncertainty is uncorrelated between tW and $t\bar{t}$.

DR vs DS

The results obtained using the DR and DS schemes implemented in POWHEG+PYTHIA6 are evaluated and the difference is symmetrised.

PDF

The uncertainty due to the choice of PDF is evaluated using the PDF4LHC15 combined PDF set [199] with 30 symmetric eigenvectors. The tW and $t\bar{t}$ AMC@NLO+HERWIG++ samples are reweighted with the central value and eigenvectors of the combined PDF set to construct a set of varied systematic templates. The full difference between the central value of the PDF4LHC15 prediction and the central AMC@NLO+HERWIG++ with CT10 prediction is symmetrised and taken as an uncertainty. Each additional eigenvector is then compared with the central PDF4LHC15 prediction, symmetrised, and taken as an additional uncertainty. The dominant uncertainty is the difference in central values, though overall the PDF uncertainty is a relatively small effect on the final result. Therefore, the PDF uncertainty is negligible in the thesis.³

6.3.2 Detector and algorithm performances uncertainties

Uncertainties arising from the imperfect detector and reconstruction algorithms are derived and studied by the various performance groups in ATLAS. These uncertainties are correlated between tW and $t\bar{t}$ samples.

Lepton efficiencies

The mis-modelling of the electron and muon reconstruction/isolation efficiencies in the MC simulation is corrected by scale factors including

- reconstruction scale factors (reco. SF),
- identification scale factors (ID SF),
- isolation efficiency scale factors (isol. SF),
- trigger scale factors (trigger SF).

Varying each factor within its uncertainty and re-performing the measurement gives an estimation of the uncertainties from that source.

³ The PDF uncertainty is studied and included in the published results of the measurements. Overall it contributes < 3 % to the total uncertainties.

Lepton energy scale and resolution

Uncertainties in the muon energy scale and resolution are negligible and therefore are not taken into account, while those in the electron energy scale (E_e scale) and resolution (E_e resolution) are considered. They stem from the uncertainties in the lepton energy scale calibration.

JES and JER

As introduced in Section 3.3.5, 21 sources [200] are considered in the thesis. They include

- 8 sources from the in-situ calibration (Eff1–8),
- 3 sources from η -intercalibration: modelling (η model), statistics/method (η stat.) and calibration non-closure (η noncl.),
- 4 sources from the PU (PU μ , PU NPV, PU p_T , PU ρ),
- 1 source for high- p_T jets, from the propagation of single hadron calibration (single particle),
- 1 source from the flavour composition (flav. comp.),
- 1 source from the flavour response (flav. resp.),
- 1 source for b -jets (BJES),
- 1 source for punch-through jets (punch-through),
- 1 source for the JER.

b -tagging

The uncertainties due to the identification of b -jets are estimated by varying the efficiency corrections in the MC simulation. They include

- 6 independent eigenvectors for the b -jets efficiency (b0–b5),
- 3 independent eigenvectors for the c -jets efficiency (c0–c2),
- 16 independent eigenvectors for the light-jets efficiency (light0–light15),
- 2 factors for the extrapolation uncertainty (extrap. and extrap. c).

These 27 variations are filtered from a list of about 100 variations by removing variations with negligible impact. These uncertainty factors alter the b -tagging scale factor on a per-event basis.

E_T^{miss} calculation

The hard objects (leptons and jets) energy scale and resolution uncertainties are propagated to the E_T^{miss} through the re-computation of its corresponding terms. Therefore the impact from the hard objects is evaluated when shifting lepton and jet energy. The soft track related uncertainties, however, need to be considered. The p_T balance between hard and soft E_T^{miss} components [201] are examined on data and MC, and deviations are treated as uncertainties. Three sources, with respect to the vector sum of hard p_T objects (p_T^{hard}), are:

- the soft-track p_T (p_T^{soft}) scale shift along the p_T^{hard} axis (MET scale),
- the p_T^{soft} resolution smearing along the p_T^{hard} axis with the p_T^{soft} positive scale shift being applied⁴ (MET reso. ||),
- the p_T^{soft} resolution smearing perpendicular to the p_T^{hard} axis with the p_T^{soft} positive scale shift being applied (MET reso. $\perp$).

⁴ The p_T^{soft} positive scale produces a better data-MC agreement. Hence the p_T^{soft} resolutions are estimated on the positive scaled sample to avoid double-counting.

Luminosity

The central value and uncertainty of the measured integrated luminosity are estimated by Van-der-Meer scans [202], performed in August 2015 and May 2016. A variation of 2.1 % on event yields for signal and all background samples was applied. Luminosity uncertainty introduces only rate uncertainty.

6.3.3 Cross-section normalisation on small backgrounds

Additionally, normalisation uncertainties on yields for small backgrounds are estimated. A 100 % uncertainty is assumed for the fake-lepton backgrounds. An uncertainty of 50 % is assigned to the Z + jets backgrounds with one b -tagged jet, while a 100 % uncertainty is assumed for Z + jets events with two b -tagged jets. These uncertainties are chosen to be consistent with previous ATLAS measurements of these processes. For diboson backgrounds, an uncertainty of 25 % is assigned to cover the difference between the predictions of the SHERPA and POWHEG-Box generators. These uncertainties are treated as uncorrelated across the various fit regions.

6.4 Application of maximum likelihood fit

A binned likelihood model is used in the fit to extract the cross-section for tW .

6.4.1 Simultaneous fit

Fits are simultaneously performed in three different regions for both Asimov data and real data. The $t\bar{t}$ normalisation factor is floated in the fit, to be constrained by the 2j2b control region, while the other normalisation factors are fixed with uncertainties set to the values discussed above. The BDT distributions before and after fit are shown in Figs. 6.1 to 6.2 for Asimov data and real data, respectively. For Asimov data, the ratios of the data and the prediction are lined at one in the prefit plots by definition, and the fit gives out the input values so they are again lined at one in the postfit plots. The total uncertainties are reduced after fit due to the constraint from (Asimov) data. Note that the scales of the ratio panels before and after fit are different.

The predicted and fitted yields for each process and the observed yields are shown in Table 6.1. The “Pred. events” are the total predictions of all processes and the errors are the quadratic sum of statistical uncertainty and all systematic uncertainties. The fitted events of individual processes are calculated by the interpolation and extrapolation method described earlier, assuming no correlations among NPs. The tW and $t\bar{t}$ fitted yields are adjusted according to the signal strength and $t\bar{t}$ normalisation factor, respectively. The errors in each process are calculated from the errors of the POIs and the NPs. Note that the errors of POIs only exist after fit.⁵ However, because of the correlations among the POIs and the NPs, it makes more sense to quote one number than to split up to two parts. The “Fitted events” are the summation over all processes. Due to many negative correlated systematic uncertainties among the processes (constrained by the total yields), the errors on the “Fitted events” are largely reduced, which can also be seen in Figs. 6.1 to 6.2.

The maximum likelihood fit gives the following results, where fitting to Asimov (expected) data and real (observed) data.

⁵ In fact 1 % on POIs is assumed before fit but is negligible.

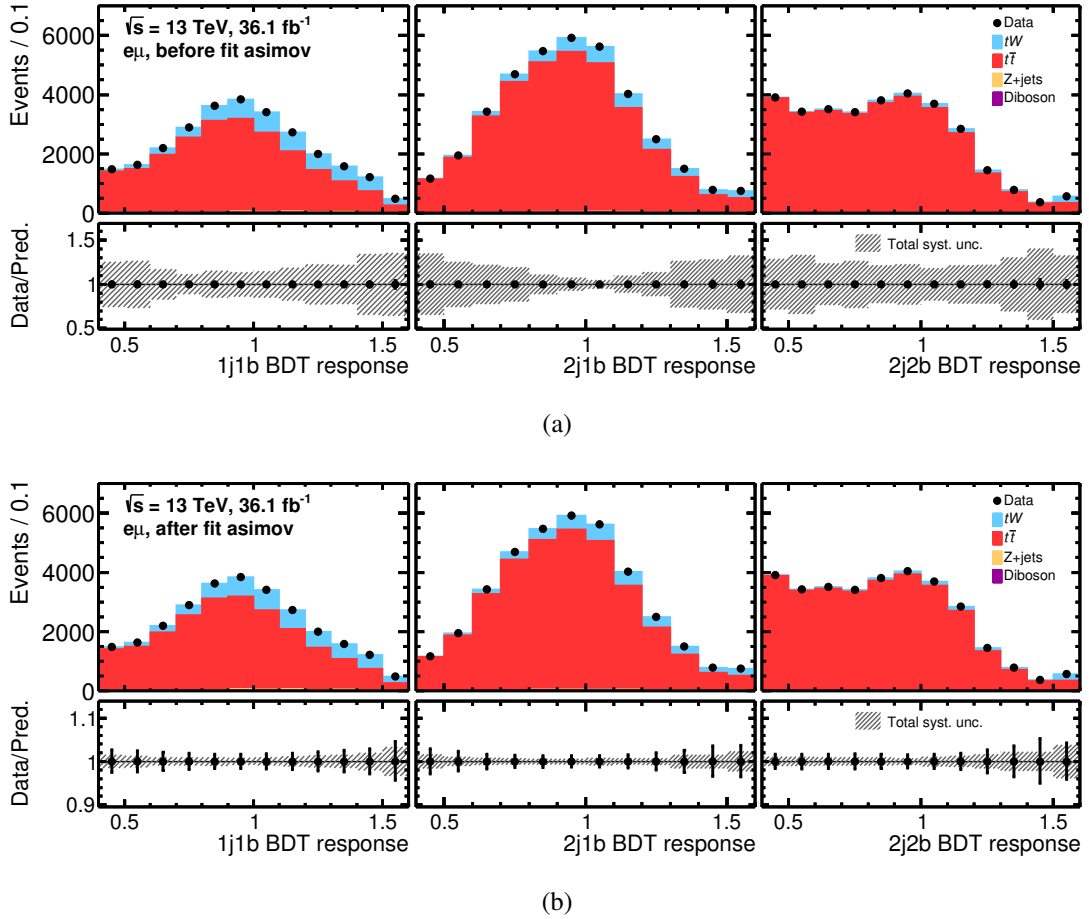
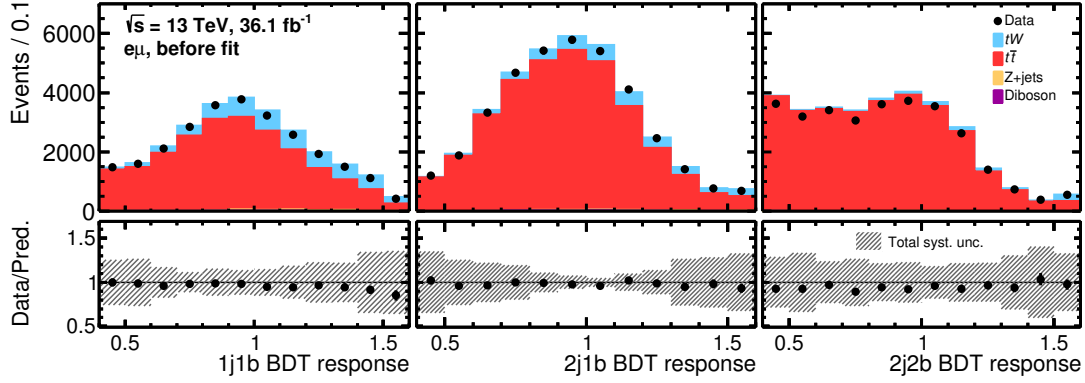
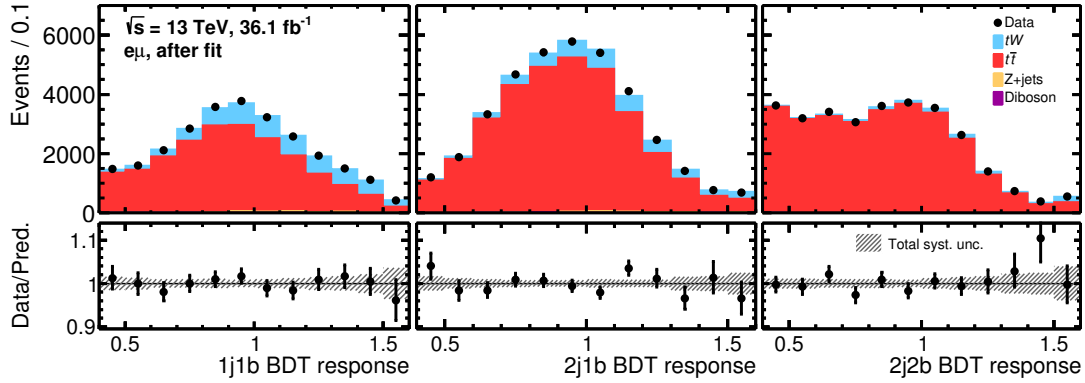


Figure 6.1: BDT response in three regions 1j1b, 2j1b, and 2j2b before (a) and after (b) fit to the Asimov dataset. Vertical bars on data represent the statistical uncertainties and the error bands represent the total systematic uncertainties. The upper panels give the yields in the number of events per bin, while the lower panels give the ratios of the numbers of Asimov data events to the total prediction in each bin.



(a)



(b)

Figure 6.2: BDT response in three regions 1j1b, 2j1b, and 2j2b before (a) and after (b) fit to real data. Vertical bars on data represent the statistical uncertainties and the error bands represent the total systematic uncertainties. The upper panels give the yields in the number of events per bin, while the lower panels give the ratios of the numbers of observed events to the total prediction in each bin.

	1j1b	2j1b	2j2b
Pred. events	27100 ± 4400	37800 ± 5100	31900 ± 7700
Pred. tW events	4910 ± 800	3370 ± 640	1080 ± 290
Pred. $t\bar{t}$ events	21700 ± 3700	34000 ± 4900	30700 ± 7500
Pred. Z + jets events	152 ± 76	116 ± 58	10 ± 10
Pred. diboson events	169 ± 42	168 ± 42	9.5 ± 2.4
Pred. fake events	120 ± 120	180 ± 180	92 ± 92
Observed events	26171	37147	29874
Fitted events	26150 ± 200	37160 ± 230	29890 ± 200
Fitted tW events	5300 ± 1000	3900 ± 680	1050 ± 140
Fitted $t\bar{t}$ events	20300 ± 1000	32820 ± 760	28740 ± 270
Fitted Z + jets events	154 ± 75	99 ± 53	10 ± 10
Fitted diboson events	168 ± 42	162 ± 41	10.0 ± 2.4
Fitted fake events	130 ± 120	140 ± 140	79 ± 79

Table 6.1: Predicted event yields and fit results. The errors on predicted events include the statistical and systematic uncertainties before fit, assuming no correlations among systematic uncertainties or among processes. The errors on fitted events include the statistical and systematic uncertainties after fit, propagated from the errors in the POIs and the NPs. Correlations are taken into account among systematic uncertainties and among processes.

	Expected	Observed
μ_{sig}	$1.00^{+0.20}_{-0.17}$	$1.14^{+0.26}_{-0.22}$
$\mu_{t\bar{t}}$	$1.000^{+0.032}_{-0.031}$	$1.004^{+0.033}_{-0.032}$

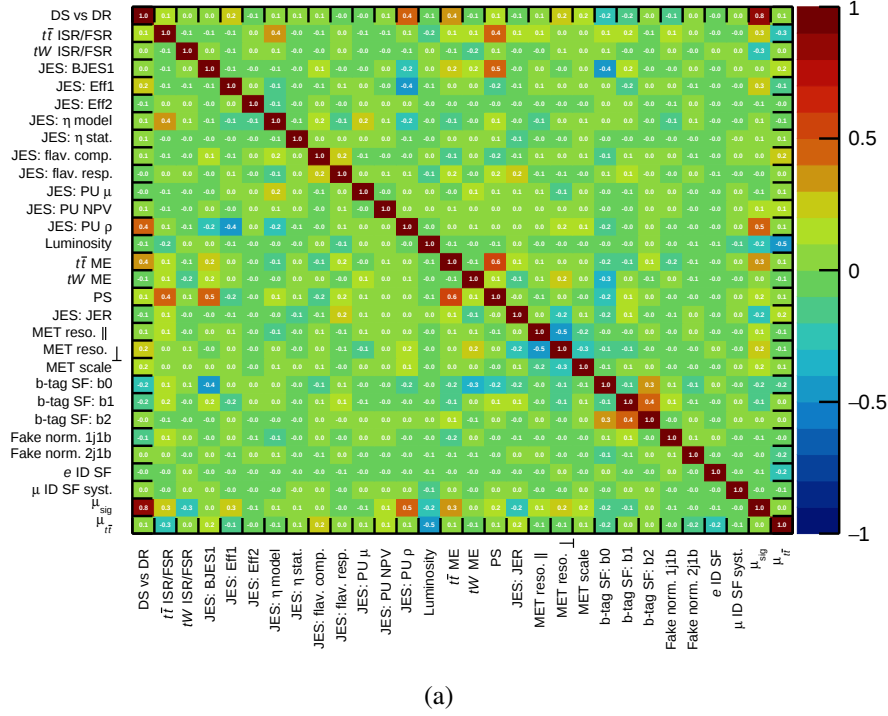
The observed signal strength corresponds to a measured cross-section of $\sigma_{tW} = 82^{+19}_{-16}$ pb.

Table 6.2 shows the breakdown of the uncertainties in the measured cross-section. Uncertainties in each category are obtained by subtracting in quadrature the error of the POI with other sources floated from the error of the global estimate. Due to the missing correlations among categories, the total systematic uncertainty is not necessarily the quadratic sum of individual errors. The correlation matrices for all parameters are given in Fig. 6.3. Theory uncertainties tend to present some correlations, for instance, PS with ISR/FSR and ME. Rate uncertainty sources such as the luminosity are highly correlated with the $t\bar{t}$ normalisation, which is rate dominant due to the large contribution of $t\bar{t}$ events. JES components have small correlations among themselves, indicating that they are properly decoupled.

6.4.2 Validations and discussions

In order to further verify the fit results and understand the systematic uncertainties, the pull and impact of each NP are inspected. In Fig. 6.4, the ranking plots with the NP floating (in yellow) and fixed (in violet) are presented. Only the top 10 uncertainties, ordered by their impact after fit (hatched boxes) in

Correlation Matrix



Correlation Matrix

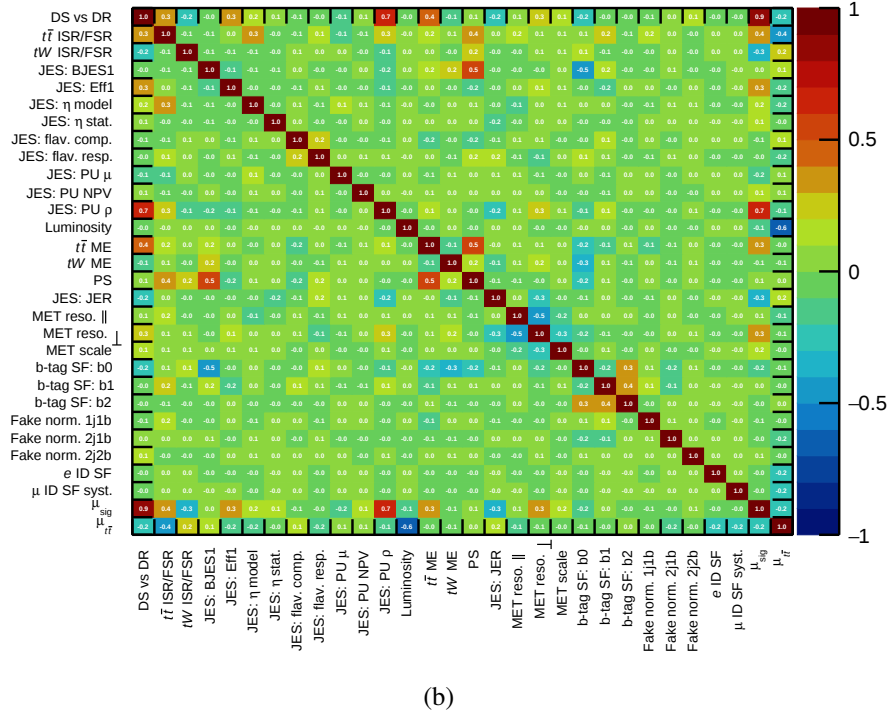


Figure 6.3: Correlation matrix of parameters used in the profile likelihood fit. The results shown represent fits to the Asimov dataset (a) and real data (b). Rows or columns with all correlations lower than 0.2 are pruned to simplify the plots.

Source	$\Delta\mu_{\text{sig}}[\%]$
ME	4.2
PS	3.6
ISR/FSR	11
DR vs DS	19
PDF	$\simeq 0$
Non- $t\bar{t}$ background normalisation	$\simeq 0$
Lepton efficiency, energy scale and resolution	4.1
JES	21
JER	8.9
b -tagging	8.2
$E_{\text{T}}^{\text{miss}}$ calculation	9.3
Luminosity	4.2
Total systematic uncertainty	23
Data statistics	3.4
Total uncertainty	23

Table 6.2: Relative uncertainties in the tW cross-section. They are estimated by fixing systematic sources to their best fit values in each category, re-fitting, and subtracting refitted uncertainty from the total uncertainty in quadrature. Due to correlations between categories, the individual uncertainties are not expected to add up to the total systematic uncertainty. The statistical uncertainty is evaluated by fitting without any NP corresponding to systematic uncertainties, and the total systematic uncertainty is evaluated by subtracting the statistical uncertainty from the total uncertainty in quadrature. For simplicity, the uncertainties in the table are symmetrised.

the floated version, are shown. The full list can be found in Appendix C. Constraints on the NP lead to a shrinkage of impact after fit.

The NPs with a large effect are also expected to have a shift away from zero by $(1 - (\Delta\hat{\theta})^2)^{1/2}$. In practice, the values shift less, primarily because many uncertainties are estimated in a conservative way. “DR vs DS” is the dominant systematic uncertainty due to its large shape uncertainties. It has a big impact in the 2j2b region. The preferred value sits between DR and DS⁶ indicating some mixture of the two schemes, i.e. neither of them is ideal but the reality is enclosed. “JES PU ρ ” is pulled to the boundary of the interval. In fact, a known issue [203] in the estimation procedure is fixed, which is expected to reduce the uncertainty, yet not propagated to this measurement. Both $t\bar{t}$ and tW “ISR/FSR” are pulled towards negative values. Overall, lower level radiation is significantly favoured. “JER”, however, is pulled to the positive direction and is also constrained. In order to investigate this more closely, a fit without $t\bar{t}$ control region was performed and a smaller constraint and smaller shift were seen—the shift and constraint come from the 2j2b region. It also reveals that “ b -tag SF: b0” is constrained when adding the $t\bar{t}$ control region, where two b -jets are required. Modest constraints related with $E_{\text{T}}^{\text{miss}}$ calculation, e.g. “MET reso. $\perp$ ”, is also seen.

ME and PS are significantly constrained in the fit but not much pulled away. This indicates a good modelling of the nominal sample. Alternative generators predict different shapes, which is strongly

⁶ When symmetrising the uncertainty, DS is set to $\theta = +1$.

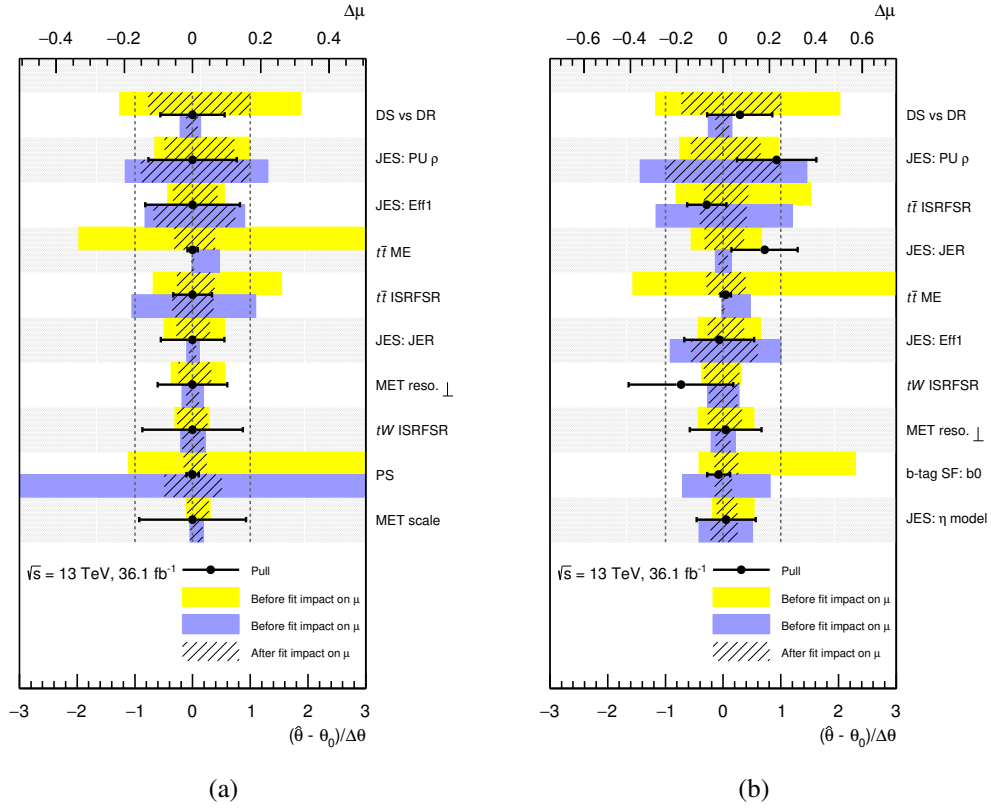


Figure 6.4: Pull (corresponding to the lower x -axis) and impact on μ_{sig} (corresponding to the upper x -axis) of the top 10 NPs ordered by the post-fit impact. The ranking in Asimov (a) and in data (b) are similar, but not identical. Bands in yellow represent the floated version of the impact, i.e. the NP in study is fixed to its up/down variations while others are floated. Bands in violet represent the fixed version of impact, i.e. the NP in study is fixed to its up/down variations while others are fixed to their global fit values. The differences between the yellow and the violet bands comes from the correlations among the NPs. The hatched boxes follow the same definitions of the bands albeit presenting after fit impact.

disfavoured by data. Thus, the likelihood function is very narrow, resulting in a very precise parameter estimation.

It is noticeable that some of the impacts are asymmetric. This is not due to asymmetric up/down variations entering into the impact calculation—the variations are given in a symmetrised manner. In fact, it turns out to be very natural in the maximum likelihood fit, accompanied with large background over signal ratio. As mentioned earlier, signal strength can be solved by setting the first order derivative of log-likelihood to zero, resulting in a function of signal and background yields that are modified respectively by the signal strength and the NP (considering one systematics here for simplicity). Ignoring the Gaussian constraint term, it is easy to work out with only the Poisson part:

$$\begin{aligned}
 & \left. \frac{\partial \sum_{b \in \text{bins}} \ln \mathcal{P}(n_b | \mu \lambda_{\text{sig}}^{\text{pred.}} + \lambda_{\text{bkg}}^{\text{pred.}})}{\partial \mu} \right|_{\mu=\hat{\mu}} = 0, \\
 \Leftrightarrow & \left. \frac{\partial \sum_{b \in \text{bins}} \left(n_b \ln(\mu \lambda_{\text{sig}}^{\text{pred.}} + \lambda_{\text{bkg}}^{\text{pred.}}) - (\mu \lambda_{\text{sig}}^{\text{pred.}} + \lambda_{\text{bkg}}^{\text{pred.}}) + \ln(n_b!) \right)}{\partial \mu} \right|_{\mu=\hat{\mu}} = 0, \\
 \Leftrightarrow & \sum_{b \in \text{bins}} \left(\frac{n_b \lambda_{\text{sig}}^{\text{pred.}}}{\hat{\mu} \lambda_{\text{sig}}^{\text{pred.}} + \lambda_{\text{bkg}}^{\text{pred.}}} - \lambda_{\text{sig}}^{\text{pred.}} \right) = 0, \\
 \Leftrightarrow & \sum_{b \in \text{bins}} \left(\frac{n_b}{\hat{\mu} \lambda_{\text{sig}}(\theta) + \lambda_{\text{bkg}}(\theta)} - 1 \right) = 0,
 \end{aligned}$$

The last step shows an explicit dependence on θ . Even though the predicted yield linearly depends on θ , $\hat{\mu}$ is not linear as a function of θ .

Another source of the asymmetry comes from the large background-over-signal ratio (in this analysis $N_{\bar{t}\bar{t}} : N_{tW} \simeq 8 : 1$). The fit will constrain the total number of events to be the data yields N_{data} (with some uncertainties). Let N_{others} being the yields for other backgrounds, one could write out the following two equations:

$$(1 + \Delta\mu_{tW}^+) \times N_{tW} + (1 + \Delta\mu_{\bar{t}\bar{t}}^+) \times N_{\bar{t}\bar{t}}^+ = N_{\text{data}} - N_{\text{others}},$$

$$(1 + \Delta\mu_{tW}^-) \times N_{tW} + (1 + \Delta\mu_{\bar{t}\bar{t}}^-) \times N_{\bar{t}\bar{t}}^- = N_{\text{data}} - N_{\text{others}}.$$

$\Delta\mu_{tW}^\pm$ are the impact of $\theta = \pm 1$ while $\Delta\mu_{\bar{t}\bar{t}}^\pm$ are variations of fitted $\mu_{\bar{t}\bar{t}}$ from one in up/down cases. Summing up the two equations and using $N_{\bar{t}\bar{t}}^{(+)} + N_{\bar{t}\bar{t}}^{(-)} \simeq 2N_{\bar{t}\bar{t}}$ implied by a symmetric NP, one gets

$$(\Delta\mu_{tW}^+ + \Delta\mu_{tW}^-) \times N_{tW} = -(\Delta\mu_{\bar{t}\bar{t}}^+ \times N_{\bar{t}\bar{t}}^+ + \Delta\mu_{\bar{t}\bar{t}}^- \times N_{\bar{t}\bar{t}}^-) + 2 \times (N_{\text{data}} - N_{\text{others}} - N_{tW} - N_{\bar{t}\bar{t}}).$$

If the fit converges, $N_{\text{data}} - N_{\text{others}} - N_{tW} - N_{\bar{t}\bar{t}} \simeq 0$,

$$\Delta\mu_{tW}^+ + \Delta\mu_{tW}^- = - \left(\Delta\mu_{\bar{t}\bar{t}}^+ \times \frac{N_{\bar{t}\bar{t}}^+}{N_{tW}} + \Delta\mu_{\bar{t}\bar{t}}^- \times \frac{N_{\bar{t}\bar{t}}^-}{N_{tW}} \right) \neq 0.$$

Even if the fitted $t\bar{t}$ scale factor is symmetric, $\Delta\mu_{t\bar{t}}^+ \simeq \Delta\mu_{t\bar{t}}^-$, a small deviation from one will be amplified by the large value of $N_{t\bar{t}}^\pm/N_{tW}$, resulting in $\Delta\mu_{tW}^+$ and $\Delta\mu_{tW}^-$ asymmetric around zero.

Many other NPs, however, are not significantly constrained by the fit. This is the expected behaviour for the parameters with a small effect on the discriminant. The estimate of such parameters is expected to be close to zero because almost no information can be extracted from data that would make the value shift away from zero.

6.4.3 Results

A measured cross-section for $\sigma_{tW} = 82 \pm 2 \text{ (stat.)}_{-15}^{+18} \text{ (syst.)} \pm 3 \text{ (lumi.) pb}$ is obtained by simultaneously fitting the BDT response in the three regions, corresponding to an observed (expected) significance of 10σ (11σ).

Differential cross-section measurements

This chapter presents the tW differential cross-section measurements. The measurements are motivated in Section 7.1. Methodologies and technique details are introduced in Section 7.2. An application of the technique is illustrated in Section 7.3, and results are shown in Section 7.4.

7.1 Motivation

Differential cross-sections convey more information than the inclusive cross-section, since they allow the detailed behaviour in a certain phase space to be studied. New physics models may have different impacts on the kinematic distributions. Therefore, differential cross-section measurements test models in a more precise way. In addition, the results help to tune the parameters of MC generators to achieve a more accurate simulation. Experimentally, differential cross-section measurements are possible once the amount of dataset is sufficiently large.

7.2 Unfolding

7.2.1 Deconvolution problem

In the inclusive cross-section measurement, the physical objects used are at the *reconstruction level*. It measures the kinematic quantities, such as p_T and η , that are after the detector detection and object reconstruction. These reconstruction-level kinematic distributions are distorted by the limited acceptance and the finite resolution of the detectors. It is acceptable for the inclusive cross-section measurement since a migration effect between neighbouring bins would not give any difference. However, it becomes a crucial problem of differential cross-section measurements where the kinematic quantities before being “smeared” by the detector, the so-called *particle-level* quantities, are of more interest. Therefore, a deconvolution from the measured distributions to the corresponding particle-level distributions is required. In particle physics, this process is usually called *unfolding*. The unfolded distributions can be directly compared with the results of other experiments and with theoretical predictions.

Mathematically the probability density functions for a measured value x and its corresponding true

value y are related by a convolution function [204],

$$f_{\text{meas}}(x) = \int \mathcal{R}(x|y)f_{\text{true}}(y)dy.$$

For a discrete case where the probability densities are represented by histograms with a finite number of bins, it becomes

$$x_i = \sum_{j=1}^{\text{bins}} \mathcal{R}_{ij}y_j.$$

The response matrix $\mathcal{R}_{ij}$ can be interpreted as a conditional probability:

$$\mathcal{R}_{ij} = \mathcal{P}(\text{observed in bin } i \mid \text{true value in bin } j).$$

It includes the inefficiency, the bias and the smearing effects from the detector. Note that $\mathcal{R}_{ij}$ can be non-square matrix when different bin numbers are used in true and measured histograms.

$\mathcal{R}_{ij}$ is usually computed from the MC simulation, and x_i is represented by the observed data yields, n_i . In addition, background events, b_i , should be subtracted. Therefore,

$$\vec{x} = E[\vec{n}] = \mathcal{R}\vec{y} + \vec{b}.$$

The key point is how to solve the true value $\vec{y}$ from $\mathcal{R}$, $\vec{x}$ and $\vec{b}$.

7.2.2 Bin-by-bin corrections

When true and measured histograms have the same binning and the bin-to-bin migrations are negligible, $\mathcal{R}$ is approximately a diagonal square matrix and the correction factors can be evaluated for each bin using MC and applied to the measurements. However, this method is subject to additional random fluctuations from data and MC. It is not used in this thesis, but rather serves as a cross check.

7.2.3 Matrix inversion

An alternative idea to solve $\vec{y}$ is by inverting the matrix $\mathcal{R}$:

$$\begin{aligned} \hat{\vec{y}} &= \mathcal{R}^{-1}(\vec{x} - \vec{b}) \\ &= \mathcal{R}^{-1}(\vec{n} - \vec{b}), \end{aligned}$$

if one realises that $\vec{n}$ is the maximum likelihood estimates for $\vec{x}$.

The estimates $\hat{\vec{y}}$ is unbiased [204]. Although it has the smallest possible variance among all unbiased estimates, the variance is still large. Besides, there are other disadvantages. Firstly, the method would not work for a singular matrix whose inversion does not exist. Secondly, it cannot handle large negative correlations between neighbouring bins [205].

7.2.4 Regularised unfolding

The large variances in the matrix inversion method can be regularised by including an additional term (a bias) on the maximum log-likelihood $\mathcal{L}(\vec{y})$:

$$\Phi(\vec{y}) = \mathcal{L}(\vec{y}) + \alpha S(\vec{y}),$$

It is known as bias–variance tradeoff. $S(\vec{y})$ is the regularisation function and α is the regularisation parameter. One can choose $S(\vec{y})$ as a measure of the smoothness of the distribution, for instance the square sum of the differences in adjacent bins, such that the fluctuations are disfavoured in the estimation by a factor of α . If α is zero, it degenerates to the matrix inversion unfolding; if α is positive infinity, any steps of the bin heights are disfavoured—the unfolded distribution will be flat.

By choosing the regularisation function and parameter carefully, this method can produce satisfactory results. However, it has not been widely used, probably due to technical complexity. Also notice that it works only with one dimensional distributions.

7.2.5 Iterative unfolding

A different approach is an iterative method proposed by D’Agostini [206], formulated using the Bayesian language. The particle-level value is a “cause” (C) while the observed/reconstructed value is an “effect” (E). The response matrix element can be interpreted as the conditional probability of the i -th cause to produce the j -th effect:

$$\mathcal{R}_{ij} = \mathcal{P}(\text{reconstruction-level value in bin } j \mid \text{particle-level value in bin } i) = \mathcal{P}(E_j|C_i).$$

According to Bayes’ theorem, the posterior, $\mathcal{P}(C_i|E_j)$, is proportional to the prior, $\mathcal{P}(C_i)$:

$$\mathcal{P}(C_i|E_j) \propto \mathcal{P}(E_j|C_i)\mathcal{P}(C_i).$$

On the other hand, once the posterior is known, the expected number of events in the bin i of the cause-distribution can be expressed as

$$\hat{n}(C_i) = \frac{1}{\hat{\epsilon}_i} \sum_j n(E_j)\mathcal{P}(C_i|E_j).$$

Dividing by the total number of events $\hat{N}(C) = \sum_i \hat{n}(C_i)$ on both sides, and taking the estimate of the efficiency $\hat{\epsilon}_i = \hat{n}(C_i)/\hat{n}(E_i)$, one retrieves another relation between the prior, $\mathcal{P}(C_i)$, and the posterior, $\mathcal{P}(C_i|E_j)$:

$$\hat{\mathcal{P}}(C_i) = \frac{1}{\hat{\epsilon}_i} \sum_j \mathcal{P}(E_j)\mathcal{P}(C_i|E_j),$$

which is different from the one given by Bayes’ theorem. With the two relations between $\mathcal{P}(C_i)$ and $\mathcal{P}(C_i|E_j)$, an iterative procedure can be applied to solve $\mathcal{P}(C_i|E_j)$.

The method does not require a matrix inversion, and is theoretically well grounded. Moreover, it can cope with multidimensional problems and different binning in true and observed distributions. Due to these advantages, the iterative method is used in this thesis.

7.2.6 RooUnfold package

The iterative unfolding technique is implemented in the RooUnfold software package [207]. To make the calculation more reliable, the unfolding procedure is factorised into a migration part and two additional corrections. The reconstructed distribution $\vec{x}$ of the events that pass both the fiducial selection (will be discussed in Section 7.3.1) and reconstruction-level requirements is related to the true distribution $\vec{y}$ by the response matrix $\mathcal{R}$:

$$\vec{x} = \mathcal{R} \cdot \vec{y}.$$

In addition, $\vec{y}$ is corrected to the number of events in the full fiducial phase space (N^{fid}) by a bin-by-bin correction to the efficiency, ϵ_i^{eff} :

$$\epsilon_i^{\text{eff}} = \frac{y_i}{N_i^{\text{fid}}},$$

while $\vec{x}$ is corrected to the number of events in the reconstruction phase space (N^{reco}) by a bin-by-bin acceptance correction, ϵ_j^{oof} , to account for out-of-fiducial events that are reconstructed but fall outside the fiducial acceptance at the particle level:

$$\frac{1}{\epsilon_j^{\text{oof}}} = \frac{N_j^{\text{reco}}}{x_j}.$$

Taking the background contributions into account, the full unfolding procedure is then described by

$$N_j^{\text{reco}} = N_j^{\text{obs}} - B_j = \frac{1}{\epsilon_j^{\text{oof}}} \sum_i \mathcal{R}_{ji} \cdot (\epsilon_i^{\text{eff}} N_i^{\text{fid}}),$$

where i (j) denotes the bin at the particle (reconstruction) level, N_j^{obs} the observed number of events and B_j the background contributions. A schematic picture is shown in 7.1.

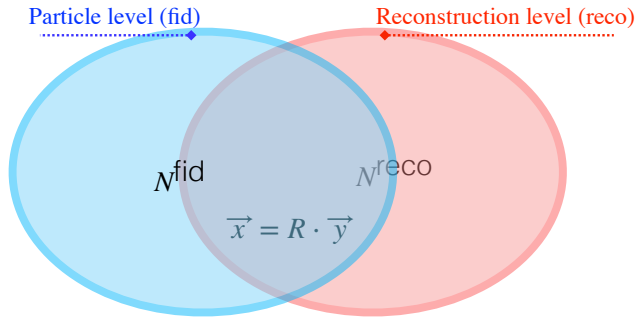


Figure 7.1: A schematic picture showing the relations of fiducial phase space in blue, and reconstruction phase space in red.

In order to validate the unfolding procedure, MC samples are split into a training set, which is used to determine the response matrices and correction factors, and a test set, which is used to validate the results. Instead of using a flat distribution, the training truth is taken as a prior to accelerate the convergence. Unfolding errors are calculated by error propagation. As pointed out in [207], the errors stem not only from the observed events, but also from the correlation between the response matrix and the observed events. That is because with the iterative updates of the response matrix, the uncertainty in the observed events propagates to the matrix repeatedly. Therefore, the unfolding error increases

with the number of iterations.

7.3 Cross-section determination

7.3.1 Particle-level objects and fiducial region

Particle-level objects in the MC simulation consider all stable ($c\tau > 10$ mm) outgoing particles [208]. Particle-level prompt charged leptons and neutrinos that arise from decays of W bosons or Z bosons are accepted. The charged leptons are then dressed with nearby photons, considering all photons that satisfy $\Delta R(\ell, \gamma) < 0.1$ and do not originate from hadrons, adding the four-momenta of all selected photons to the bare lepton to obtain the dressed lepton four-momentum. Particle-level jets are built from all remaining stable particles in the event after excluding leptons, and the photons used to dress the leptons, by clustering them using the anti- k_t algorithm with $R = 0.4$. Particle-level b -jets are built by associating jets to any b hadron with $p_T > 5$ GeV. This is done by reclustering jets with b hadrons included in the input list of particles, but with their p_T scaled down to negligibly small values. Then the jets containing b hadrons after this reclustering are considered to be b -jets.

The definition of the fiducial region is chosen to match the lepton and jet requirements at the reconstruction level. Two leptons with $p_T > 27$ GeV and $|\eta| < 2.5$ are required, rejecting events with a third lepton with $p_T > 20$ GeV and $|\eta| < 2.5$. Furthermore, exactly one jet must be found which is b -tagged and satisfies $p_T > 25$ GeV and $|\eta| < 2.5$. No cuts are placed on event-level particle quantity E_T^{miss} .

7.3.2 Choice of variables and binning

Although the top quark and the two W bosons cannot be directly reconstructed due to insufficient kinematic constraints, one can select a list of observables that are correlated with kinematic properties of tW production and sensitive to the differences in theoretical modelling. While keeping the list of observables as short as possible, unfolded distributions are measured for:

- the energy of the b -jet, $E(b)$ ¹;
- the mass of the leading lepton and b -jet, $m(\ell_1 b)$;
- the mass of the sub-leading lepton and the b -jet, $m(\ell_2 b)$;
- the energy of the system of the two leptons and b -jet, $E(\ell\ell b)$;
- the transverse mass of the leptons, b -jet and neutrinos, $m_T(\ell\ell\nu\nu b)$; and
- the mass of the two leptons and the b -jet, $m(\ell\ell b)$.

The top-quark production is probed most directly by $E(b)$, the only final-state object that can unambiguously be matched to the decay products of the top quark. The top-quark decay is probed by $m(\ell_1 b)$ and $m(\ell_2 b)$, which are sensitive to angular correlations of decay products due to production spin correlations. The combined tW -system is probed by $E(\ell\ell b)$, $m_T(\ell\ell\nu\nu b)$, and $m(\ell\ell b)$. Here the transverse mass is preferred if a neutrino is involved. At the particle level the vector summed transverse momenta of simulated neutrinos are used in $m_T(\ell\ell\nu\nu b)$; while at the reconstruction level, the transverse momenta of the neutrinos in $m_T(\ell\ell\nu\nu b)$ are represented by the measured E_T^{miss} . All quantities for leptons and jets are taken simply from the relevant reconstructed or particle-level objects.

¹ Hereinafter, the jet is denoted as b , which, however, is identical to j_1 in the previous chapters.

These observables are selected to minimise the bias introduced by the BDT requirement, since it makes unfolding more difficult to unfold observables with high correlations with the BDT discriminant.

The same binning of the variables are chosen at the reconstruction- and particle-level to not introduce unnecessary confusion to the unfolding procedure. Although any binning is conceptually possible to use, the decision is made to balance statistical and systematic uncertainties in each bin, and to create an unfolding procedure stable against statistical fluctuations. In practice, at most approximately 20 % statistical uncertainty in each bin and at least approximately 60 % of events per column of the response matrix lie on the diagonal entry. The range is trying to cover the majority of simulated and data events such that the variables measured tend to peak somewhere near the centre of the range, with tails near the minimum and maximum. Underflow and overflow events, if any, are added into the first and last bin, respectively. The binning selected for each variable is given in Table 7.1.

Variable	Bins	Bin boundaries [GeV]	Number of iterations	Figure reference
$E(b)$	5	25, 60, 100, 135, 175, 500	15	Fig. 7.2(a)
$m(\ell_1 b)$	6	0, 60, 100, 150, 200, 250, 400	7	Fig. 7.2(b)
$m(\ell_2 b)$	4	0, 50, 100, 150, 400	5	Fig. 7.2(c)
$E(\ell\ell b)$	6	50, 175, 275, 375, 500, 700, 1200	5	Fig. 7.2(d)
$m_T(\ell\ell\nu\nu b)$	4	50, 275, 375, 500, 1000	7	Fig. 7.2(e)
$m(\ell\ell b)$	6	0, 125, 175, 225, 300, 400, 1000	5	Fig. 7.2(f)

Table 7.1: Summary of the number of bins, bin boundaries, and the number of iterations chosen in the unfolding procedure.

7.3.3 Signal-enriched region

The 1j1b region contains the largest fraction of the signal events. To further enrich the signal events, a requirement can be placed on the BDT response (see Fig. 5.6(a)). It is optimised to reduce the total uncertainty of the measurement, considering both statistical and systematic uncertainties, and the decision is made to be larger than 0.1 to give on average the smallest uncertainty over all bins.

The effect of this requirement on event yields is shown in Table 7.2 and the distributions are shown in Fig. 7.2. The BDT requirement lowers systematic uncertainties by reducing contributions from the $t\bar{t}$ background, which is subject to large modelling uncertainties. For example, the total systematic uncertainty in the fiducial cross-section is reduced by 18% of the total when applying the BDT response requirement, compared to having no requirement.

7.3.4 Number of iterations

As the number of iteration grows, the statistical uncertainty inflates, while the bias of the results away from the true values, which indicates the accuracy of the unfolding, generally tends to reach a plateau. Thus the number of iterations is chosen to minimise the growth of the statistical uncertainty propagated through the unfolding procedure while operating in a regime where the bias is sufficiently independent of the number of iterations. The numbers used in the measurements are summarised in Table 7.1.

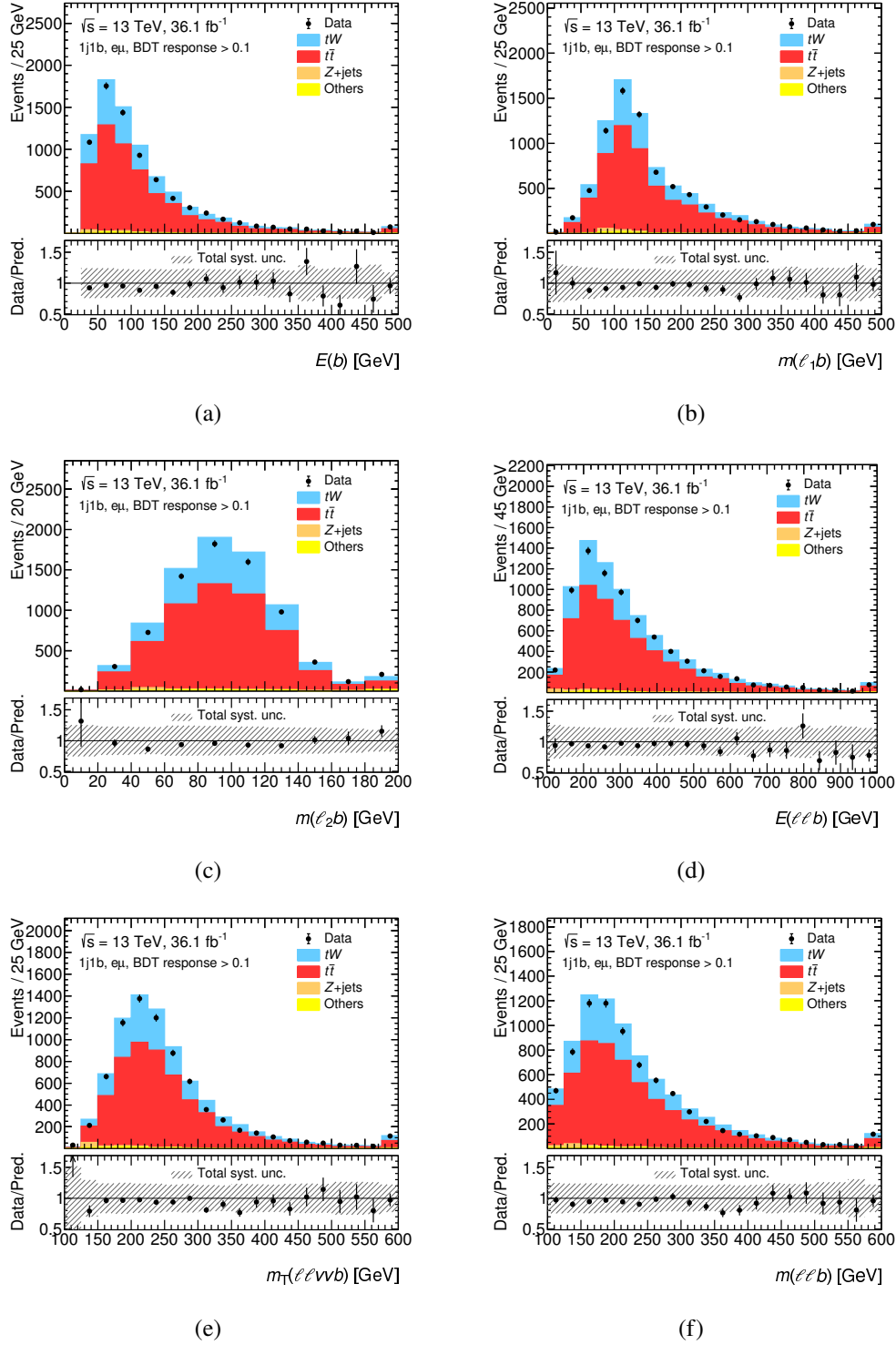


Figure 7.2: Distributions of the variables chosen for the differential cross-section measurements. The BDT response is required to be larger than 0.1. The signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties in the MC predictions. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panel gives the yields in the number of events per bin, while the lower panel gives the ratios of the numbers of observed events to the total prediction in each bin.

Process	Events	Events BDT response > 0.1
tW	4910 ± 800	2350 ± 510
$t\bar{t}$	21700 ± 3600	5500 ± 1400
$Z + \text{jets}$	152 ± 76	88 ± 44
Diboson	169 ± 42	75 ± 18
Fakes	120 ± 120	28 ± 28
Predicted	27100 ± 3700	8000 ± 1500
tW fraction [%]	18 ± 4	29 ± 8
Observed	26171	7545

Table 7.2: Predicted and observed yields in the 1j1b region before and after applying the BDT requirement. After the BDT requirement, the tW fraction increases.

7.3.5 Unfolding and cross-section calculation

The unfolding procedure is performed for each of the observables. The acceptances (ϵ^{oof}), efficiencies (ϵ^{eff}), and response matrices ($\mathcal{R}$) are calculated and summarised in Fig. 7.3.

Unfolded event yields in the i -th bin (N_i^{fid}) are converted to cross-section values as a function of an observable X using the expression:

$$\frac{d\sigma_i}{dX} = \frac{N_i^{\text{fid}}}{L\Delta_i},$$

where L is the integrated luminosity of the data sample and Δ_i is the width of bin i of the particle-level distribution. The fiducial cross-section is the sum of the differential cross-sections in each bin multiplied by the corresponding bin widths:

$$\sigma^{\text{fid}} = \sum_i \left(\frac{d\sigma_i}{dX} \cdot \Delta_i \right) = \sum_i \frac{N_i^{\text{fid}}}{L}.$$

Differential cross-sections are divided by the fiducial cross-section to get *normalised* differential cross-sections:

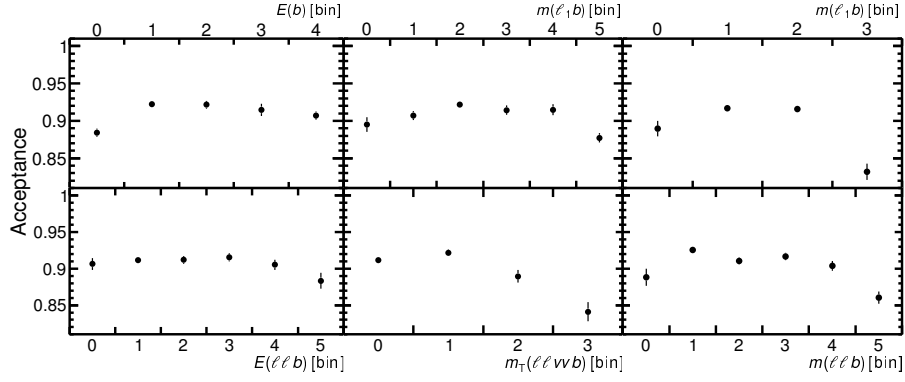
$$\frac{1}{\sigma^{\text{fid}}} \cdot \frac{d\sigma_i}{dX} = \frac{N_i^{\text{fid}}}{\Delta_i \cdot \sum_i N_i^{\text{fid}}},$$

7.3.6 Validation of unfolding methods

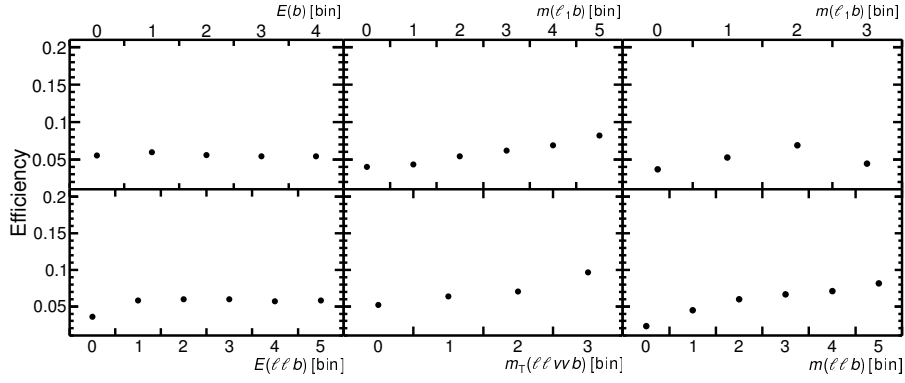
Before unfolding real data, a number of statistical tests are performed to validate the implementation of the unfolding technique.

First, the test sample is picked to be identical to the training sample. In these trivial tests, the unfolded data matches exactly the truth values of the test data, within limits of numerical machine precision.

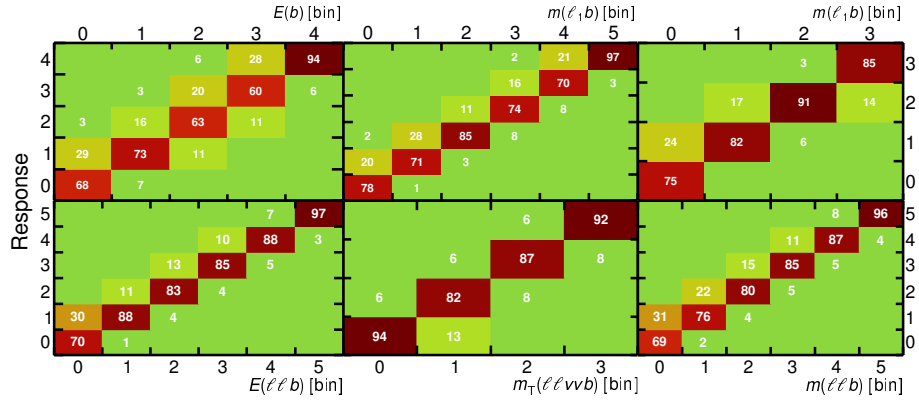
Next, a series of tests is performed to check that the statistical uncertainties assigned to the unfolded



(a)



(b)



(c)

Figure 7.3: The acceptance corrections (a), ϵ^{oof} , the efficiency factors (b), ϵ^{eff} , and the response matrices (c), $\mathcal{R}$, in bins of the variables. Error bars subjected to ϵ^{oof} and ϵ^{eff} correspond to statistical uncertainties from the MC sample.

data represent the true uncertainty in the test sample. The MC reconstruction-level distribution is interpreted as a probability density function, which is then used to generate 2000 pseudo-data distributions. The training sample is also taken from the same underlying nominal MC sample. Each pseudo-data distribution is then used as a test sample, unfolded and compared to the corresponding MC truth. For each bin of the observables, the residual is divided by the assigned statistical uncertainty (see Section 7.3.7) to obtain a pull value, then a Gaussian is fitted to this pull distribution to extract the mean and width. The means and widths of each pull distribution are summarised in Fig. 7.4. The pulls are centred around zero with widths close to one, indicating that the statistical uncertainties are correctly estimated.

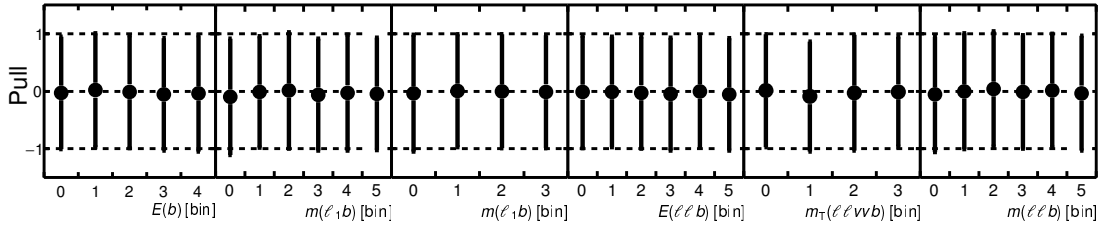


Figure 7.4: The means and widths calculated for each bin of each observable. The means are centred around zero and the widths are close to one.

7.3.7 Uncertainty estimation

Sources

In addition to the sources of systematic uncertainties for the inclusive cross-section measurement listed in Section 6.3, three additional sources are considered related with the unfolding procedure. Uncertainties due to the size of the MC samples are estimated using pseudo-experiments. An ensemble of pseudo-data is created by fluctuating the MC samples within the statistical uncertainties. Each set of pseudo-data is used to construct ϵ^{oof} , ϵ^{eff} , and $\mathcal{R}$, and the nominal MC sample is unfolded. The width of the distribution of unfolded values from this ensemble is taken as the statistical uncertainty. Additional non-closure uncertainties are added after the *stress* test of the unfolding procedure with injected Gaussian or linear functions. Each distribution is tested by reweighting the input MC sample according to the injected function, unfolding, and checking that the weights are recovered in the unfolded distribution. The extent to which the unfolded weighted data are biased with respect to the underlying weighted generator-level distribution is taken as the unfolding non-closure uncertainty. Lastly, an uncertainty of 5.5 % is applied to the $t\bar{t}$ normalisation to account for the renormalisation and factorisation scale, and PDF uncertainties in the NNLO cross-section calculation.

Procedure for the uncertainty estimation

The propagation of uncertainties through the unfolding process is computed by constructing the migration matrix and correction factors with the baseline sample, and performing the unfolding with the varied sample as an input. In most cases, the baseline sample is POWHEG-Box+PYTHIA 6 FULLSIM, but in those cases where the varied sample uses the Af2, the baseline sample is also changed to Af2.

For uncertainties modifying background processes, varied samples take into account the changes in the background. Experimental uncertainties are treated as correlated between signal and background in this procedure. The varied samples are unfolded using the response matrices and correction factors from the baseline sample, and compared to the corresponding particle-level distributions from the MC event generator; the relative difference in each bin is the estimated systematic uncertainty.

7.4 Results

Normalised differential cross-sections as a function of the particle-level observables are given in Table 7.3. In Figs. 7.5 to 7.7, the cross-sections are compared to the predictions of various MC event generators in the same fiducial region. The results show that the largest uncertainties come from the size of the data sample and $t\bar{t}$ and tW modelling. The detailed uncertainty tables for each observable can be found in Appendix D. Both the statistical and systematic uncertainties have a significant impact on the result. The exact composition varies bin-to-bin but there is no single source of uncertainty that dominates each normalised measurement. Some of the largest systematic uncertainties are those related to $t\bar{t}$ modelling. The cancellation in the normalised differential cross-sections is very effective at reducing a number of systematic uncertainties. The most notable cancellation is related to the $t\bar{t}$ parton shower model uncertainty, which is quite dominant prior to dividing by the fiducial cross-section.

Differences between the MC predictions are smaller than the uncertainty on the data. The predictions of DS and DR samples give very similar results for all observables as expected from the fiducial selection. The predictions of POWHEG-Box+PYTHIA 6 with varied ISR/FSR tuning were also examined but not found to give significantly different distributions in the fiducial phase space of this analysis.

Furthermore, the covariance matrix, $\mathbf{C}$, for each differential cross-section measurement is computed following a procedure similar to the one used in [209]. Two covariance matrices are constructed in different ways and summed to form the final covariance. The first one is computed using 10 000 pseudo-experiments and includes statistical uncertainties on data as well as systematic uncertainties from experimental sources. The statistical uncertainties are included by independently fluctuating each bin of the data distribution according to Poisson distributions for each pseudo-experiment. Each bin of the resulting pseudo-data distribution is then fluctuated according to a Gaussian distribution for each experimental uncertainty, preserving bin-to-bin correlation information for each uncertainty. The other matrix includes the systematic uncertainties from event generator model uncertainties, unfolding non-closure uncertainties, and MC statistical uncertainties. In this second matrix, the bin-to-bin correlation value is set to zero for the non-closure and MC statistical uncertainties, and set to unity for the other uncertainties. The impact of setting the bin-to-bin correlation value to unity was compared for the non-closure uncertainty, and this choice was found to have negligible impact on the results. This covariance matrix is used to compute a χ^2 and corresponding p -value to assess how well the measurements agree with the predictions. The χ^2 values are computed using the expression:

$$\chi^2 = \mathbf{v}^T \mathbf{C}^{-1} \mathbf{v},$$

where $\mathbf{v}$ is the vector of differences between the measured cross-sections and predictions. The comparison between the data and MC predictions is summarised in Fig. 7.8, where χ^2 values and corresponding p -values are presented. Most of the MC models show fair agreement with the measured cross-sections, with no particularly low p -values observed.

$E(b)$ bin [GeV]	[25, 60]	[60, 100]	[100, 135]	[135, 175]	[175, 500]	
$(1/\sigma) d\sigma/dx$ [GeV ⁻¹]	0.00422	0.00937	0.00428	0.00169	0.000801	
Stat. unc. [%]	22	12	27	48	11	
Total syst. unc. [%]	27	28	30	59	28	
Total unc. [%]	35	31	40	76	30	
$m(\ell_1 b)$ bin [GeV]	[0, 60]	[60, 100]	[100, 150]	[150, 200]	[200, 250]	[250, 400]
$(1/\sigma) d\sigma/dx$ [GeV ⁻¹]	0.00105	0.0036	0.00887	0.00328	0.00163	0.000692
Stat. unc. [%]	29	25	9.6	18	25	16
Total syst. unc. [%]	54	48	21	29	26	47
Total unc. [%]	61	54	23	34	36	50
$m(\ell_2 b)$ bin [GeV]	[0, 50]	[50, 100]	[100, 150]	[150, 400]		
$(1/\sigma) d\sigma/dx$ [GeV ⁻¹]	0.00205	0.00885	0.00654	0.000512		
Stat. unc. [%]	26	9.9	9.9	17		
Total syst. unc. [%]	83	8.4	9.9	65		
Total unc. [%]	87	13	14	67		
$E(\ell\ell b)$ bin [GeV]	[50, 175]	[175, 275]	[275, 375]	[375, 500]	[500, 700]	[700, 1200]
$(1/\sigma) d\sigma/dx$ [GeV ⁻¹]	0.00069	0.00364	0.0025	0.0014	0.000435	7.39×10^{-5}
Stat. unc. [%]	30	9.7	12	15	22	34
Total syst. unc. [%]	95	12	22	21	66	74
Total unc. [%]	99	16	25	26	70	81
$m_T(\ell\ell\nu\nu b)$ bin [GeV]	[50, 275]	[275, 375]	[375, 500]	[500, 1000]		
$(1/\sigma) d\sigma/dx$ [GeV ⁻¹]	0.00353	0.00121	0.000498	4.25×10^{-5}		
Stat. unc. [%]	6.1	22	22	28		
Total syst. unc. [%]	9.6	31	43	78		
Total unc. [%]	11	38	48	83		
$m(\ell\ell b)$ bin [GeV]	[0, 125]	[125, 175]	[175, 225]	[225, 300]	[300, 400]	[400, 1000]
$(1/\sigma) d\sigma/dx$ [GeV ⁻¹]	0.00112	0.0043	0.00586	0.00315	0.000486	0.000112
Stat. unc. [%]	24	17	11	12	35	17
Total syst. unc. [%]	130	21	23	14	54	84
Total unc. [%]	140	27	25	18	64	86

Table 7.3: Summary of the measured normalised differential cross-sections, with uncertainties shown as percents. The uncertainties are divided into statistical and systematic contributions.

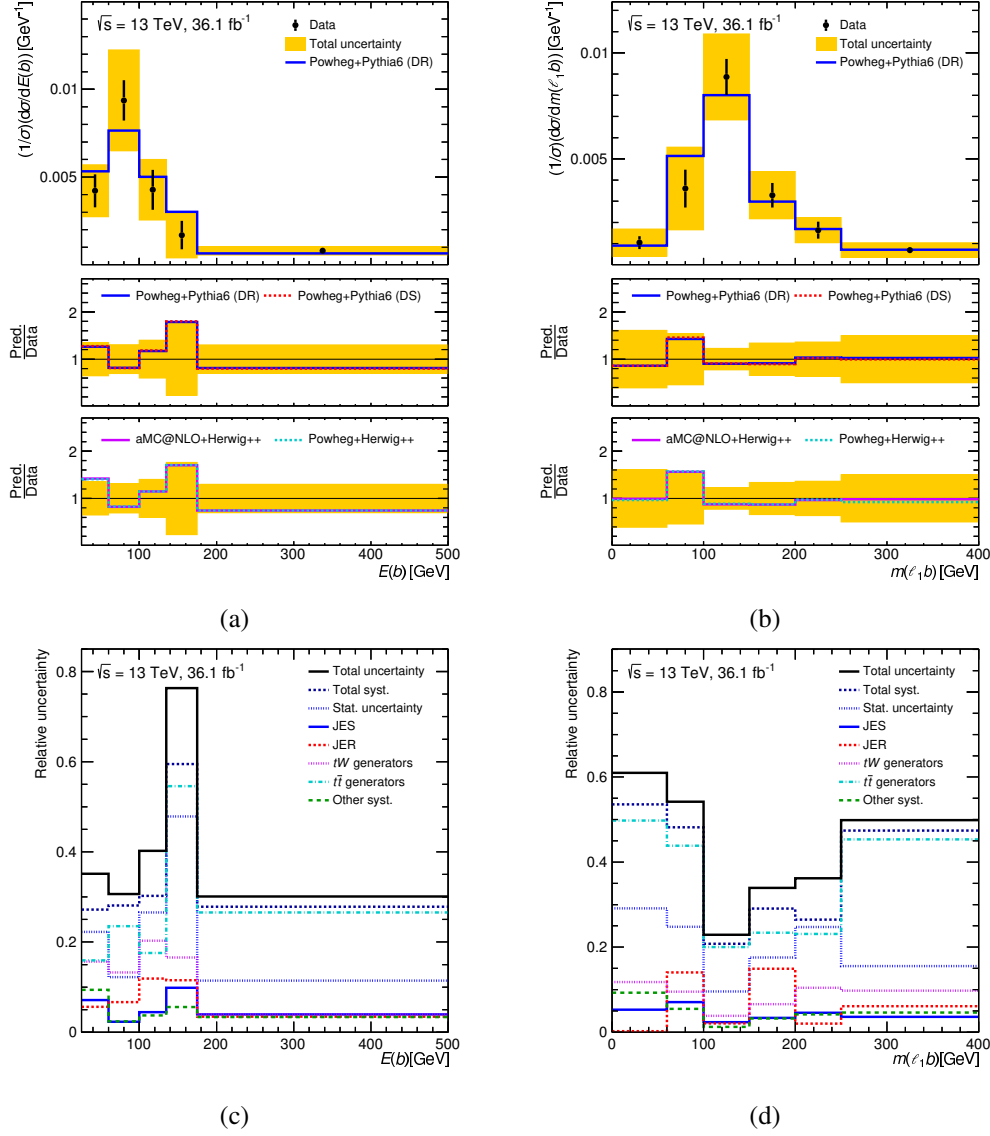


Figure 7.5: Normalised differential cross-sections unfolded from data, compared with selected MC models (a) (b), with respect to $E(b)$ and $m(\ell_1 b)$, and the main uncertainties (c) (d). The error bars on the data points show the statistical uncertainties. Values of differential cross-sections and uncertainties can be found in Table 7.3.

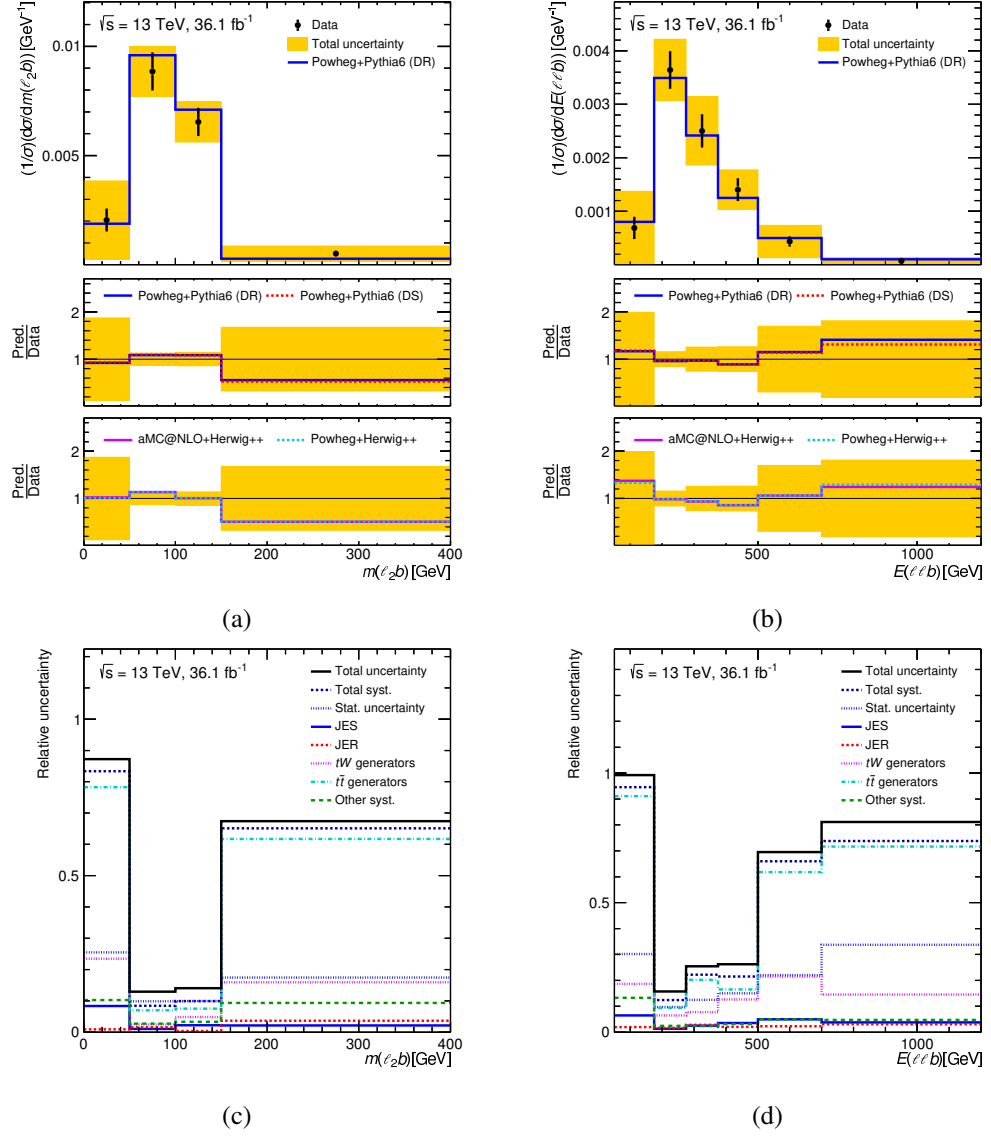


Figure 7.6: Normalised differential cross-sections unfolded from data, compared with selected MC models (a) (b), with respect to $m(\ell_2 b)$ and $E(\ell\ell b)$, and the main uncertainties (c) (d). The error bars on the data points show the statistical uncertainties. Values of differential cross-sections and uncertainties can be found in Table 7.3.

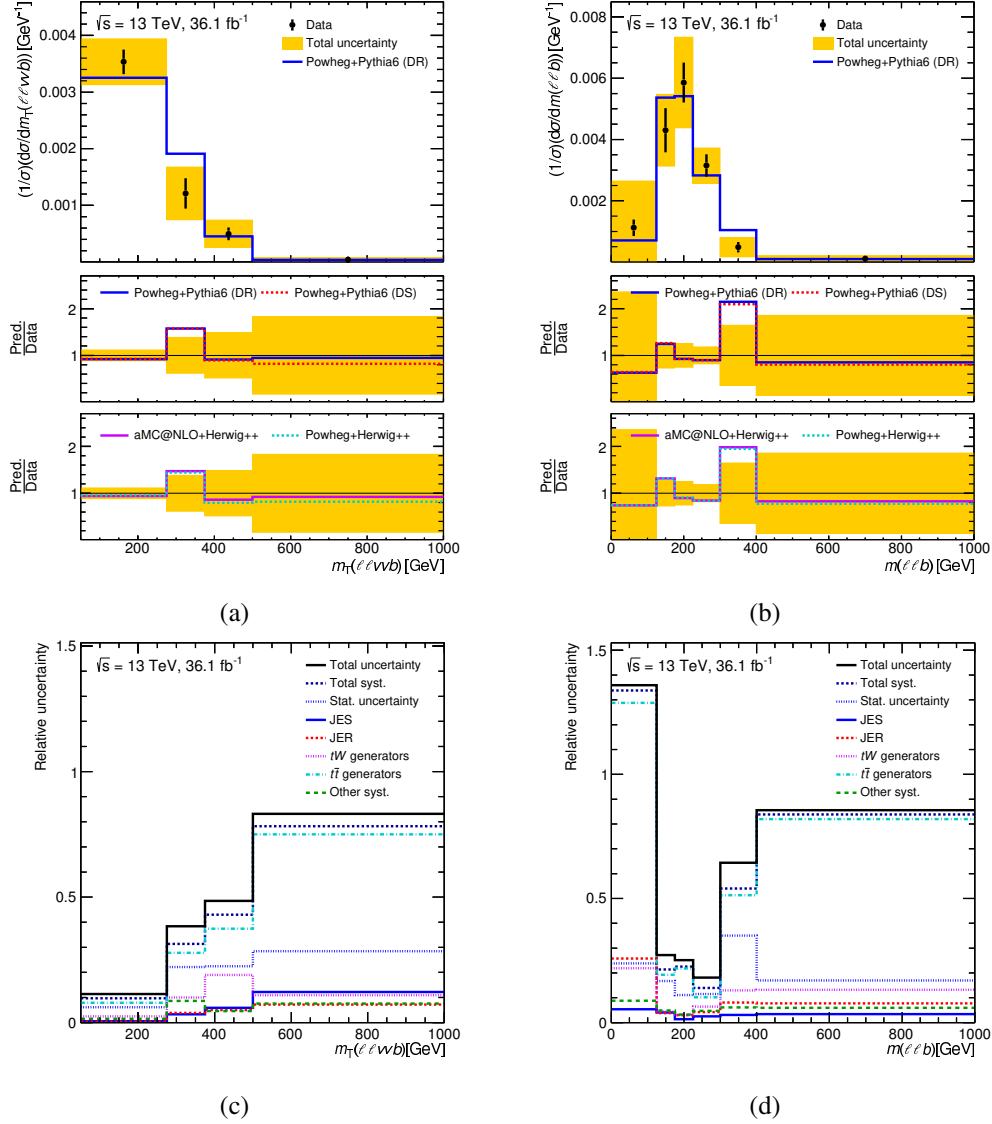


Figure 7.7: Normalised differential cross-sections unfolded from data, compared with selected MC models (a) (b), with respect to $m_T(\ell\ell\nu\nu b)$ and $m(\ell\ell b)$, and the main uncertainties (c) (d). The error bars on the data points show the statistical uncertainties. Values of differential cross-sections and uncertainties can be found in Table 7.3.

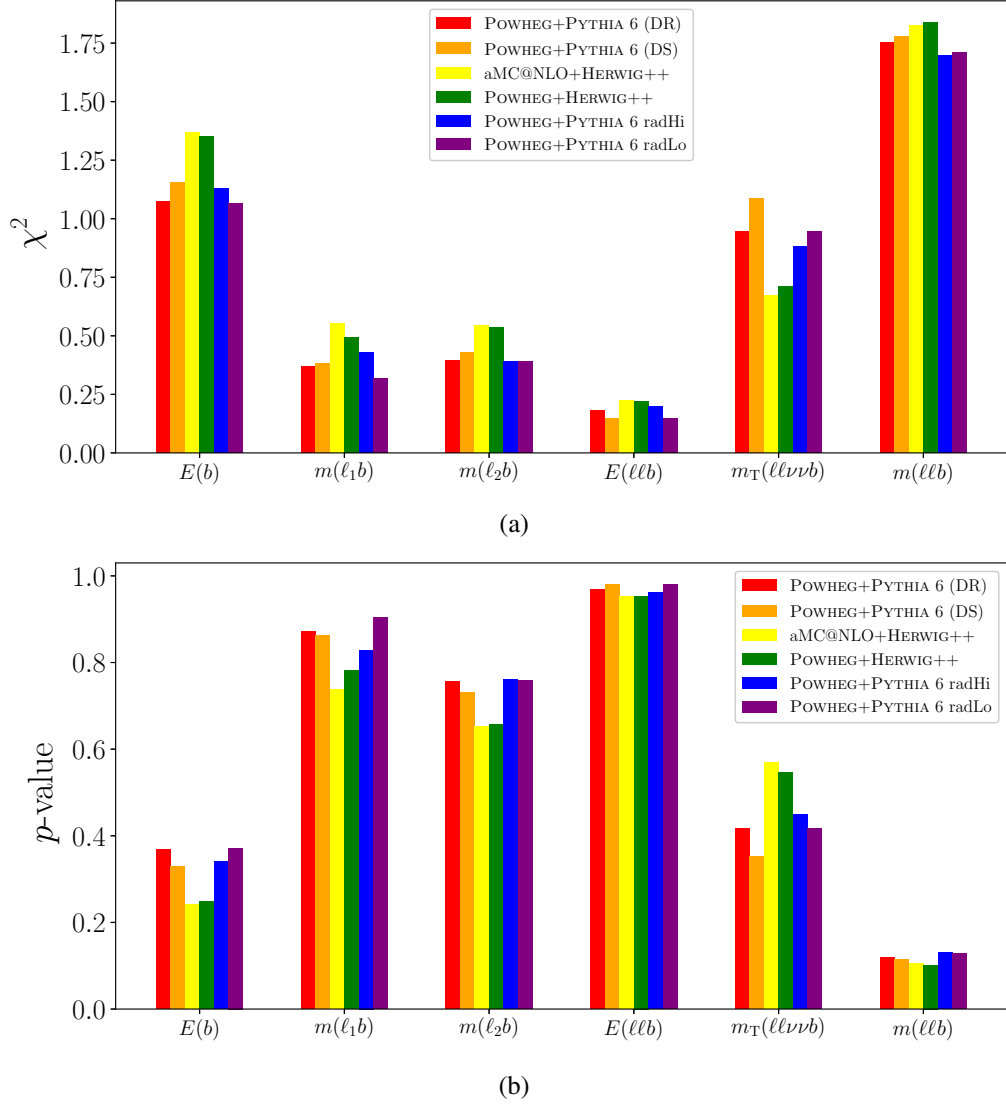


Figure 7.8: Reduced χ^2 and p -values for the measured normalised cross-sections compared to various particle-level MC predictions.

Adversarial neural network training

This chapter explores a deep learning technique for a further reduction of the systematic uncertainties in the inclusive cross-section measurement. A brief introduction to deep learning is given in Section 8.1, followed by the concept of adversarial networks which can include systematic deviations during training in Section 8.2. Lastly, a training is performed in the 2j2b region and an improvement in overall precision of the measurement is seen, comparable with the previous results from the BDT discriminant.

8.1 Introduction

The field of machine learning, such as aforementioned BDT technique, has been assisting high energy physics for decades. It is developing rapidly and provides advanced and intelligent methods for data analysis, pattern recognition, and model inference [210, 211]. Deep learning, a technique for implementing machine learning,¹ has revealed extraordinary performance in several domains. Applying deep learning to high energy physics promises improvements in many fronts, such as detector simulation and calibration, jet reconstruction, flavour tagging and data analysis.

8.1.1 Information theory

Information theory is a branch of applied mathematics that quantifies the amount of information an event conveys. Intuitively, a more likely event should have less information content and independent events should have an additive information content. An operation that satisfies both properties is the negative logarithm of the probability density function $P(x)$ for the event $x = x$, called the *self-information*:

$$I(x) = -\log P(x).$$

The self-information deals only with a single outcome. Information of an entire probability distribution is the expectation of the self-information, also known as the differential entropy (continuous case) or

¹ In the thesis, deep learning and machine learning are treated interchangeably despite their different scopes.

the Shannon entropy (discrete case):

$$H(P(x)) = - \int P(x) \log P(x) dx = - \sum_x P(x) \log P(x).$$

It is analogous to the entropy in statistical thermodynamics and gets maximum if $P(x)$ is a uniform distribution.

Furthermore, a measure of the difference between two probability distributions $P(x)$ and $Q(x)$ is defined by the Kullback-Leibler (KL) divergence:

$$D_{\text{KL}}(P||Q) = - \sum_x P(x) \log \frac{P(x)}{Q(x)} = - \sum_x [P(x) \log P(x) - P(x) \log Q(x)]$$

In other words, it is the expectation of the logarithmic difference between $P(x)$ and $Q(x)$, where the expectation is taken using the probabilities $P(x)$. Without losing too much generalisation, we define the *cross-entropy* as

$$H(P, Q) = H(P) + D_{\text{KL}}(P||Q) = \sum_x P(x) \log Q(x).$$

Minimising the cross-entropy with respect to $Q(x)$ is equivalent to minimising the KL divergence, because $Q(x)$ does not participate in the omitted term. Moreover, minimising the cross-entropy is also equivalent to maximising the likelihood of a model Q_{model} with a parameter θ given $\mathcal{P}_{\text{data}}$. The reason is the following: a maximum likelihood estimator, $\hat{\theta}$, is given by the following equations:

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta} \prod_i Q_{\text{model}}(x_i, \theta) \\ &= \arg \max_{\theta} \sum_i \log Q_{\text{model}}(x_i, \theta). \end{aligned}$$

Since x_i follows an underlying distribution $\mathcal{P}_{\text{data}}(x)$, the summation can be expressed as an expectation of the distribution:

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta} \sum_x \mathcal{P}_{\text{data}}(x) \log Q_{\text{model}}(x, \theta) \\ &= \arg \max_{\theta} H(\mathcal{P}_{\text{data}}, Q_{\text{model}}). \end{aligned}$$

8.1.2 Feed-forward networks

Although artificial neural networks were initially inspired by the biological neural networks that constitute animal brains, modern deep learning goes beyond the neuroscientific perspective. Instead of trying to simulate the brain, neural networks attempt to approximate the underlying functions relating inputs and outputs. Deep feed-forward networks are the quintessential deep learning models. An example is shown in Fig. 8.1. It comprises an input layer, an output layer, and several hidden layers, and each layer contains a certain number of nodes, where the nodes in one layer are connected to the nodes in adjacent layers. In a dense network, the nodes in one layer are connected to all nodes in adjacent layers. Information flows through the layers via the connections between the nodes and no circles or feedback connections exist. A linear combination of the weighted inputs is converted to an

output in a node via an *activation* function. Each connection is associated with a weight which will be updated during training to reduce the loss value, a value that measures the deviation of the prediction and the true value.

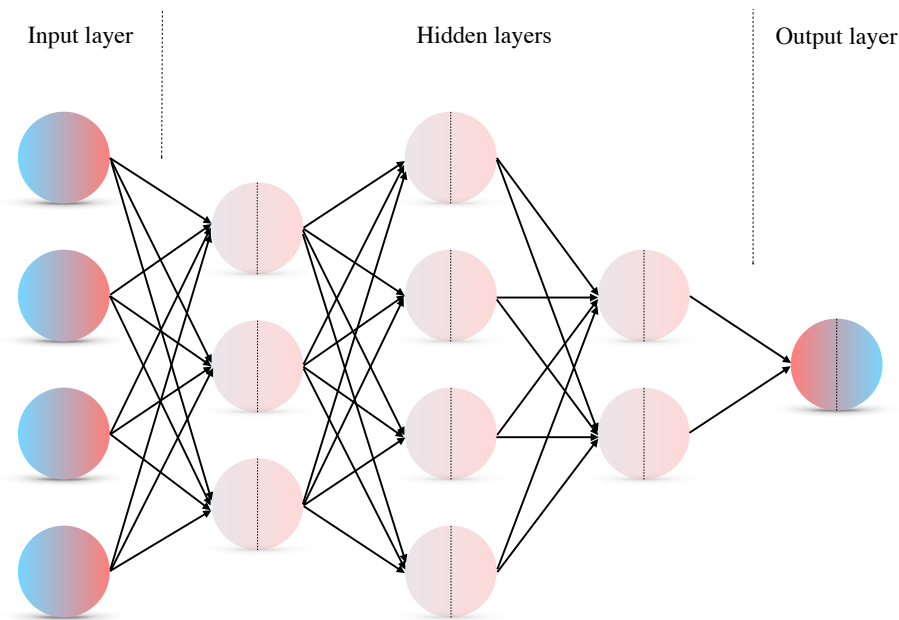


Figure 8.1: A graph of a five-layer feed-forward network. Nodes are represented by circles and weights are associated with arrows. Number of nodes and layers can be customised depending on the problem to be solved. Information flows from left to right.

The underlying function that a feed-forward network tries to represent is not necessarily linear. Nonlinear elements are incorporated via the activation functions. Table 8.1 summarises some typical activation functions. A binary step function would be the simplest one. It outputs 1, if the input sum is above a certain threshold, or 0 if the input sum is below a certain threshold. In practice, however, zero derivatives almost everywhere make the network impossible to learn with the gradient-based learning methods (see Section 8.1.5). The sigmoid and hyperbolic tangent smooth the behaviour close to the threshold (usually at 0) and add an appreciable derivative in a wider range. These functions are valuable for the output layers, but are still not ideal for the hidden layers. Gradients in large absolute values are asymptotically vanishing, which makes the learning very slow in these regions. In addition, certain information from the input is lost since the output is squeezed to a small range. The rectified linear units (ReLU) give a satisfactory solution: the output value is zero when the input value is less than zero, and greater than or equal to zero when the output is equal to the input. When the input value is positive, the derivative is 1, hence there is no squeezing effect. The gradient at zero is not defined, but it is not worrisome since taking the exact value zero would never happen. With so many nice features, the ReLU are recommended as the default for hidden layers. Nevertheless, because of the zero gradients in the negative regime, once the inputs get negative, they will refuse to update. This problem, called dying ReLU problem, can cause neurons to not respond, making a substantial part of the network passive. Several generalised functions try to overcome this shortcoming, for instance, leaky rectified linear unit (leaky ReLU), exponential linear unit (eLU) and so on, by adding finite

slopes in the negative regime in different ways.

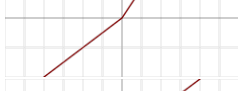
Name	Plot	Equation
Binary step		$g(z) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$
Sigmoid		$g(z) = \sigma(z) = \frac{1}{1 + e^{-x}}$
Hyperbolic tangent		$g(z) = \tanh(z) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
ReLU		$g(z) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$
Leaky ReLU		$g(z) = \begin{cases} 0.01x & x < 0 \\ x & x \geq 0 \end{cases}$
ELU		$g(z; \alpha) = \begin{cases} \alpha(e^x - 1) & x < 0 \\ x & x \geq 0 \end{cases}$

Table 8.1: Distributions and mathematical expressions of some example activation functions in deep learning. Figures are taken from [212].

8.1.3 Model capacity

Training error is the discrepancy between predictions and the true values when the trained model is applied to the training sample. Although the training data has already been used to train the model, it does not mean that the model will accurately perform when applied back on the training data itself. The test error, on the other hand, is the discrepancy measured using a test sample that was collected separately from the training set. The model is predictive if the two errors are small and close to each other. Similar to what is described in Section 5.1.6, underfitting occurs when the model is unable to obtain a sufficiently low error on the training set, while overfitting occurs when the gap between the training error and test error is too large.

To avoid overfitting and underfitting of a model, its capacity should be considered and altered. The capacity of a model is loosely defined as how complex a relationship this model can represent. A sufficient capacity leads to an ability to solve complex tasks, but when the capacity is higher than needed, it may overfit. On the other hand, a model with a lower capacity is unable to solve complex tasks, bringing an underfit. The capacity of a neural network can be adjusted by altering its number of hidden layers and number of nodes in each layer. Roughly speaking, a model with more layers and more nodes has a higher capacity.

8.1.4 Optimisation

After choosing an optimal capacity for a task, the network is trained to reduce the training error by updating the weights. Optimisation is a procedure of minimising or maximising a certain objective function or *loss* function, $f(w)$, by altering w . The loss function can be the cross-entropy of the predicted distribution and the true distribution. The minimum of the cross-entropy gives the smallest training error by the model.

Gradient descent is an optimisation algorithm to minimise a function by an iterative procedure. To minimise a function $f(\vec{w})$ with a starting point $\vec{w}_0 = \vec{a}_0$, we would like to find the direction which causes the most rapid decline in the function value. The direction of the negative gradient of $\vec{w}_n$ at $\vec{a}_n$ gives a proposal for an update:

$$\vec{w}_{n+1} = \vec{w}_n - \epsilon \nabla_{\vec{w}} f(\vec{a}_n).$$

The learning rate, ϵ , is a hyperparameter of the method to determine the size of the moving step.

If the training sample is too large, the learning will be very slow because it updates the parameters with the complete training set. The stochastic gradient descent (SGD) updates the parameters with a subset of the training sample. It relies on the fact that the gradient is an expectation, which can be estimated using a subset. The size of the sub-sample, usually called the batch size, is a hyperparameter in SGD.

In some poorly conditioned objectives, the minimum may not lie on the direction of the negative gradient descent, sometimes even much further away. For example in a long, narrow valley, the minimum is towards the end of the valley. The SGD will lead to oscillation across the valley since the negative gradient will point down one of the steep sides rather than along the valley towards the optimum. The *momentum* of the SGD corrects the current gradient in each update by adding some contributions from previous gradients and introduces a hyperparameter to control the decay of the contributions with the updates.

8.1.5 Back-propagation

Back-propagation, or *backprop* for short, is a way to calculate the gradient descent of a loss function. The procedure is based on the application of the chain rule and computationally proceeds “backwards” with respect to the computations of the loss value. To demonstrate this, a simplified version of Fig. 8.1 is depicted in Fig. 8.2, with additional details. The input x that represents the initial information is propagates forward to the nodes at each hidden layer, where a linear summation $z(w)$ and a nonlinear activation $g(z)$ are sequentially performed. Finally the network outputs $\hat{y}$, and loss value can be computed. To update w_1 , the gradient can be expressed using the chain rule as

$$\nabla_{w_1} \text{Loss} = \frac{\partial \text{Loss}}{\partial w_1} = \frac{\partial \text{Loss}}{\partial g_4} \cdot \frac{\partial g_4}{\partial z_4} \cdot \frac{\partial z_4}{\partial g_3} \cdot \frac{\partial g_3}{\partial z_3} \cdot \frac{\partial z_3}{\partial g_2} \cdot \frac{\partial g_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial g_1} \cdot \frac{\partial g_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}.$$

With known $\vec{g}(z)$ and $\vec{z}(w)$, these derivatives can be calculated analytically or numerically.

8.2 Invariant training on nuisance parameters

One of the challenges of applying machine learning to physics measurements is the need to incorporate systematic uncertainties. Previously, as demonstrated in Chapter 5, a discriminator is built on nominal

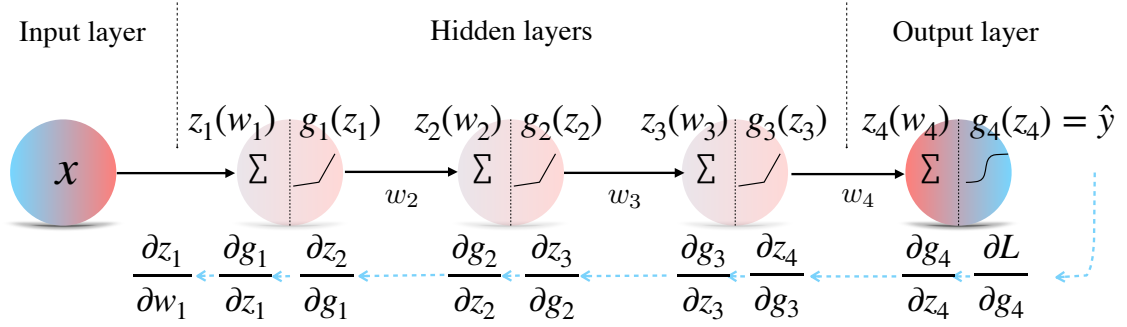


Figure 8.2: A schematic picture showing how backprop works.

samples and applied on systematic deviations. It lacks a cross-talk between the optimisation of the separation and systematic deviations. Various ideas were explored to balance the separation and the size of the uncertainty. For instance, certain variables that introduce large uncertainties can be excluded from the variable list used in the training. Or use mixed nominal and systematic samples for training, so that the model could also pick the features of systematic variations. However, it is hard to disentangle the explicit effect of the systematic uncertainties in the results with these ideas.

A new idea to deal with this problem is to perform an adversarial training [213], implementing a pivotal discriminant with respect to the NPs that parametrise the systematic uncertainties [214]. The adversarial neural networks (ANN) comprise two networks f and r , where f plays the role of discriminating signal and backgrounds, and r tries to predict the NP values based on the output of f . If the discriminant is sensitive to the NPs, a well trained r can tell whether an event is from the nominal or systematic samples and gives a negative feedback to alter f . After some iterations, the responses of f for the nominal and systematic samples get closer and closer and it becomes harder and harder for r to predict the NP value. Thereafter the discriminant f is insensitive to the NPs and the systematic deviation on the final ANN output is small.

To formulate the method, a joint loss function $\mathcal{L}(\theta_f, \theta_r)$ is constructed from the individual losses $\mathcal{L}_f(\theta_f)$ and $\mathcal{L}_r(\theta_r)$:

$$\mathcal{L}(\theta_f, \theta_r) = \mathcal{L}_f(\theta_f) - \lambda \mathcal{L}_r(\theta_r).$$

λ is a positive value that controls relative contributions of both networks. Model parameters θ_f and θ_r are therefore estimated by finding the minimax solution

$$\hat{\theta}_f, \hat{\theta}_r = \arg \min_{\theta_f} \max_{\theta_r} \mathcal{L}(\theta_f, \theta_r).$$

With respect to f , minimising the joint loss is equivalent to minimising $\mathcal{L}_f(\theta_f)$ and maximising $\mathcal{L}_r(\theta_r)$ (due to the negative sign). In other words, f performs well at the classification task while r performs as poorly as possible. With respect to r , maximising the joint loss is equivalent to minimising $\mathcal{L}_r(\theta_r)$. So we want r to be as good as possible at the task of predicting NPs.

8.3 Implementation

The “DR vs DS” uncertainty is one of the dominant uncertainties in the result obtained from the BDT (see Table 6.2 and Fig. 6.4). Further investigation shows that this mainly comes from the $2j2b$ region, as this region is close to the $t\bar{t}$ phase space and thus is sensitive to the interference. Therefore, the implemented ANN mainly focuses on the “DR vs DS” uncertainty in the $2j2b$ region. However, the final results are extracted using all available systematic uncertainties for a fair comparison with the preceding BDT results.

8.3.1 Setup and training

The training sample is comprised of half of the $t\bar{t}$ nominal as the background sample and half of tW nominal (DR) and DS signal samples (see Tables 4.2 to 4.3). Net f is trying to separate tW nominal from $t\bar{t}$ nominal while net r differentiates tW nominal and DS samples. Other backgrounds are not included in the training due to their low number of events. The second halves are used for test purposes.²

KERAS [215] with the THEANO [216] backend is used to build an ANN and perform the training. Classifier f is a fully connected deep network with nine layers (seven hidden layers plus one input and one output layer).³ The number of nodes in the input layer is customised to be the number of input variables. Hidden layers have 40 nodes each, and use eLU as the activation function. The output layer contains 1 node for the ANN response, using the sigmoid activation. Adversary r is a shallow network⁴ that takes the output of f as its input. Technically, r contains the full architecture of f and shares parameters⁵ with it. Its unique hidden layer contains 100 nodes and the output layer has 1 node. The ReLU is used in the hidden nodes while the sigmoid is used in the output node. The architectures of f and r are tabulated in Table 8.2. The number of parameters in each layers is a multiplication of the number of nodes in the previous layer plus one times the number of nodes in that layer, where the additional one counts for a bias node in the previous layer. During the training of r , parameters that are shared with f are fixed. Parameter λ is found to be insensitive to the final results, and chosen to be 10 in this thesis. The input variable list simply takes the optimised variable list from the BDT (see Table 5.2), as they have quite good separation power.

Training starts with pretrainings of f and r to obtain reasonable initial values. Afterwards, f is trained to allow a good signal-background separation.⁶ Then r is optimised with f fixed to discriminate the systematic distribution from the nominal. After optimising r , the joint loss may not retain its minimum and therefore f is retrained. The iterative procedure gives the optima of both f and r , and the prediction of f is the response of the ANN. To find out the best number of iterations, a large number of iterations are checked and results are shown in Fig. 8.3. The number of iteration is chosen to be 100 such that $\mathcal{L}(\theta_f)$ (in red) reaches a minimum while $\mathcal{L}_r(\theta_r)$ (in green) reaches a maximum.

² In practice, the fractions for the training and test samples are 0.6 and 0.4, respectively.

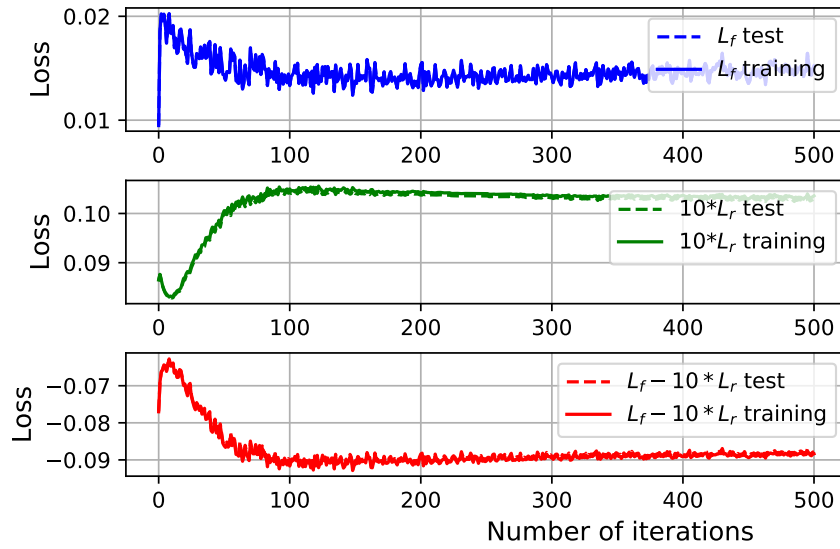
³ The number of layers and nodes are studied using a feed-forward network that tries to separate signal from background.

⁴ A shallow architecture is chosen here to reduce the risk of overtraining and the results are satisfactory. Deep networks should also work in theory.

⁵ Parameter sharing is a way to force sets of parameters to be equal. It is an important method of regularisation.

⁶ Technically, f is trained via a composed network fr constructed from f and r , in order to compute the joint loss function that includes both f and r parts.

Layer (type)	Output shape	Parameters No.	Free?	Layer (type)	Output shape	Parameters No.	Free?
f input (InputLayer)	(11)	0	—	f input (InputLayer)	(11)	0	—
f layer1 (Dense)	(40)	480	Yes	f layer1 (Dense)	(40)	480	No
f layer2 (Dense)	(40)	1640	Yes	f layer2 (Dense)	(40)	1640	No
f layer3 (Dense)	(40)	1640	Yes	f layer3 (Dense)	(40)	1640	No
f layer4 (Dense)	(40)	1640	Yes	f layer4 (Dense)	(40)	1640	No
f layer5 (Dense)	(40)	1640	Yes	f layer5 (Dense)	(40)	1640	No
f layer6 (Dense)	(40)	1640	Yes	f layer6 (Dense)	(40)	1640	No
f layer7 (Dense)	(40)	1640	Yes	f layer7 (Dense)	(40)	1640	No
f output (Dense)	(1)	41	Yes	f output (Dense)	(1)	41	No
Total parameters: 10361				r layer1 (Dense)	(100)	200	Yes
Trainable parameters: 10361				r output (Dense)	(1)	101	Yes
Non-trainable parameters: 0				Total parameters: 10662			
				Trainable parameters: 301			
				Non-trainable parameters: 10361			

Table 8.2: The architecture of network f (left) and r (right).Figure 8.3: The evolution of values of $\mathcal{L}_f$, $\lambda\mathcal{L}_r(\theta_r)$ and $\mathcal{L}(\theta_f, \theta_r) = \mathcal{L}_f - \lambda\mathcal{L}_r(\theta_r)$, respectively, with 500 iterations and $\lambda = 10$. In all cases, both training losses (solid lines) and test losses (dashed lines) are compared and the differences are small.

8.3.2 ANN response

The “DR vs DS” deviation on the ANN response in the 2j2b region is shown in Fig. 8.4(a), compared with the same systematic variations for the BDT response (Fig. 8.4(b)). The ratio distribution is flatter, which means the feedback from network r has an effect. Especially in the signal (larger response) region, smaller deviations indicates a lower uncertainty from this source.

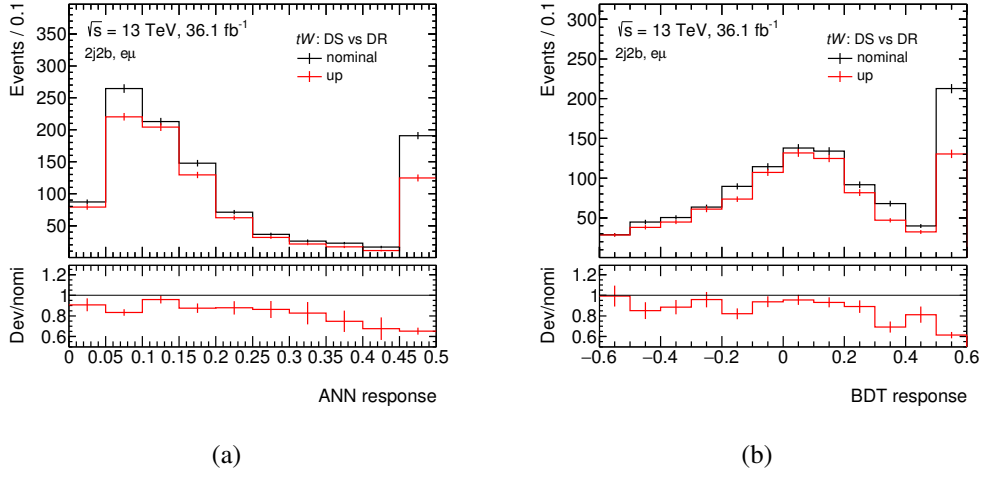


Figure 8.4: The ANN responses (a) and the BDT responses (b) of the tW process nominal and DR in the 2j2b region.

8.3.3 Fit results and discussions

To qualify the performance of the ANN, a maximum likelihood fit is performed on the Asimov dataset and real data with all available systematics and all three regions. The BDT distributions are used in the 1j1b and 2j1b regions while the ANN distributions are used in the 2j2b region. The fit strategy is the same as the one used in Section 6.4. The following ANN results are obtained (compared to the BDT results in Section 6.4.1):

(ANN)	Expected	Observed	(BDT)	Expected	Observed
μ_{sig}	$1.00^{+0.18}_{-0.16}$	$0.99^{+0.18}_{-0.17}$	μ_{sig}	$1.00^{+0.20}_{-0.17}$	$1.14^{+0.26}_{-0.22}$
$\mu_{t\bar{t}}$	$1.000^{+0.033}_{-0.031}$	$1.005^{+0.033}_{-0.032}$	$\mu_{t\bar{t}}$	$1.000^{+0.032}_{-0.031}$	$1.004^{+0.033}_{-0.032}$

The observed signal strength corresponds to a measured cross-section of $\sigma_{tW} = 71^{+13}_{-12}$ pb, while the preceding result given by the BDT discriminant is $\sigma_{tW} = 82^{+19}_{-16}$ pb. The two results are consistent within the uncertainty, while the one from ANN gets more precise (smaller errors).

The pull and impact summary plots for the NPs in the fit are given in Fig. 8.5 (the full list can be found in Appendix E). Compared with the results from the BDT response (Fig. 6.4), the “DR vs DS” after-fit impact, although still at the top, is reduced. Although all the other systematics except for the “DR vs DS” were not fed to the training, due to the correlations, some of them are reduced as well. For example, the $t\bar{t}$ ME and PS were in the top ten list while now they drop out of the list. The tW

and $t\bar{t}$ ISR/FSR are reduced as well. This effect can be also seen from the correlation matrices of the systematic uncertainties given in Fig. 8.6. Compared with Fig. 6.3, it is clear that the correlation between the “DR vs DS” with other systematics reduced a lot. For example, in the data fit, the correlations with the $t\bar{t}$ ISR/FSR is reduced to from 0.3 to 0.1, while the one with the $t\bar{t}$ ME is reduced to from 0.4 to 0.3. The correlation with the “JES: Eff1” is also slightly reduced. The one with the “JES: PU ρ ” goes down dramatically, from 0.7 to 0.3. The distributions of the systematic deviations are shown in Figs. 8.7 to 8.8. In general, the ratios are flatter in the ANN response compared with the BDT response.

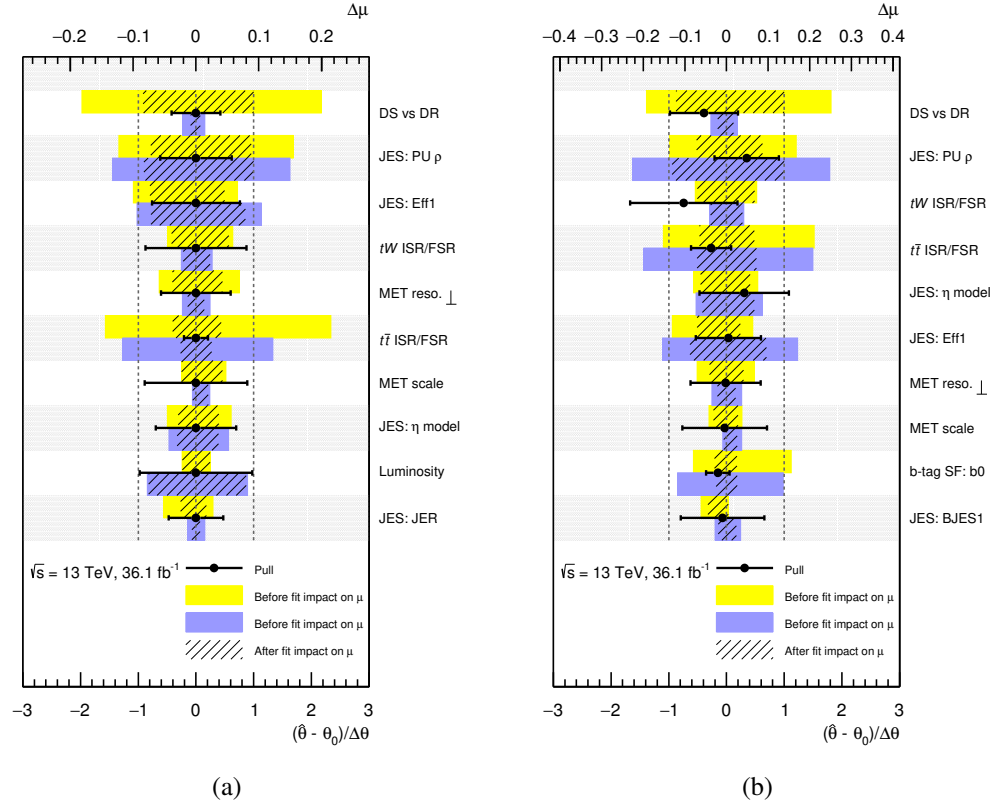
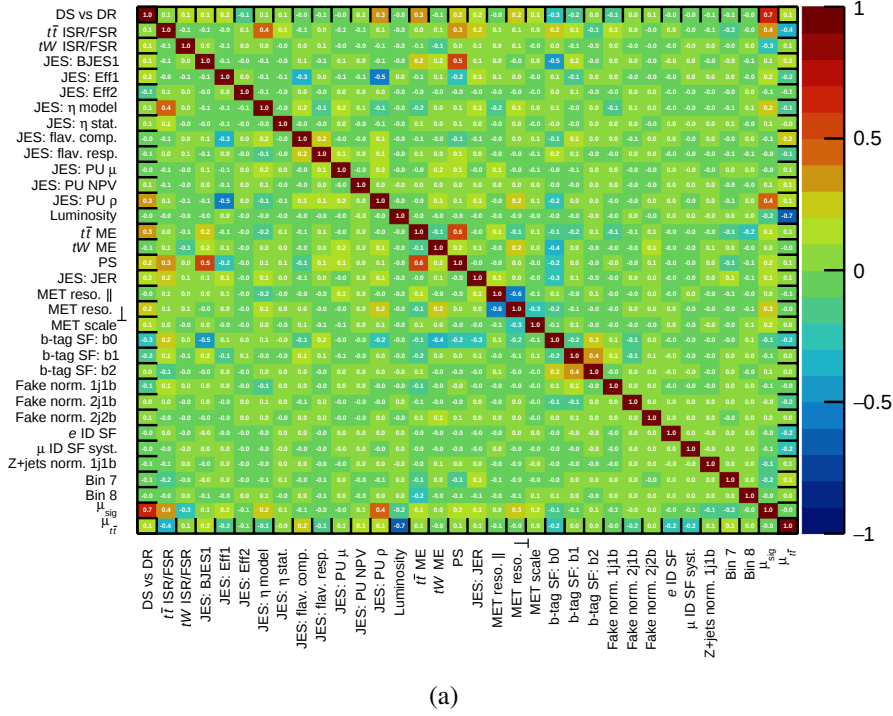


Figure 8.5: Pull (corresponding to the lower x -axis) and impact on μ_{sig} (corresponding to the upper x -axis) of the top 10 NPs ordered by the post-fit impact. The ranking in Asimov (a) and in data (b) are similar, but not identical. Bands in yellow represent the floated version of the impact, i.e. the NP being studied is fixed to its up/down variations while others are floated. Bands in violet represent the fixed version of impact, i.e. the NP being studied is fixed to its up/down variations while others are fixed to their global fit values. The hatched lines follows the same definitions of the bands albeit presenting after fit impact.

Correlation Matrix



Correlation Matrix

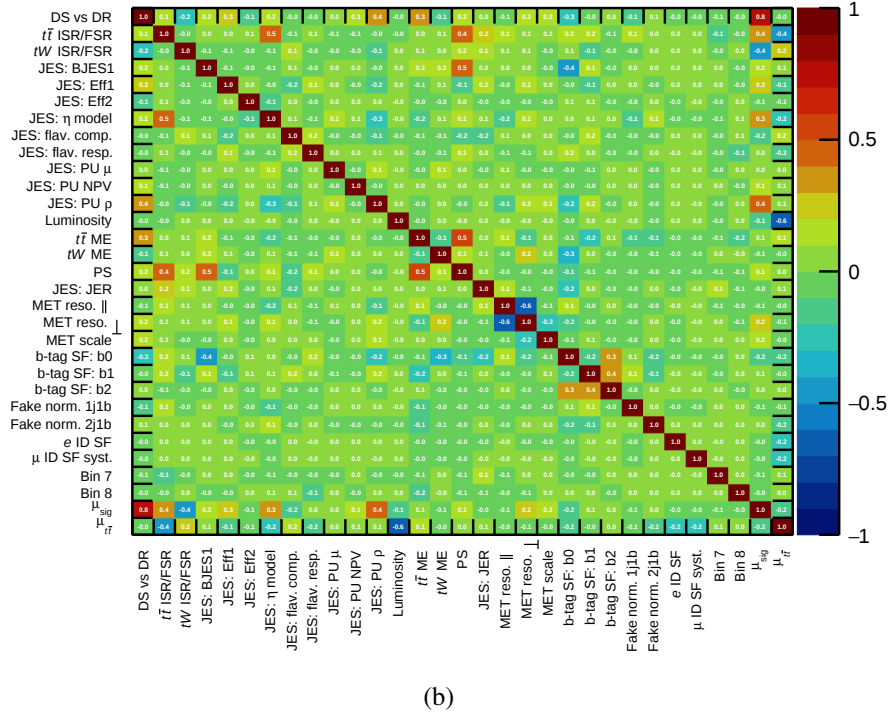


Figure 8.6: Correlation matrices of the systematic uncertainties. The results shown represent fits to the Asimov dataset (a) and real data (b). Rows or columns with no entries of magnitude greater than 0.2 are pruned to simplify the plots.

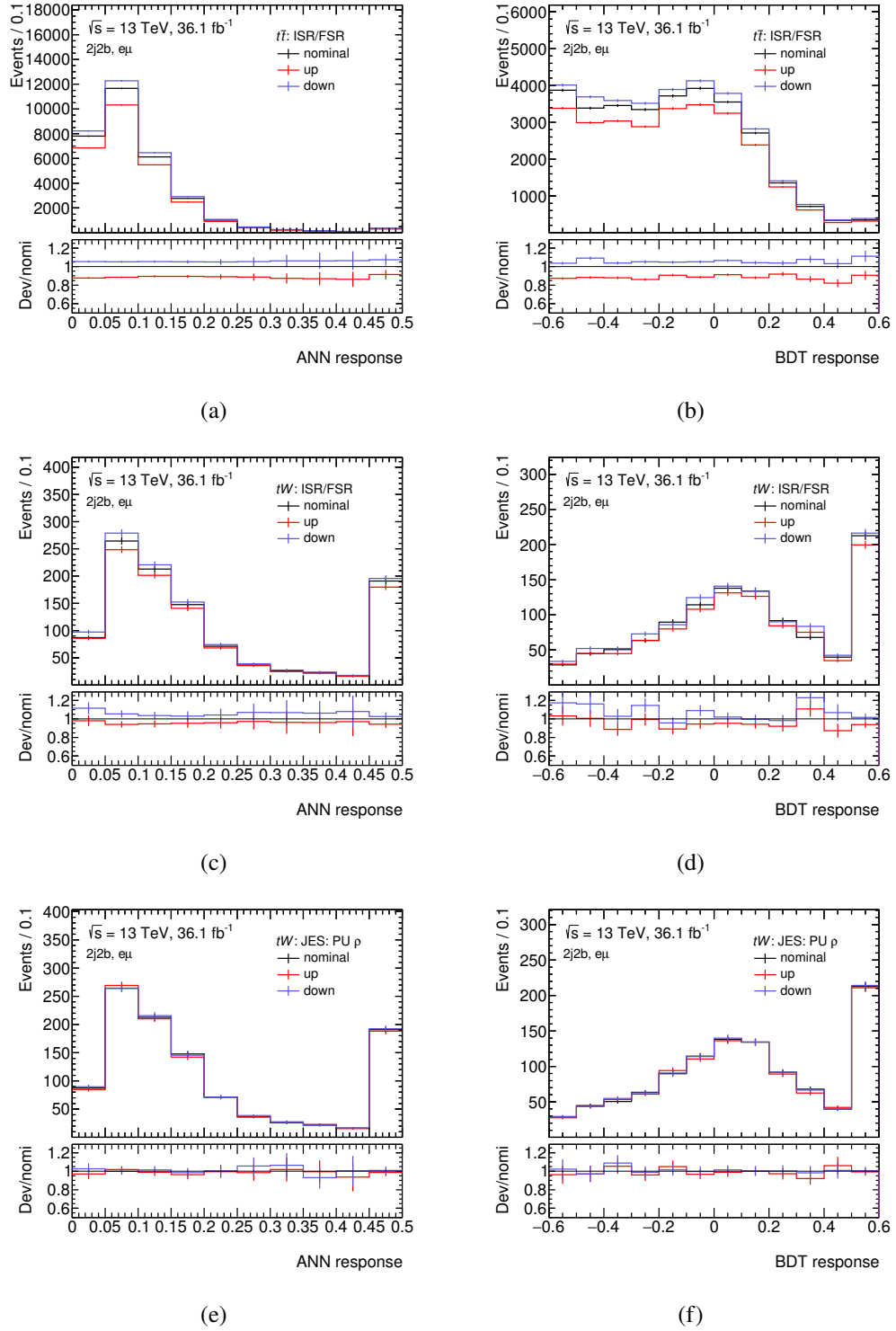


Figure 8.7: Comparison of ANN response (left) and BDT response (right) of some systematic deviations in the 2j2b region.

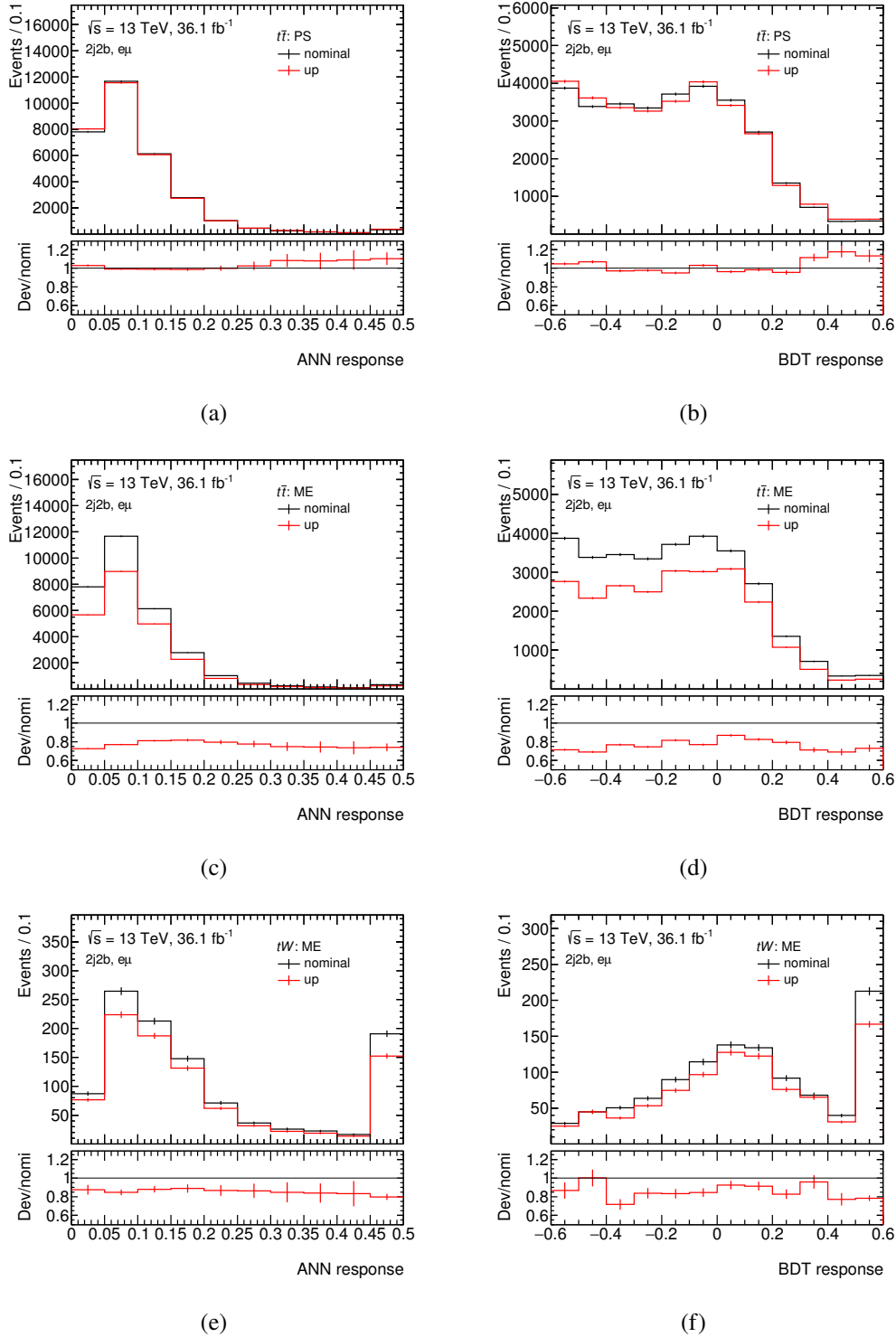


Figure 8.8: Comparison of ANN response (left) and BDT response (right) of some systematic deviations in the $2j2b$ region. These systematic deviations only contain one side and symmetrised to the other side in the fit.

Conclusions

This thesis presents the inclusive cross-section measurement and so far the first differential cross-section measurement for the associated production of a top quark and a W boson, known as the tW -channel, at $\sqrt{s} = 13$ TeV. The dataset corresponds to an integrated luminosity of 36.1 fb^{-1} , collected by the ATLAS detector at the LHC in 2015 and 2016. This process is predicted by the SM. Any deviation from the SM prediction could indicate modifications from new physics.

The analysis uses dilepton events with at least one b -tagged jet to suppress backgrounds. Events are divided into signal and control regions based on the number of jets and b -jets. The tW signal is separated from the $t\bar{t}$ background using a BDT discriminant, further improved by an ANN discriminant. A likelihood fit is performed on the discriminant response to extract the inclusive cross-section for the signal process. The inclusive cross-section is measured to be (using the ANN discriminant)

$$\sigma_{\text{meas.}}(\sqrt{s} = 13 \text{ TeV}) = 71 \pm 2 \text{ (stat.)}_{-11}^{+12} \text{ (syst.)} \pm 3 \text{ (lumi.) pb},$$

while the theoretical prediction for this process is

$$\sigma_{\text{theo.}}(\sqrt{s} = 13 \text{ TeV}) = 71.7 \pm 1.8 \text{ (scale)} \pm 3.4 \text{ (PDF) pb}.$$

The measured value agrees quite well with the theoretical prediction, indicating no new physics observed.

Dominant uncertainties arise from JES, signal modelling, and $t\bar{t}$ background modelling. The result of the measurement presented in this thesis is summarised together with other measurements of the single top-quark production at the LHC in Fig. 9.1. The cross-section measured in this analysis is consistent with the theory predictions and with a recent CMS result.

Furthermore, a tW enriched region is selected and the differential cross-section is measured for several particle-level observables. Six observables are chosen, constructed from the masses and energies of leptons and jets as well as the transverse momenta of neutrinos. Measurements are normalised with the fiducial cross-section, causing several of the main uncertainties to cancel out. Dominant uncertainties arise from limited data statistics, signal modelling, and $t\bar{t}$ background modelling. Results are found to be in good agreement with predictions from several MC event generators.

The results of this thesis are updates of the preceding results published in the two papers [180, 209]. The first paper reported the inclusive cross-section using only 3.2 fb^{-1} data, taken in 2015. The second

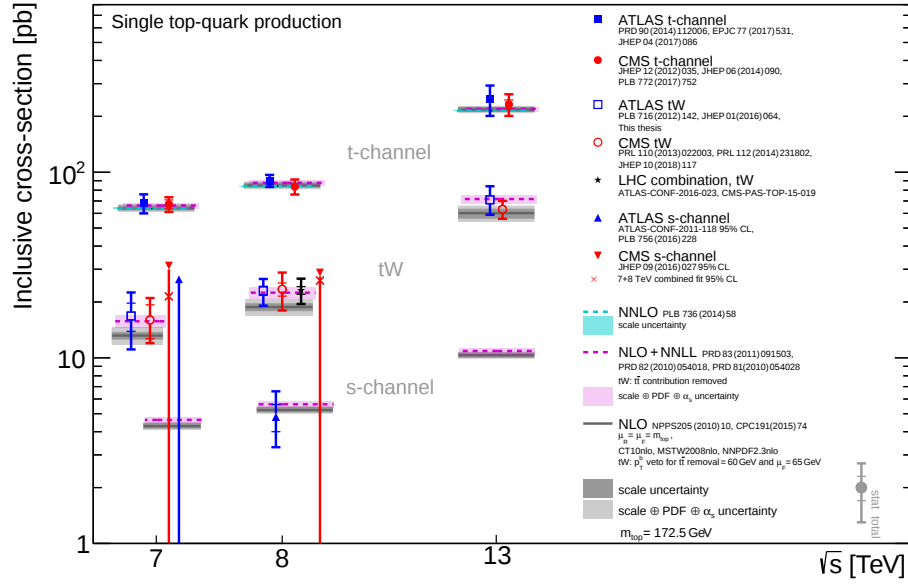


Figure 9.1: The results of this and other measurements of the production of single top-quarks at the LHC, compared to theory predictions. Two measurements of tW at $\sqrt{s} = 13$ TeV are added with respect to Fig. 2.7. The one in blue represents the results from this thesis, while the one in red is a recent result from CMS [217].

paper presented the differential cross-section results with the same amount of data as that used in the thesis. However, this thesis is an update with better object reconstruction and calibration and better systematic uncertainties. The study using the ANN is newly introduced in this thesis.

Appendix

Generator information of samples used in the fake-lepton events estimates

Jet filter	W decay	Cross-section [pb]					
		Jet p_T [GeV]					
		0–70	70–140	140–280	280–500	500–1 000	> 1 000
Light-jet	$e\nu_e$	15 300	619	197	38.3		
	$\mu\nu_\mu$	15 300	619	198	38.3		
	$\tau\nu_\tau$	15 300	619	197	38.4		
c jet	$e\nu_e$	2 430	224	95.2	22.4		
	$\mu\nu_\mu$	2 430	223	96.5	22.4		
	$\tau\nu_\tau$	2 440	223	94.1	22.3		
b jet	$e\nu_e$	832	94.9	35.9	9.59		
	$\mu\nu_\mu$	829	76.7	36.3	8.77		
	$\tau\nu_\tau$	838	95.3	34.8	9.49		
—	$e\nu_e$					14.6	1.20
	$\mu\nu_\mu$					14.6	1.20
	$\tau\nu_\tau$					14.6	1.20

Table A.1: The cross-sections for W + jets as a function of jet p_T slices and lepton-jet flavours of the W decay. They are normalised to NLO (LO for more than 2 jets) theoretical calculation. Jet flavours are simulated inclusively when jet $p_T > 500$ GeV. Numbers are extracted from [218].

Jet filter	Lepton	Cross-section [pb]		
		Jet p_T [GeV]		
		0–70	70–280	> 280
Light-jet	ee	2 270	43.9	2.71
	$\mu\mu$	2 270	43.8	2.70
	$\tau\tau$	2 270	43.9	2.74
b jet	ee	81.7	5.35	0.487
	$\mu\mu$	81.1	5.50	0.513
	$\tau\tau$	81.6	5.41	0.499

Table A.2: The cross-sections for $\ell\ell$ + jets as a function of jet p_T slices and lepton-jet flavours. They are normalised to NLO (LO for more than 2 jets) theoretical calculation. Numbers are extracted from [219].

APPENDIX **B**

Rest BDT input variables

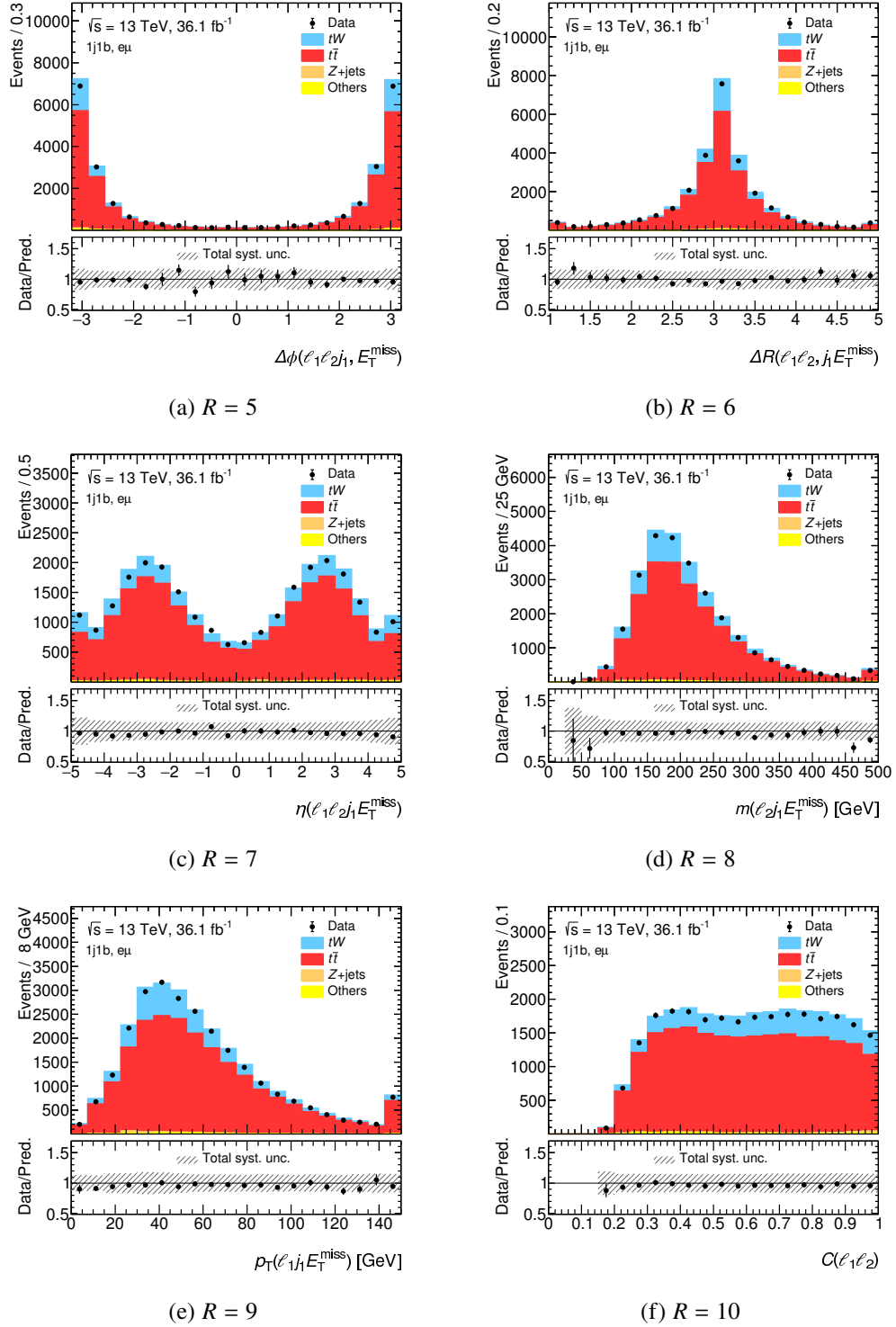
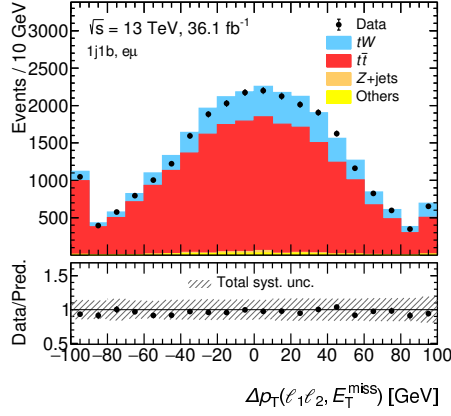
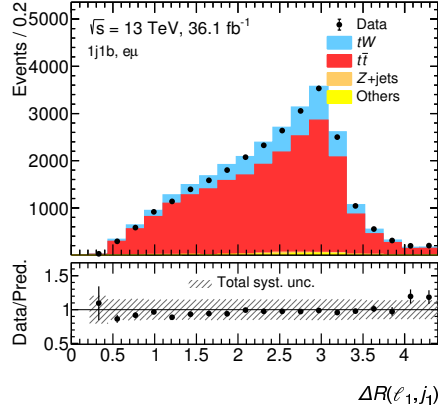


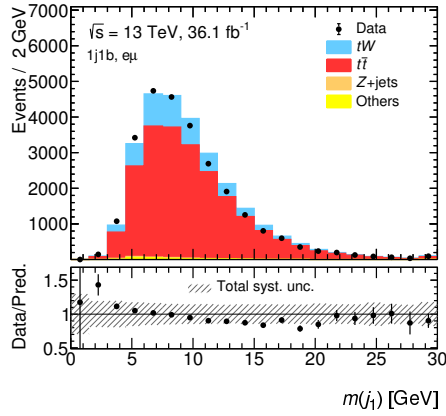
Figure B.1: The rest BDT input variables used in the 1j1b region (supplementary to Fig. 5.3). Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panel gives the yields in the number of events per bin, while the lower panel gives the ratios of the numbers of observed events to the total prediction in each bin. The sub-caption gives the rank R defined in Table 5.2.



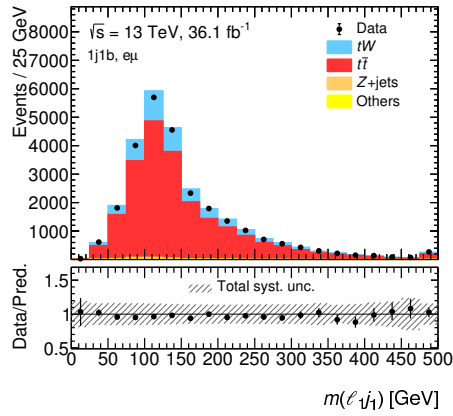
(a) $R = 11$



(b) $R = 12$



(c) $R = 13$



(d) $R = 14$

Figure B.2: The rest BDT input variables used in the 1j1b region (supplementary to Fig. 5.3 and Fig. B.1). Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panel gives the yields in the number of events per bin, while the lower panel gives the ratios of the numbers of observed events to the total prediction in each bin. The sub-caption gives the rank R defined in Table 5.2.

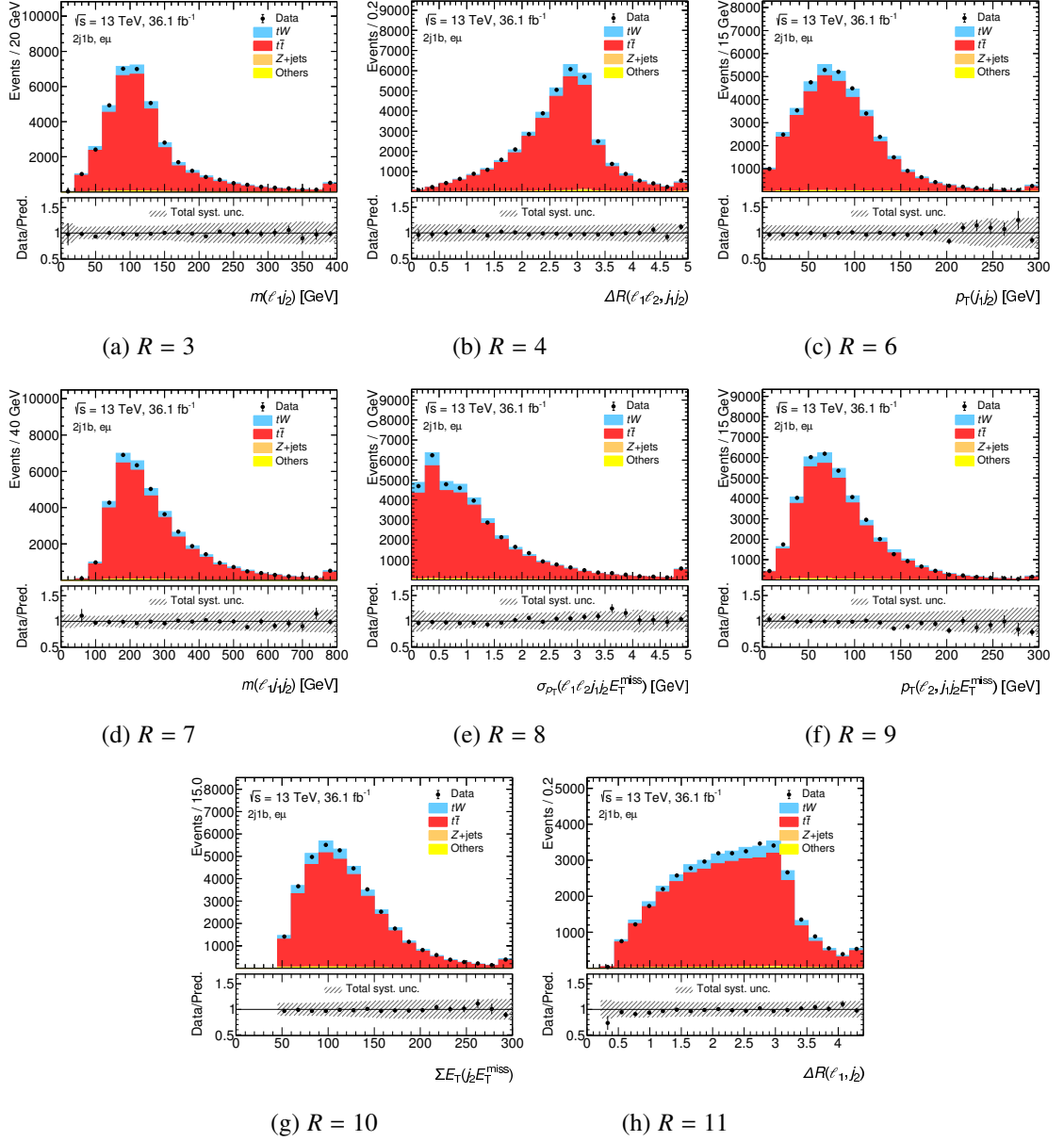


Figure B.3: The rest BDT input variables used in the 2j1b region (supplementary to Fig. 5.3). Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panel gives the yields in the number of events per bin, while the lower panel gives the ratios of the numbers of observed events to the total prediction in each bin. The sub-caption gives the rank R defined in Table 5.2.

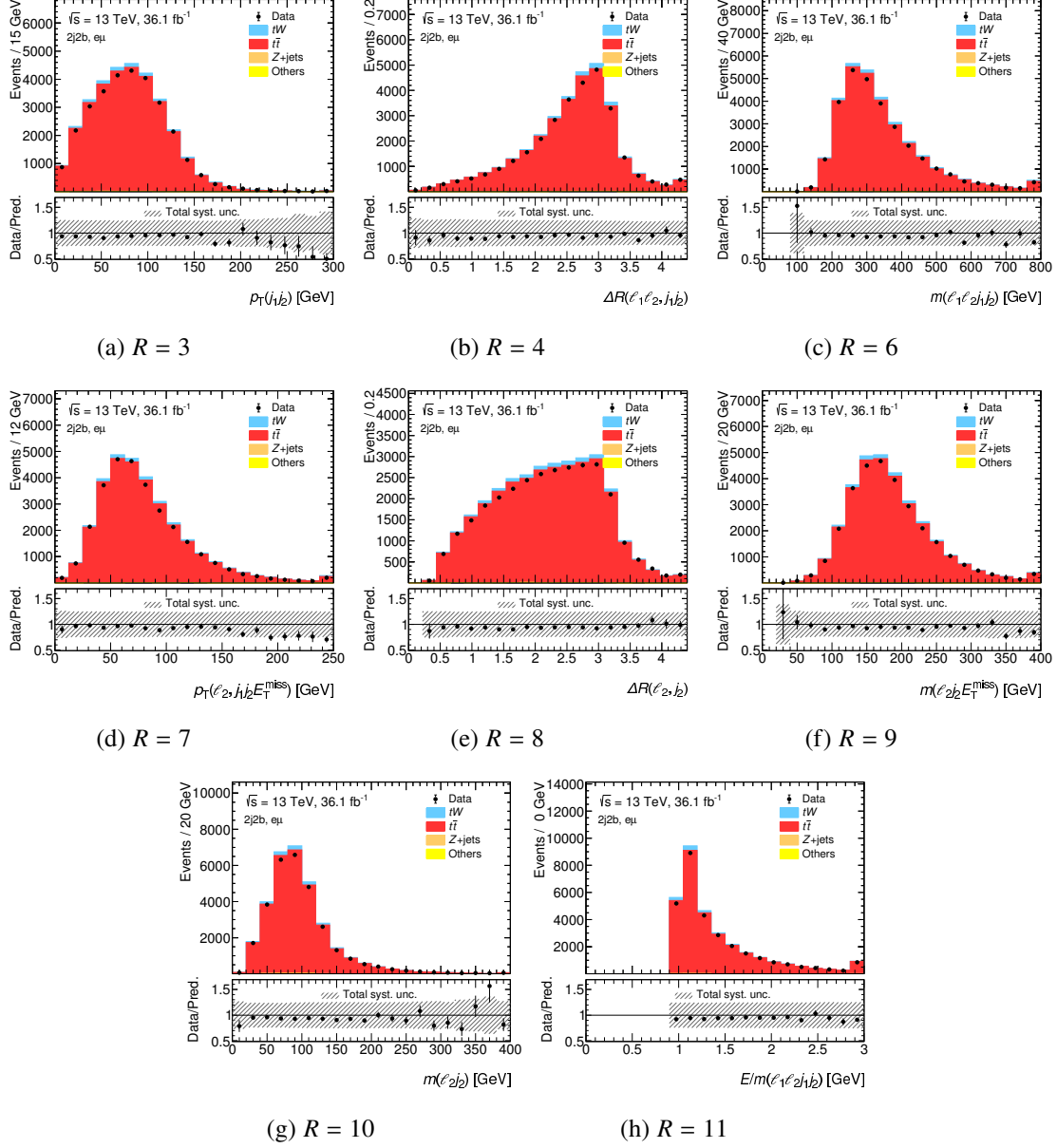
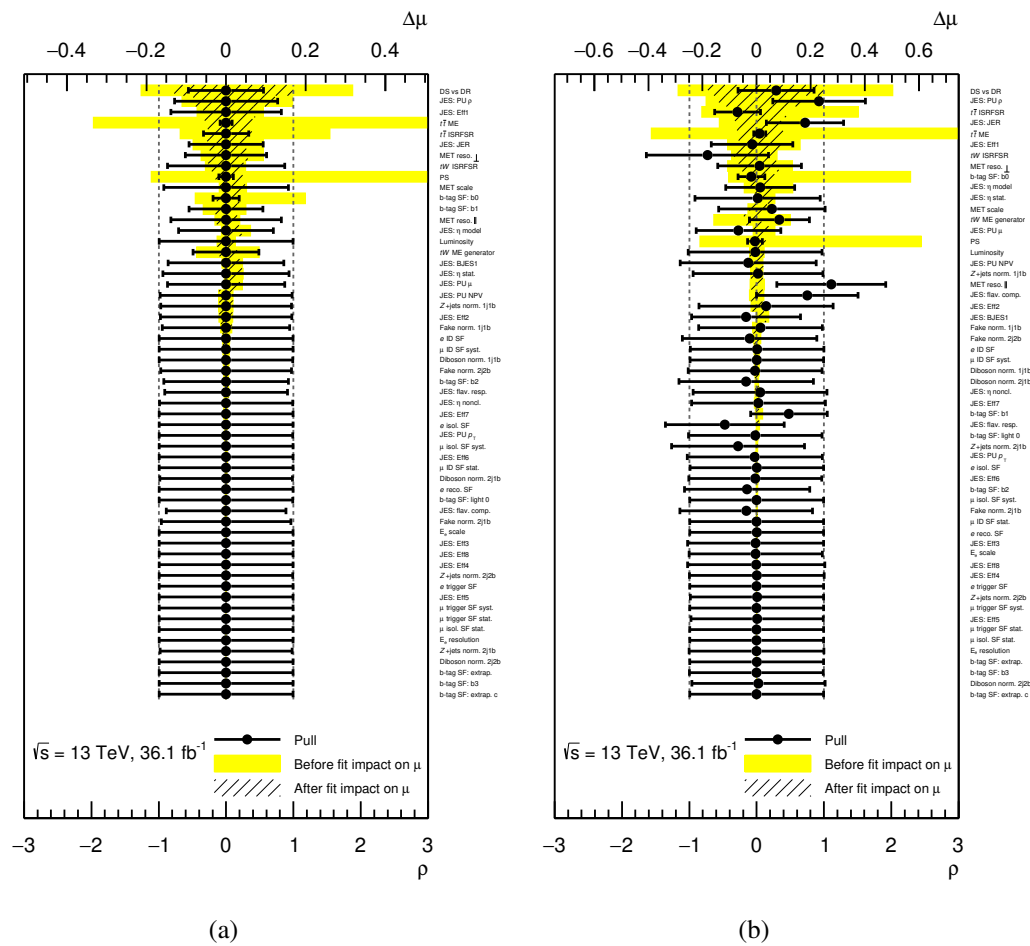


Figure B.4: The rest BDT input variables used in the 2j2b region (supplementary to Fig. 5.3). Signal and backgrounds are normalised to their theoretical predictions, and the error bands represent the total systematic uncertainties used in the thesis. The first and last bins of each distribution include underflow and overflow events, respectively. The upper panel gives the yields in the number of events per bin, while the lower panel gives the ratios of the numbers of observed events to the total prediction in each bin. The sub-caption gives the rank R defined in Table 5.2.

Pull and impact of the NPs (BDT)



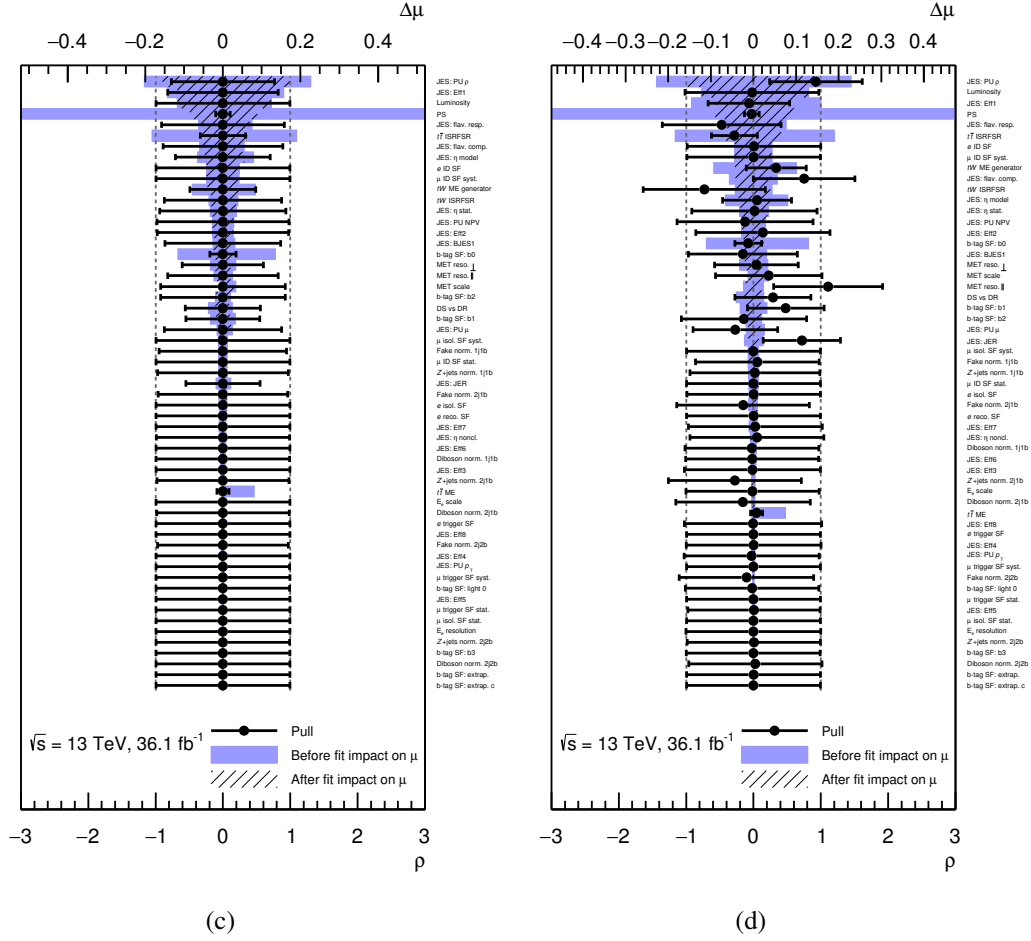


Figure C.1: Pull (corresponding to the lower x -axis) and impact on μ_{sig} (corresponding to the upper x -axis) of the NPs ordered by the post-fit impact. The ranking in asimov (a) and in data (b) are similar, but not identical. Floated version of impact for asimov (a) and data (b) and fixed version of impact for asimov (c) and data (d) are shown. The hatched lines follows the same definitions albeit presenting before fit impact.

APPENDIX **D**

Detailed uncertainties on the differential cross-sections

$E(b)$ bin [GeV]	[25, 60]	[60, 100]	[100, 135]	[135, 175]	[175, 500]
$(1/\sigma) d\sigma/dx$ [GeV $^{-1}$]	0.00422	0.00937	0.00428	0.00169	0.000801
Statistical uncertainty	22	12	27	48	11
MC stat. uncertainty (bootstrap)	2.9	2.5	4.3	5.1	1.4
Unfolding non-closure	-7.8	0	0	0	3.6
E_e ID SF	0.076	0.03	0.067	-0.47	-0.0016
E_e resolution	0.17	-0.11	0.2	0.65	-0.21
tW ISRFSR	7.3	-2.1	-2.8	2.6	-0.22
tW ME	11	-11	18	-15	3.1
tW PS	-7.1	6.1	-7.7	5	-1.6
DR vs DS	-5.3	3.9	-2.6	2.9	-1.8
Diboson norm.	-0.087	0.22	0.36	0.66	-0.66
Fake norm.	2.1	-0.5	-0.71	2.3	-0.67
JER	5.6	-6.7	12	12	-3.4
JES: η model	1.2	0.54	-0.07	-0.28	-1.3
JES: η stat.	0.97	0.14	-0.57	-0.12	-0.38
JES: Eff1	3.9	-1	-0.5	4.9	-1.8
JES: Eff2	1.7	-1.4	0.84	2.7	-0.17
JES: JES	4.5	-0.27	-2.6	2.4	-1.3
JES: PU μ	-1	0.19	0.51	-1.9	0.52
JES: PU ρ	2.4	0.77	-1.9	5.7	-2.8
JES: PU NPV	1.7	-0.66	-1.1	4.3	-0.51
JES: flav. comp.	0.14	0.83	-1.7	1	-0.53
JES: flav. resp.	0.2	-0.59	1.8	-1.8	0.17
Luminosity	-0.14	-0.033	-0.21	-0.77	0.45
MET reso. $\parallel$	0.88	0.5	-1.5	4.3	-1.5
MET reso. $\perp$	5.3	-2.1	2.6	-0.71	-1.3
MET scale	0.98	-0.3	-0.61	2.1	-0.32
Z + jets norm.	-3.2	0.63	1.7	-0.22	-0.019
$t\bar{t}$ ISRFSR	-15	0.94	5.4	12	1.2
$t\bar{t}$ ME	4.3	22	-2.1	-50	-19
$t\bar{t}$ PS	-0.26	-9	-17	18	18
$t\bar{t}$ norm.	-0.42	-0.097	-0.6	-2.2	1.3
b-tag SF: b0	6.5	-0.5	-0.76	-0.14	-2.5
b-tag SF: b1	-1.2	0.11	0.24	0.54	0.25
Total syst. uncertainty	27	28	30	59	28
Total uncertainty	35	31	40	76	30

 Table D.1: Normalised cross-sections differential in $E(b)$. Uncertainties are shown in percent.

$E(\ell\ell b)$ bin [GeV]	[50, 175]	[175, 275]	[275, 375]	[375, 500]	[500, 700]	[700, 1200]
$(1/\sigma) d\sigma/dx$ [GeV $^{-1}$]	0.00069	0.00364	0.0025	0.0014	0.000435	7.39e-05
Statistical uncertainty	30	9.7	12	15	22	34
MC stat. uncertainty (bootstrap)	3.2	1.3	1.9	2.3	2.5	2.6
Unfolding non-closure	-8.4	-3.4	0	0	3.3	6.2
E_e ID SF	0.33	0.13	0.087	-0.39	-0.18	-0.36
E_e resolution	-0.9	0.52	0.076	-1.1	0.64	0.41
tW ISRFSR	5.7	-0.48	-0.39	-3.7	5	-0.13
tW ME	7.7	-5.6	-4.6	6.5	21	-11
tW PS	13	-2.2	4.6	-8.5	-0.75	2.2
DR vs DS	-9.2	2.4	-4.1	5.6	3.7	-9.5
Diboson norm.	-0.25	0.27	0.13	-0.2	-0.23	-1.5
Fake norm.	4	-0.2	-0.11	-0.52	-0.49	-3.1
JER	-2	-1.6	2.9	2	-2.3	-3.1
JES: η model	0.55	0.76	-1	1.3	-3.6	0.65
JES: η stat.	4.3	-0.38	-1.3	0.48	-0.5	1.4
JES: Eff1	-0.53	0.16	-0.68	2.3	-2.4	-0.85
JES: Eff2	-2.8	0.1	1.1	-0.59	0.31	-0.023
JES: JES	1.4	0.035	-0.028	-0.056	-1.2	-0.25
JES: PU μ	-2.7	0.28	-1.1	1.5	1.1	1.2
JES: PU ρ	-0.51	0.7	-0.62	0.52	-1.2	-1.3
JES: PU NPV	2	-0.47	-0.53	0.033	0.27	2.8
JES: flav. comp.	-0.29	0.32	0.25	-1.3	0.63	0.53
JES: flav. resp.	-1	0.25	-0.24	0.88	-0.86	-0.54
Luminosity	0.5	-0.096	0.11	-0.59	0.44	0.79
MET reso. $\parallel$	-1.3	0.86	-1.9	0.1	2.9	-0.032
MET reso. $\perp$	2.1	-1.7	0.8	2.4	-2.5	0.78
MET scale	-2.8	-0.23	0.48	0.26	1.8	0.14
Z + jets norm.	-12	1.1	0.98	1.1	1.4	1.3
$t\bar{t}$ ISRFSR	-7	-6.1	2.1	12	1.8	0.29
$t\bar{t}$ ME	84	1.4	-19	10	-36	-46
$t\bar{t}$ PS	-34	-7	-6.1	3.3	50	55
$t\bar{t}$ norm.	1.5	-0.28	0.32	-1.7	1.3	2.3
b-tag SF: b0	1.8	0.75	-0.41	-0.3	-2	-2.7
b-tag SF: b1	-0.73	-0.094	-0.0052	0.47	0.11	0.16
Total syst. uncertainty	95	12	22	21	66	74
Total uncertainty	99	16	25	26	70	81

Table D.2: Normalised cross-sections differential in $E(\ell\ell b)$. Uncertainties are shown in percent.

$m_T(\ell\ell\nu\nu b)$ bin [GeV]	[50, 275]	[275, 375]	[375, 500]	[500, 1000]
$(1/\sigma) d\sigma/dx$ [GeV $^{-1}$]	0.00353	0.00121	0.000498	4.25e-05
Statistical uncertainty	6.1	22	22	28
MC stat. uncertainty (bootstrap)	0.65	2.4	3.1	2.5
Unfolding non-closure	-4.6	0	0	10
E_e ID SF	0.1	-0.38	-0.62	0.15
E_e resolution	-0.14	1	-0.59	1.3
tW ISRFSR	-0.27	1.3	-0.19	3
tW ME	-2.4	8.2	16	-4.4
tW PS	0.19	-5	6.2	3.3
DR vs DS	-0.82	2.4	8.8	-8.9
Diboson norm.	-0.019	0.33	0.35	-2.2
Fake norm.	0.11	-0.24	0.46	-4.1
JER	-0.4	3.9	-5	7.2
JES: η model	0.21	1.2	-3	-5.8
JES: η stat.	0.076	-0.32	0.92	-3.7
JES: Eff1	-0.018	0.98	0.42	-6.1
JES: Eff2	-0.011	-0.49	-0.047	3.4
JES: JES	0.3	-1.7	-0.2	-0.8
JES: PU μ	0.025	-1.8	3.3	-0.36
JES: PU ρ	0.29	0.51	-2.5	-6.4
JES: PU NPV	-0.071	-0.39	1.2	1.4
JES: flav. comp.	-0.18	0.48	2	-2.1
JES: flav. resp.	0.25	-1.3	-1.3	1.9
Luminosity	0.036	-0.6	0.012	2
MET reso. $\parallel$	-0.44	4	-1.2	-2.5
MET reso. $\perp$	-0.96	6.2	0.62	-1.1
MET scale	-0.81	4	3.1	-1.6
Z + jets norm.	-0.31	1.3	1.6	-0.39
$t\bar{t}$ ISRFSR	-1.5	7.2	-4.7	29
$t\bar{t}$ ME	-0.011	11	-24	11
$t\bar{t}$ PS	-7.7	25	28	68
$t\bar{t}$ norm.	0.11	-1.8	0.036	5.9
b-tag SF: b0	0.52	-1.3	-2.5	-4.5
b-tag SF: b1	-0.14	0.66	0.61	-0.2
Total syst. uncertainty	9.6	31	43	78
Total uncertainty	11	38	48	83

 Table D.3: Normalised cross-sections differential in $m_T(\ell\ell\nu\nu b)$. Uncertainties are shown in percent.

$m(\ell_1 b)$ bin [GeV]	[0, 60]	[60, 100]	[100, 150]	[150, 200]	[200, 250]	[250, 400]
$(1/\sigma) d\sigma/dx$ [GeV ⁻¹]	0.00105	0.0036	0.00887	0.00328	0.00163	0.000692
Statistical uncertainty	29	25	9.6	18	25	16
MC stat. uncertainty (bootstrap)	3.6	2.4	1.5	3.2	3.9	1.4
Unfolding non-closure	-11	-5	-1.7	0	0	4.7
E_e ID SF	1.1e-05	0.024	0.14	-0.1	-0.13	-0.38
E_e resolution	-1.7	0.52	-0.45	1.6	-0.95	0.52
tW ISRFSR	-0.15	4.4	-2	1.9	-4.1	2.6
tW ME	7.8	-2.9	-3.1	-2.2	9.2	8.8
tW PS	-2.2	7.9	-1.1	-4.1	-1.7	2.8
DR vs DS	-8.6	0.53	-0.57	4.3	-2.3	1.9
Diboson norm.	-0.64	-0.14	0.26	-0.27	0.7	-0.68
Fake norm.	6	0.42	-0.6	0.17	-0.76	-1.3
JER	0.22	-14	-2	15	-2	6.1
JES: η model	1.9	2.4	-0.27	-1.1	1.2	-2.5
JES: η stat.	0.91	2.2	-0.26	-0.67	-1.5	-0.19
JES: Eff1	0.69	2.9	-1.2	1.7	-1.8	-0.66
JES: Eff2	-2.6	-1.7	0.74	-0.14	0.74	0.35
JES: JES	1.8	3.2	-1.1	-2	3.2	-0.069
JES: PU μ	-0.35	-2.1	0.038	0.066	1.3	1.9
JES: PU ρ	2.4	3.3	-1.5	1.1	-1.3	-0.55
JES: PU NPV	2.4	-1.5	0.28	-0.88	-0.53	1.3
JES: flav. comp.	-1.1	-0.069	0.16	0.21	-0.43	0.14
JES: flav. resp.	-0.22	0.38	-0.099	0.46	0.13	-0.81
Luminosity	-0.094	-0.26	0.17	-0.25	-0.14	0.2
MET reso. $\parallel$	4.3	-0.76	0.12	-1.4	0.79	-0.39
MET reso. $\perp$	1.8	0.25	-0.87	1.4	3.1	-2.5
MET scale	-3.1	1.4	0.2	0.15	0.44	-1.6
Z + jets norm.	2.4	-3.5	0.25	1.1	0.53	0.11
$t\bar{t}$ ISRFSR	6.7	-18	1.5	3.5	7.9	3.2
$t\bar{t}$ ME	43	23	-7.2	-2.4	14	-34
$t\bar{t}$ PS	-25	-33	19	-23	-16	30
$t\bar{t}$ norm.	-0.27	-0.76	0.5	-0.74	-0.4	0.59
b-tag SF: b0	3	3.8	-0.026	-1.4	-2.2	-3.1
b-tag SF: b1	-0.57	-0.6	-0.11	0.43	0.47	0.6
Total syst. uncertainty	54	48	21	29	26	47
Total uncertainty	61	54	23	34	36	50

Table D.4: Normalised cross-sections differential in $m(\ell_1 b)$. Uncertainties are shown in percent.

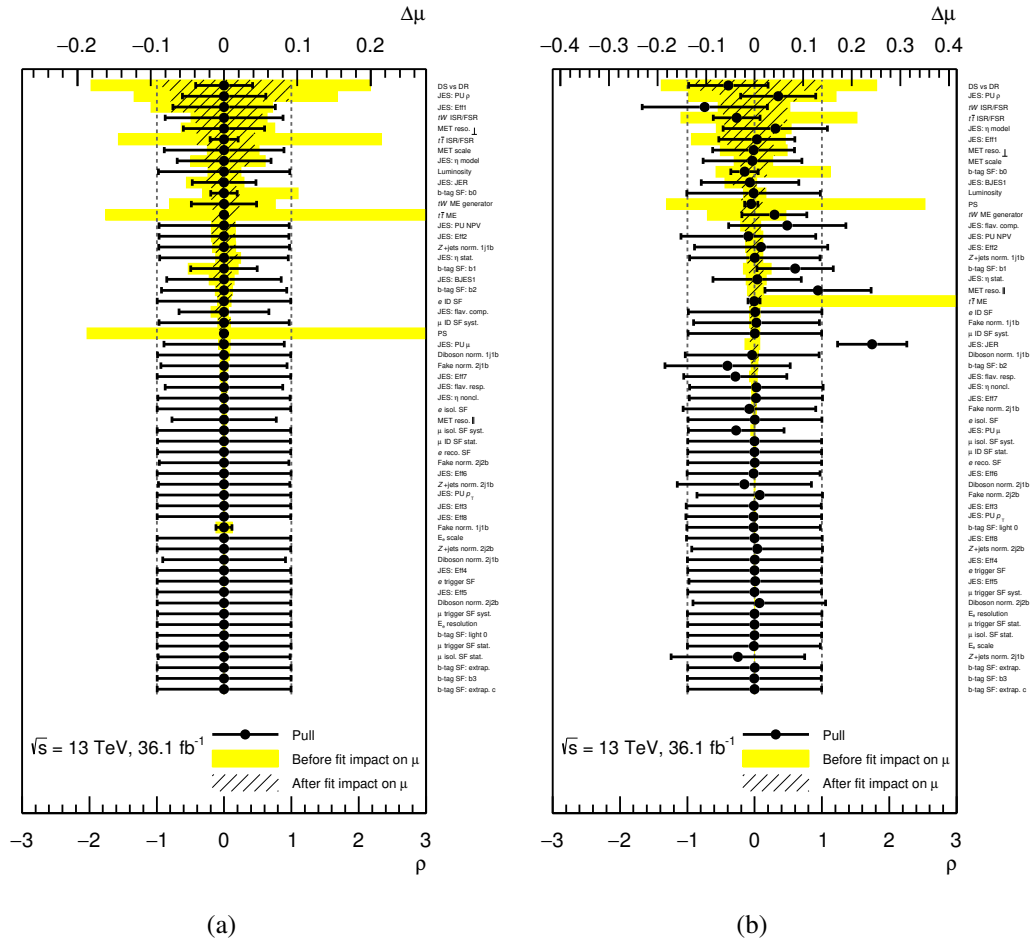
$m(\ell\ell b)$ bin [GeV]	[0, 125]	[125, 175]	[175, 225]	[225, 300]	[300, 400]	[400, 1000]
$(1/\sigma) d\sigma/dx$ [GeV $^{-1}$]	0.00112	0.0043	0.00586	0.00315	0.000486	0.000112
Statistical uncertainty	24	17	11	12	35	17
MC stat. uncertainty (bootstrap)	3.1	1.8	1.8	1.9	2.8	1.8
Unfolding non-closure	-7.2	-2.9	0	0	0	5
E_e ID SF	-0.39	0.23	0.15	0.025	-0.53	-0.32
E_e resolution	-1.9	-0.21	0.86	-0.068	0.6	0.7
tW ISRFSR	2.1	-0.77	0.093	-0.87	-2.1	2.3
tW ME	-6.1	-3	-1.7	3.1	11	11
tW PS	21	-3	-2.3	-5.3	1.1	-6
DR vs DS	0.1	-0.83	1.2	-1.6	7.1	-2
Diboson norm.	-0.5	-0.032	0.58	0.049	0.079	-1.6
Fake norm.	5.4	-0.32	-0.92	-0.86	-1.1	-2.5
JER	-26	3.8	3	4.2	8.1	7.7
JES: η model	0.55	1.3	0.082	-0.54	-1.2	-3
JES: η stat.	1.4	0.51	0.39	-1.8	0.43	-0.31
JES: Eff1	-2.6	1.4	-0.44	0.82	-1.3	0.74
JES: Eff2	-2.1	-0.1	0.42	0.73	0.55	-0.015
JES: JES	0.3	1.2	-0.32	-0.83	-0.4	0.26
JES: PU μ	0.72	-1.8	0.24	0.36	1.7	0.62
JES: PU ρ	-2.6	2.6	-0.99	0.25	0.27	0.33
JES: PU NPV	2	-0.59	-0.33	-0.6	1.1	0.56
JES: flav. comp.	-1.5	0.64	0.38	-0.37	0.31	0.43
JES: flav. resp.	1.3	-0.8	-0.0057	0.41	-1.2	-0.83
Luminosity	-0.47	0.24	0.035	-0.037	-0.63	0.64
MET reso. $\parallel$	3.2	-2.1	-2.7	3.9	-1.4	-0.66
MET reso. $\perp$	4.1	-2.1	-0.27	-0.29	5.4	-3.5
MET scale	0.23	-1	-0.73	1.9	1.4	-1.5
Z + jets norm.	2.1	-3.5	0.92	0.66	0.26	0.46
$t\bar{t}$ ISRFSR	-19	-6	9.5	-1	15	9.7
$t\bar{t}$ ME	91	-13	-15	-9.1	-26	-33
$t\bar{t}$ PS	-89	13	13	-4.6	41	75
$t\bar{t}$ norm.	-1.4	0.7	0.1	-0.11	-1.8	1.9
b-tag SF: b0	3.5	0.81	-0.05	-1.6	-1.5	-3.2
b-tag SF: b1	-0.43	-0.33	-0.026	0.33	0.73	0.41
Total syst. uncertainty	130	21	23	14	54	84
Total uncertainty	140	27	25	18	64	86

 Table D.5: Normalised cross-sections differential in $m(\ell\ell b)$. Uncertainties are shown in percent.

$m(\ell_2 b)$ bin [GeV]	[0, 50]	[50, 100]	[100, 150]	[150, 400]
$(1/\sigma) d\sigma/dx$ [GeV ⁻¹]	0.00205	0.00885	0.00654	0.000512
Statistical uncertainty	26	9.9	9.9	17
MC stat. uncertainty (bootstrap)	3.2	1.1	1.4	2.7
Unfolding non-closure	-9.5	-1.9	0	7.7
E_e ID SF	-0.51	0.00051	0.041	0.3
E_e resolution	2.1	-0.4	0.79	-2.3
tW ISRFSR	-1.4	1.7	0.66	-6.6
tW ME	-20	1.1	3	4.7
tW PS	9.1	-1.4	-2.1	3
DR vs DS	7.6	-0.17	3.1	-13
Diboson norm.	-0.24	0.7	1.3	-5.6
Fake norm.	-0.55	0.79	-0.21	-1.8
JER	0.91	-1.5	0.37	3.7
JES: η model	1.4	0.32	-0.55	-0.79
JES: η stat.	0.99	0.23	-0.47	-0.41
JES: Eff1	4.2	-0.33	-0.77	-0.25
JES: Eff2	-1.2	-0.15	0.79	-0.53
JES: JES	2.1	0.12	-0.66	-0.43
JES: PU μ	1.4	-0.82	0.82	-0.41
JES: PU ρ	4.9	0.037	-1.2	-0.89
JES: PU NPV	2.2	0.16	-0.7	-0.53
JES: flav. comp.	2.2	-0.12	-0.077	-1.1
JES: flav. resp.	-2.8	0.19	0.3	0.84
Luminosity	-1.2	-0.021	0.21	0.51
MET reso. $\parallel$	4.8	-1.4	-1.2	4
MET reso. $\perp$	6.7	-1	-1.1	1.1
MET scale	1.4	-1.3	-0.28	3.9
Z + jets norm.	2	-0.69	1.5	-3
$t\bar{t}$ ISRFSR	-9.6	-3.1	4.6	6.7
$t\bar{t}$ ME	74	-3.7	-3.7	-37
$t\bar{t}$ PS	-24	-5.1	-4.7	49
$t\bar{t}$ norm.	-3.5	-0.061	0.6	1.5
b-tag SF: b0	4.9	0.93	-1.9	-2.2
b-tag SF: b1	-0.19	-0.14	0.28	-0.08
Total syst. uncertainty	83	8.4	9.9	65
Total uncertainty	87	13	14	67

Table D.6: Normalised cross-sections differential in $m(\ell_2 b)$. Uncertainties are shown in percent.

Pull and impact of the NPs (ANN)



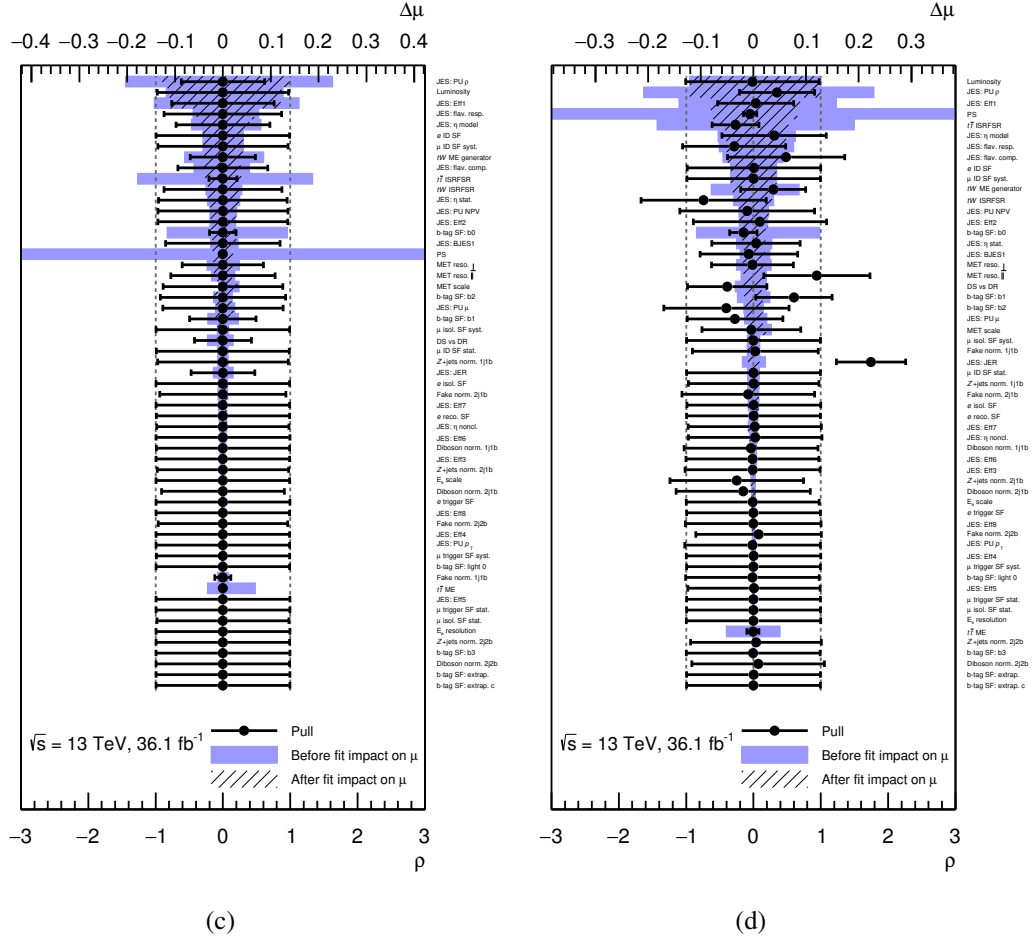


Figure E.1: Pull (corresponding to the lower x -axis) and impact on μ_{sig} (corresponding to the upper x -axis) of the NPs ordered by the post-fit impact. The ranking in asimov (a) and in data (b) are similar, but not identical. Floated version of impact for asimov (a) and data (b) and fixed version of impact for asimov (c) and data (d) are shown. The hatched lines follows the same definitions albeit presenting before fit impact.

Bibliography

- [1] T. Arabatzis, *Representing Electrons: A Biographical Approach to Theoretical Entities*, University of Chicago Press, 2006 70, ISBN: 0-226-02421-0 (cit. on p. 1).
- [2] J. J. Thomson and M. F.R.S., *XL. Cathode Rays*, *Philosophical Magazine* **44** (1897) 293, URL: <https://doi.org/10.1080/14786449708621070> (cit. on p. 1).
- [3] P. J. Nolan, *Fundamentals of College Physics*, vol. 2, Addison-Wesley, 1995 962, ISBN: 978-0697121455 (cit. on p. 1).
- [4] J. Chadwick, *Possible Existence of a Neutron*, *Nature* **129** (1932) 312 (cit. on p. 1).
- [5] P. A. M. Dirac, *The quantum theory of the electron*, *Proceedings of the Royal Society of London A* **117** (1928) 610, ISSN: 0950-1207, URL: <http://rspa.royalsocietypublishing.org/content/117/778/610> (cit. on p. 1).
- [6] C. D. Anderson, *The Positive Electron*, *Phys. Rev.* **43** (1933) 491, URL: <https://link.aps.org/doi/10.1103/PhysRev.43.491> (cit. on p. 1).
- [7] K. v. Meyenn, ed., *Das Jahr 1932 Die Entdeckung des Neutrons*, Springer, 1985 105, ISBN: 978-3-540-78801-0, URL: https://doi.org/10.1007/978-3-540-78801-0_3 (cit. on p. 1).
- [8] F. Reines and C. Cowan, *The Neutrino*, *Nature* **178** (1956) 446 EP, URL: <http://dx.doi.org/10.1038/178446a0> (cit. on p. 1).
- [9] M. Gell-Mann, *A schematic model of baryons and mesons*, *Physics Letters* **8** (1964) 214, ISSN: 0031-9163, URL: <http://www.sciencedirect.com/science/article/pii/S0031916364920013> (cit. on p. 1).
- [10] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 1*, (1964) (cit. on p. 1).
- [11] G. Zweig, “An SU(3) model for strong interaction symmetry and its breaking. Version 2”, *DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978*, ed. by D. Lichtenberg and S. P. Rosen, 1964 22, URL: <http://inspirehep.net/record/4674/files/cern-th-412.pdf> (cit. on p. 1).
- [12] R. Feynman, *QED : the strange theory of light and matter*, Princeton University Press, 2006, ISBN: 9780691125756 (cit. on p. 1).

- [13] P. A. M. Dirac, *The quantum theory of the emission and absorption of radiation*, *Proceedings of the Royal Society of London A* **114** (1927) 243, ISSN: 0950-1207, URL: <http://rspa.royalsocietypublishing.org/content/114/767/243> (cit. on p. 1).
- [14] E. Fermi, *Quantum Theory of Radiation*, *Rev. Mod. Phys.* **4** (1932) 87, URL: <https://link.aps.org/doi/10.1103/RevModPhys.4.87> (cit. on p. 1).
- [15] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264> (cit. on p. 1).
- [16] S. L. Glashow, *The renormalizability of vector meson interactions*, *Nuclear Physics* **10** (1959) 107, ISSN: 0029-5582, URL: <http://www.sciencedirect.com/science/article/pii/0029558259901968> (cit. on p. 1).
- [17] A. Salam and J. C. Ward, *Weak and electromagnetic interactions*, *Il Nuovo Cimento* (1955-1965) **11** (1959) 568, ISSN: 1827-6121, URL: <https://doi.org/10.1007/BF02726525> (cit. on p. 1).
- [18] R. Oerter, *The theory of almost everything: the Standard Model, the unsung triumph of modern physics*, Pi Press, 2006, ISBN: 0132366789 (cit. on p. 2).
- [19] A. Steere, *A Timeline of Major Particle Accelerators*, Michigan State University. Department of Physics and Astronomy, 2005, URL: <https://books.google.de/books?id=YRNonQEACAAJ> (cit. on p. 2).
- [20] P. Bryant, *A brief history and review of accelerators*, (1994) (cit. on p. 2).
- [21] J. D. Cockcroft and E. T. Walton, *Experiments with high velocity positive ions*, *Proc. R. Soc. Lond. A* **129** (1930) 477 (cit. on p. 2).
- [22] E. O. Lawrence and M. S. Livingston, *The production of high speed light ions without the use of high voltages*, *Phys. Rev.* **40** (1932) 19 (cit. on p. 2).
- [23] R. Wideröe, *Some memories and dreams from the childhood of particle accelerators*, *Europhysics News* **15** (1984) 9 (cit. on p. 2).
- [24] V. I. Veksler, *A new method of acceleration of relativistic particles*, *J. Phys.* **9** (1945) 153, URL: <http://cds.cern.ch/record/109364> (cit. on p. 2).
- [25] E. M. McMillan, *The Synchrotron—A Proposed High Energy Particle Accelerator*, *Phys. Rev.* **68** (1945) 143, URL: <https://link.aps.org/doi/10.1103/PhysRev.68.143> (cit. on p. 2).
- [26] E. D. Courant, M. S. Livingston and H. S. Snyder, *The strong-focusing synchrotron—A new high energy accelerator*, *Phys. Rev.* **88** (1952) 1190 (cit. on p. 2).
- [27] S. Van der Meer, *Stochastic damping of betatron oscillations*, *The Development of Colliders* (1995) 261 (cit. on p. 2).

-
- [28] O. S. Brüning, P. Collier et al., *LHC Design Report*, CERN Yellow Reports: Monographs, CERN, 2004, URL: <https://cds.cern.ch/record/782076> (cit. on p. 2).
- [29] J. J. Aubert, U. Becker et al., *Experimental Observation of a Heavy Particle J*, *Phys. Rev. Lett.* **33** (1974) 1404, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.33.1404> (cit. on p. 2).
- [30] J. E. Augustin, A. M. Boyarski et al., *Discovery of a Narrow Resonance in e^+e^- Annihilation*, *Phys. Rev. Lett.* **33** (1974) 1406, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.33.1406> (cit. on p. 2).
- [31] M. L. Perl, G. S. Abrams et al., *Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation*, *Phys. Rev. Lett.* **35** (1975) 1489, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.35.1489> (cit. on pp. 2, 18).
- [32] J. Street, *Cloud chamber studies of cosmic ray showers and penetrating particles*, *Journal of the Franklin Institute* **227** (1939) 765, ISSN: 0016-0032, URL: <http://www.sciencedirect.com/science/article/pii/S0016003239908173> (cit. on p. 2).
- [33] S. W. Herb, D. C. Hom et al., *Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977) 252, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.39.252> (cit. on pp. 2, 18).
- [34] B. R. Stella and H.-J. Meyer, *$Y(9.46\text{ GeV})$ and the gluon discovery (a critical recollection of PLUTO results)*, *Eur. Phys. J. H* **36** (2011) 203, arXiv: 1008.1869 [hep-ex], URL: <https://doi.org/10.1140/epjh/e2011-10029-3> (cit. on p. 2).
- [35] UA1 Collaboration, G. Arnison et al., *Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540\text{ GeV}$* , *Physics Letters B* **122** (1983) 103, ISSN: 0370-2693, URL: <http://www.sciencedirect.com/science/article/pii/0370269383911772> (cit. on p. 2).
- [36] UA2 Collaboration, M. Banner et al., *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider*, *Physics Letters B* **122** (1983) 476, ISSN: 0370-2693, URL: <http://www.sciencedirect.com/science/article/pii/0370269383916052> (cit. on p. 2).
- [37] UA1 Collaboration, G. Arnison et al., *Experimental observation of lepton pairs of invariant mass around $95\text{ GeV}/c^2$ at the CERN SPS collider*, *Physics Letters B* **126** (1983) 398, ISSN: 0370-2693, URL: <http://www.sciencedirect.com/science/article/pii/0370269383901880> (cit. on p. 2).

- [38] UA2 Collaboration, M. Banner et al., *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider*, *Physics Letters B* **122** (1983) 476, ISSN: 0370-2693, URL: <http://www.sciencedirect.com/science/article/pii/0370269383916052> (cit. on p. 2).
- [39] CDF Collaboration, F. Abe et al., *Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab*, *Phys. Rev. Lett.* **74** (1995) 2626, arXiv: 9503002 [hep-ex], URL: <https://link.aps.org/doi/10.1103/PhysRevLett.74.2626> (cit. on pp. 2, 18).
- [40] D0 Collaboration, S. Abachi et al., *Observation of the Top Quark*, *Phys. Rev. Lett.* **74** (1995) 2632, arXiv: 9503003 [hep-ex], URL: <https://link.aps.org/doi/10.1103/PhysRevLett.74.2632> (cit. on pp. 2, 18).
- [41] M. L. Perl, G. S. Abrams et al., *Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation*, *Phys. Rev. Lett.* **35** (1975) 1489, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.35.1489> (cit. on p. 2).
- [42] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001 (cit. on pp. 2, 27).
- [43] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on pp. 2, 9).
- [44] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (cit. on pp. 2, 9).
- [45] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on pp. 2, 11).
- [46] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on pp. 2, 11).
- [47] ATLAS Collaboration, *A Particle Consistent with the Higgs Boson Observed with the ATLAS Detector at the Large Hadron Collider*, *Science* **338** (2012) 1576 (cit. on p. 2).
- [48] CMS Collaboration, *Updated measurements of the Higgs boson at 125 GeV in the two photon decay channel*, (2013), URL: <http://cds.cern.ch/record/1530524> (cit. on pp. 2, 11).
- [49] *NNLO+NNLL top-quark-pair cross sections*, URL: https://en.wikipedia.org/wiki/Standard_Model (cit. on p. 8).
- [50] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, *Phys. Lett. B* **726** (2013) 120, arXiv: 1307.1432 [hep-ex] (cit. on p. 11).
- [51] N. Cabibbo, *Unitary symmetry and leptonic decays*, *Phys. Rev. Lett.* **10** (1963) 531 (cit. on p. 11).

-
- [52] M. Kobayashi et al., *M. Kobayashi and T. Maskawa*, *Prog. Theor. Phys.* **49**, 652 (1973)., *Prog. Theor. Phys.* **49** (1973) 652 (cit. on p. 11).
- [53] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, *Phys. Rev. Lett.* **51** (1983) 1945, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.51.1945> (cit. on p. 12).
- [54] C. Patrignani et al., *Review of Particle Physics*, *Chin. Phys. C* **40** (2016) 100001 (cit. on pp. 12, 20).
- [55] Belle Collaboration, *Observation of a Narrow Charmoniumlike State in Exclusive $B^{\pm} \rightarrow K^{\pm} \pi^{+} \pi^{-} J/\psi$ Decays*, *Phys. Rev. Lett.* **91** (2003) 262001, arXiv: 1112.5675 [hep-ex] (cit. on p. 13).
- [56] BESIII Collaboration, *Observation of a Charged Charmoniumlike Structure in $e^{+} e^{-} \rightarrow \pi^{+} \pi^{-} J/\psi$ at $\sqrt{s}=4.26$ GeV*, *Phys. Rev. Lett.* **110** (2013) 252001, arXiv: 1303.5949 [hep-ex] (cit. on p. 13).
- [57] LHCb Collaboration, *Observation of the Resonant Character of the $Z(4430)^{-}$ State*, *Phys. Rev. Lett.* **112** (2014) 222002, arXiv: 1404.1903 [hep-ex] (cit. on p. 13).
- [58] I. Bigi, *CP violation - A probe of Nature's Grand Design*, *Surveys in High Energy Physics* **12** (1998) 269, arXiv: 9712475 [hep-ph] (cit. on p. 14).
- [59] M. C. Gonzalez-Garcia and M. Maltoni, *Phenomenology with massive neutrinos*, *Physics Reports* **460** (2008) 1, arXiv: 0704.1800 [hep-ph] (cit. on p. 14).
- [60] *PLANCK Publications*, URL: <https://www.cosmos.esa.int/web/planck/publications> (cit. on p. 14).
- [61] J. D. Lykken, *Beyond the Standard Model*, (2010), arXiv: 1005.1676 [hep-ph], URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-10-103 (cit. on p. 14).
- [62] S. van der Meer, *Calibration of the effective beam height in the ISR*, tech. rep. CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, 1968, URL: <http://cds.cern.ch/record/296752> (cit. on p. 16).
- [63] P. D. Group, *Review of particle physics*, *Chin. Phys. C* **40** (2016) 100001 (cit. on p. 18).
- [64] J. Espinosa, G. F. Giudice and A. Riotto, *Cosmological implications of the Higgs mass measurement*, *Journal of Cosmology and Astroparticle Physics* **2008** (2008) 002 (cit. on p. 19).
- [65] H. E. Haber and G. L. Kane, *The search for supersymmetry: probing physics beyond the standard model*, *Physics Reports* **117** (1985) 75 (cit. on p. 19).
- [66] R. Yamada, *Tevatron*, (1978) (cit. on p. 19).
- [67] M. Czakon, P. Fiedler and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$* , *Phys. Rev. Lett.* **110** (2013) 252004, arXiv: 1303.6254 [hep-ph], URL: <https://link.aps.org/doi/10.1103/PhysRevLett.110.252004> (cit. on p. 20).

- [68] *NNLO+NNLL top-quark-pair cross sections*,
URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>
(cit. on p. 20).
- [69] M. Czakon and A. Mitov,
Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders,
Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph] (cit. on p. 20).
- [70] *LHCTopWG Summary Plots*, URL:
<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots>
(cit. on p. 21).
- [71] D0 Collaboration, V. M. Abazov et al., *Observation of Single Top-Quark Production*,
Phys. Rev. Lett. **103** (2009) 092001, arXiv: 0903.0850 [hep-ex] (cit. on p. 20).
- [72] CDF Collaboration, T. Aaltonen et al.,
Observation of Electroweak Single Top-Quark Production,
Phys. Rev. Lett. **103** (2009) 092002, arXiv: 0903.0885 [hep-ex] (cit. on p. 20).
- [73] *NLO single-top channel cross sections*,
URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>
(cit. on p. 22).
- [74] N. Kidonakis, *Theoretical results for electroweak-boson and single-top production*,
PoS DIS (2015) 170, arXiv: 1506.04072 [hep-ph] (cit. on p. 22).
- [75] S. Frixione et al., *Single-top hadroproduction in association with a W boson*,
JHEP **07** (2008) 029, arXiv: 0805.3067 [hep-ph] (cit. on p. 24).
- [76] T. M. P. Tait, *The tW^- mode of single top production*, *Phys. Rev. D* **61** (1999) 034001,
arXiv: hep-ph/9909352 [hep-ph] (cit. on p. 25).
- [77] J. M. Campbell and F. Tramontano,
Next-to-leading order corrections to Wt production and decay, *Nucl. Phys. B* **726** (2005) 109,
arXiv: hep-ph/0506289 [hep-ph] (cit. on p. 25).
- [78] C. D. White et al., *Isolating Wt production at the LHC*, *JHEP* **11** (2009) 074,
arXiv: 0908.0631 [hep-ph] (cit. on p. 25).
- [79] R. Frederix, *Top Quark Induced Backgrounds to Higgs Production in the $WW^{(*)} \rightarrow ll\nu\nu$ Decay Channel at Next-to-Leading-Order in QCD*, *Phys. Rev. Lett.* **112** (2014) 082002,
arXiv: 1311.4893 [hep-ph] (cit. on p. 25).
- [80] F. Demartin et al., *tWH associated production at the LHC*, *Eur. Phys. J. C* **77** (2017) 34,
arXiv: 1607.05862 [hep-ph] (cit. on p. 25).
- [81] A. Denner et al.,
Next-to-Leading-Order QCD Corrections to $W^+W^-b\bar{b}$ Production at Hadron Colliders,
Phys. Rev. Lett. **106** (2011) 052001, arXiv: 1711.10359 [hep-ph] (cit. on p. 25).
- [82] B. Richter,
Very high-energy electron-positron colliding beams for the study of the weak interactions,
Nucl. Instrum. Meth. **136** (1976) 47 (cit. on p. 27).
- [83] *CERN Accelerating science*, URL: <https://home.cern> (cit. on p. 27).

-
- [84] *LHC Commissioning with Lead Ion Beams*, URL: <https://lhc-commissioning.web.cern.ch/lhc-commissioning/commissioning-ions.htm> (cit. on p. 27).
- [85] J. Jowett, R. Alemany-Fernandez et al., *The 2015 Heavy-Ion Run of the LHC*, (2016) TUPMW027. 4 p, URL: <http://cds.cern.ch/record/2207379> (cit. on p. 27).
- [86] *Longer term LHC schedule*, URL: <https://lhc-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-long-term.htm> (cit. on pp. 27, 29).
- [87] *For one day only LHC collides xenon beams*, URL: <https://home.cern/about/updates/2017/10/one-day-only-lhc-collides-xenon-beams> (cit. on p. 27).
- [88] J. Haffner, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, (2013), General Photo, URL: <https://cds.cern.ch/record/1621894> (cit. on p. 28).
- [89] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003 (cit. on pp. 28, 29).
- [90] CMS Collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004 (cit. on p. 28).
- [91] LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005 (cit. on p. 28).
- [92] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002 (cit. on p. 28).
- [93] TOTEM Collaboration, *The TOTEM experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08007 (cit. on p. 28).
- [94] LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08006 (cit. on p. 28).
- [95] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*, (2009) (cit. on p. 29).
- [96] *LHC sets new world record*, URL: <http://press.cern/press-releases/2009/11/lhc-sets-new-world-record> (cit. on p. 29).
- [97] *CERN achieves 7 TeV collisions at Large Hadron Collider*, URL: <https://physicsworld.com/a/cern-achieves-7-tev-collisions-at-large-hadron-collider/> (cit. on p. 29).
- [98] M. Lamont, *Status of the LHC*, *J. Phys. Conf. Ser.* **455** (2013) 012001 (cit. on p. 29).
- [99] *CERN's Large Hadron Collider gears up for run 2*, URL: <https://home.cern/about/updates/2014/12/cerns-large-hadron-collider-gears-run-2> (cit. on p. 29).
- [100] *LHC Season 2: First physics at 13 TeV to start tomorrow*, URL: <https://home.cern/about/updates/2015/06/lhc-season-2-first-physics-13-tev-start-tomorrow> (cit. on p. 29).

- [101] *LHC performance reaches new highs*, URL: <https://home.cern/about/updates/2016/07/lhc-performance-reaches-new-highs> (cit. on p. 29).
- [102] *Record luminosity: well done LHC*, URL: <https://home.cern/about/updates/2017/11/record-luminosity-well-done-lhc> (cit. on p. 29).
- [103] *LHC Report: Another run is over and LS2 has just begun. . .*
URL: https://home.cern/news/news/accelerators/lhc-report-another-run-over-and-ls2-has-just-begun?utm_source=Bulletin&utm_medium=Email&utm_content=2018-12-12E&utm_campaign=BulletinEmail (cit. on p. 29).
- [104] *Luminosity Public Results Run 2*, URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2> (cit. on pp. 29, 51).
- [105] ATLAS Collaboration,
Expected Performance of the ATLAS Experiment – Detector, Trigger and Physics, 2009,
arXiv: 0901.0512 (cit. on pp. 29, 32, 33).
- [106] *ATLAS Photos*,
URL: <http://atlasexperiment.org/photos/full-detector-cgi.html>
(cit. on p. 30).
- [107] *ATLAS Photos*,
URL: <http://atlasexperiment.org/photos/inner-detector-combined.html>
(cit. on p. 31).
- [108] G. Ripellino, *The alignment of the ATLAS Inner Detector in Run-2*, PoS LHCP (2016) 196
(cit. on p. 31).
- [109] A. Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*,
tech. rep. CERN-LHCC-2017-021. ATLAS-TDR-030, CERN, 2017,
URL: <https://cds.cern.ch/record/2285585> (cit. on pp. 32, 33).
- [110] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, (2010)
(cit. on p. 32).
- [111] ATLAS Collaboration,
Study of the material of the ATLAS inner detector for Run 2 of the LHC,
JINST 12 (2017) P12009, arXiv: 1707.02826 [hep-ex] (cit. on p. 32).
- [112] *ATLAS Photos*, URL:
<http://atlasexperiment.org/photos/calorimeters-combined-barrel.html>
(cit. on p. 34).
- [113] ATLAS Collaboration, *ATLAS calorimeter performance Technical Design Report*, (1996)
(cit. on p. 35).
- [114] ATLAS Collaboration, *ATLAS muon spectrometer: Technical design report*, (1997)
(cit. on pp. 34, 36).

-
- [115] *Magnet System*, URL: <https://atlas.cern/discover/detector/magnet-system> (cit. on p. 37).
- [116] ATLAS Collaboration, *ATLAS central solenoid: Technical design report*, (1997) (cit. on p. 37).
- [117] ATLAS Collaboration, *ATLAS barrel toroid: Technical design report*, (1997) (cit. on p. 37).
- [118] ATLAS Collaboration, *ATLAS endcap toroids: Technical design report*, (1997) (cit. on p. 37).
- [119] A. Yamamoto et al., *The ATLAS central solenoid*, *Nucl. Instrum. Meth. A* **584** (2008) 53, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900207020414> (cit. on p. 37).
- [120] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2010*, *Eur. Phys. J. C* **72** (2012) 1849, arXiv: 1110.1530 [hep-ex] (cit. on p. 38).
- [121] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, *Eur. Phys. J. C* **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on pp. 38, 57).
- [122] *ATLAS Photos*, URL: <http://atlasexperiment.org/photos/how-atlas-works.html> (cit. on p. 39).
- [123] T. Cornelissen et al., *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, tech. rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, 2007, URL: <https://cds.cern.ch/record/1020106> (cit. on p. 39).
- [124] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on pp. 39, 40).
- [125] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) 332, arXiv: 1611.10235 [hep-ex] (cit. on p. 40).
- [126] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011, 2017, URL: <https://cds.cern.ch/record/2270366> (cit. on p. 40).
- [127] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, URL: <https://cds.cern.ch/record/1099735> (cit. on p. 40).
- [128] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 40).
- [129] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2016-015, 2016, URL: <https://cds.cern.ch/record/2203514> (cit. on pp. 41, 42).

- [130] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, URL: <https://cds.cern.ch/record/2157687> (cit. on p. 42).
- [131] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015-2016 LHC proton-proton collision data*, (2018), arXiv: [1812.03848 \[hep-ex\]](#) (cit. on p. 42).
- [132] ATLAS Collaboration, *Muon reconstruction performance in early $\sqrt{s} = 13$ TeV data*, ATL-PHYS-PUB-2015-037, 2015, URL: <https://cds.cern.ch/record/2047831> (cit. on pp. 42, 43).
- [133] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, *Eur. Phys. J. C* **74** (2014) 3130, arXiv: [1407.3935 \[hep-ex\]](#) (cit. on p. 43).
- [134] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **76** (2016) 292, arXiv: [1603.05598 \[hep-ex\]](#) (cit. on p. 44).
- [135] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, arXiv: [0802.1189 \[hep-ph\]](#) (cit. on p. 44).
- [136] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083, 2013, URL: <https://cds.cern.ch/record/1570994> (cit. on p. 45).
- [137] ATLAS Collaboration, *Tagging and suppression of pileup jets*, ATL-PHYS-PUB-2014-001, 2014, URL: <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-001/> (cit. on pp. 45, 46).
- [138] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **96** (2017) 072002, arXiv: [1703.09665 \[hep-ex\]](#) (cit. on pp. 45–47).
- [139] ATLAS Collaboration, *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2015-017, 2015, URL: <https://cds.cern.ch/record/2008678> (cit. on p. 46).
- [140] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015, URL: <https://cds.cern.ch/record/2037697> (cit. on p. 47).
- [141] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: <https://cds.cern.ch/record/2160731> (cit. on pp. 47, 48).
- [142] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*, (2018), arXiv: [1802.08168 \[hep-ex\]](#) (cit. on p. 48).

-
- [143] S. Agostinelli et al., *GEANT4: a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 (cit. on p. 52).
- [144] *ATLAS Detector Simulation with Geant4*,
URL: <http://atlas-computing.web.cern.ch/atlas-computing/packages/simulation/geant4/geant4.html> (cit. on p. 52).
- [145] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, 2010,
URL: <https://cds.cern.ch/record/1300517> (cit. on p. 52).
- [146] K. Edmonds et al., *The fast ATLAS track simulation (FATRAS)*, (2008) (cit. on p. 52).
- [147] ATLAS Collaboration, *Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles*, ATL-SOFT-PUB-2014-01, 2014, URL: <https://cds.cern.ch/record/1669341> (cit. on p. 52).
- [148] ATLAS Collaboration,
Simulation of top-quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV, ATL-PHYS-PUB-2016-004, 2016, URL: <https://cds.cern.ch/record/2120417> (cit. on pp. 52, 53).
- [149] E. Re,
Single-top Wt -channel production matched with parton showers using the POWHEG method, *Eur. Phys. J. C* **71** (2011) 1547, arXiv: 1009.2450 [hep-ph] (cit. on p. 52).
- [150] H.-L. Lai et al., *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024, arXiv: 1007.2241 [hep-ph] (cit. on pp. 52, 53).
- [151] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026, arXiv: hep-ph/0603175 (cit. on p. 52).
- [152] J. Pumplin et al.,
New generation of parton distributions with uncertainties from global QCD analysis, *JHEP* **07** (2002) 012, arXiv: hep-ph/0201195 (cit. on p. 52).
- [153] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, *Phys. Rev. D* **82** (2010) 074018, arXiv: 1005.3457 [hep-ph] (cit. on p. 52).
- [154] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152 (cit. on p. 52).
- [155] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 52).
- [156] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012, URL: <https://cds.cern.ch/record/1474107> (cit. on p. 52).
- [157] A. D. Martin et al., *Parton distributions for the LHC*, *Eur. Phys. J. C* **63** (2009) 189, arXiv: 0901.0002 [hep-ph] (cit. on p. 52).
- [158] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on p. 53).

- [159] G. Marchesini et al., *HERWIG: A Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. Version 5.1 - April 1991*, *Comput. Phys. Commun.* **67** (1992) 465 (cit. on p. 53).
- [160] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419> (cit. on p. 53).
- [161] J. M. Campbell et al., *Top-pair production and decay at NLO matched with parton showers*, *JHEP* **04** (2015) 114, arXiv: 1412.1828 [hep-ph] (cit. on p. 53).
- [162] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) 2930, arXiv: 1112.5675 [hep-ph] (cit. on p. 53).
- [163] M. Botje et al., *The PDF4LHC Working Group Interim Recommendations*, (2011), arXiv: 1101.0538 [hep-ph] (cit. on p. 53).
- [164] A. D. Martin et al., *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, *Eur. Phys. J. C* **64** (2009) 653, arXiv: 0905.3531 [hep-ph] (cit. on p. 53).
- [165] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, *Phys. Rev. D* **89** (2014) 033009, arXiv: 1302.6246 [hep-ph] (cit. on p. 53).
- [166] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on p. 53).
- [167] *MC15SingleTopSamplesPMG*, URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MC15SingleTopSamplesPMG> (cit. on p. 55).
- [168] *MC15TTbarSamplesPMG*, URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MC15TTbarSamplesPMG> (cit. on p. 55).
- [169] ATLAS Collaboration, *Monte Carlo Generators for the Production of a W or Z/ γ^* Boson in Association with Jets at ATLAS in Run 2*, ATL-PHYS-PUB-2016-003, 2016, URL: <https://cds.cern.ch/record/2120133> (cit. on p. 54).
- [170] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 54).
- [171] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 54).
- [172] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, arXiv: 1111.5206 [hep-ph] (cit. on p. 54).
- [173] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038, arXiv: 0709.1027 [hep-ph] (cit. on p. 54).
- [174] S. Höche et al., *QCD matrix elements + parton showers: The NLO case*, *JHEP* **04** (2013) 027, arXiv: 1207.5030 [hep-ph] (cit. on p. 54).
- [175] R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on p. 54).

-
- [176] R. Gavin et al., *FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order*, *Comput. Phys. Commun.* **182** (2011) 2388, arXiv: [1011.3540 \[hep-ph\]](#) (cit. on p. 54).
- [177] *MC15ZjetsSherpa221LightSamplesPMG*, URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MC15ZjetsSherpa221LightSamplesPMG> (cit. on p. 56).
- [178] ATLAS Collaboration, *Multi-boson simulation for 13 TeV ATLAS analyses*, ATL-PHYS-PUB-2016-002, 2016, URL: <https://cds.cern.ch/record/2119986> (cit. on p. 54).
- [179] *MC15DibosonSherpa*, URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MC15DibosonSherpa> (cit. on p. 56).
- [180] ATLAS Collaboration, *Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s} = 13$ TeV with ATLAS*, *JHEP* **01** (2018) 063, arXiv: [1612.07231 \[hep-ex\]](#) (cit. on pp. 57, 127).
- [181] J. H. Friedman, *Stochastic gradient boosting*, *Comput. Stat. & Data Analysis* **38** (2002) 367 (cit. on p. 65).
- [182] R. Schapire, *The strength of weak learnability*, *Machine Learning* **5** (1990) 197 (cit. on pp. 65, 67).
- [183] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, *Journal of Computer and System Sciences* **55** (1997) 119, ISSN: 0022-0000, URL: <http://www.sciencedirect.com/science/article/pii/S002200009791504X> (cit. on p. 67).
- [184] J. H. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*, *Annals of Statistics* **29** (2000) 1189 (cit. on p. 67).
- [185] P. J. Huber, *Robust estimation of a location parameter*, *Ann. Math. Statist.* **35** (1964) 73, ISSN: 0003-4851, URL: <http://dx.doi.org/10.1214/aoms/1177703732> (cit. on p. 68).
- [186] L. Rosasco et al., *Are loss functions all the same?*, *Neural Computation* **16** (2004) 1063 (cit. on p. 68).
- [187] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, PoS ACAT (2007) 040, arXiv: [physics/0703039](#) (cit. on pp. 69, 70).
- [188] A. Kolmogorov, *19++*. *Sulla determinazione empirica di una legge di distribuzione*, *Giornale dell'Istituto Italiano degli Attuari* **4** (1933) 91 (cit. on p. 70).
- [189] N. Smirnov, *Sur les écarts de la courbe de distribution empirique*, *Matematicheskii Sbornik* **48** (1939) 3 (cit. on p. 70).
- [190] G. Cowan, *Statistical Data Analysis*, Oxford science publications, Clarendon Press, 1998, ISBN: 9780198501558, URL: <https://books.google.de/books?id=ff8ZyW0nIJAC> (cit. on p. 81).
- [191] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, 2012, URL: <https://cds.cern.ch/record/1456844> (cit. on p. 82).

- [192] M. Baak et al., *HistFitter software framework for statistical data analysis*, *Eur. Phys. J. C* **75** (2015) 153, arXiv: 1410.1280 [hep-ex] (cit. on p. 82).
- [193] J. Neyman and E. S. Pearson, *IX. On the problem of the most efficient tests of statistical hypotheses*, *Phil. Trans. R. Soc. Lond. A* **231** (1933) 289 (cit. on p. 82).
- [194] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *Ann. Math. Statist.* **9** (1938) 60 (cit. on p. 83).
- [195] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, *Transactions of the American Mathematical society* **54** (1943) 426 (cit. on p. 83).
- [196] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 (cit. on pp. 83, 84).
- [197] L. Demortier and L. Lyons, *Everything you always wanted to know about pulls*, *CDF note* **43** (2002) (cit. on p. 83).
- [198] E. Gross, *Practical Statistics for High Energy Physics*, *CERN Yellow Reports: School Proceedings* **4** (2017) 165, ISSN: 2519-805X, URL: <https://e-publishing.cern.ch/index.php/CYRSP/article/view/303> (cit. on p. 83).
- [199] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys.* **G43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 85).
- [200] *Uncertainty release for analyses using 2015 and 2016 data in 20.7 (MC15c) releases*, URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/JetUncertainties20152016Data20p7> (cit. on p. 86).
- [201] *MET Systematics*, URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MissingETSystematics> (cit. on p. 86).
- [202] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, *Eur. Phys. J. C* **76** (2016) 653, arXiv: 1608.03953 [hep-ex] (cit. on p. 87).
- [203] *JES uncertainties for 2018*, URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetUncertaintiesRel21Moriond2018SmallR> (cit. on p. 92).
- [204] G. Cowan, *A survey of unfolding methods for particle physics*, *Conf. Proc. C* **0203181** (2002) 248 (cit. on p. 98).
- [205] *SLUO Lectures on Statistics and Numerical Methods in HEP Lecture 9: Unfolding*, URL: http://www-group.slac.stanford.edu/sluo/Lectures/Stat_Lectures.html (cit. on p. 98).
- [206] G. D’Agostini, *A Multidimensional unfolding method based on Bayes’ theorem*, *Nucl. Instrum. Meth. A* **362** (1995) 487 (cit. on p. 99).

-
- [207] T. Adye, *Unfolding algorithms and tests using RooUnfold*, 2011, arXiv: [1105.1160 \[physics.data-an\]](#) (cit. on p. [100](#)).
- [208] ATLAS Collaboration, *Proposal for particle-level object and observable definitions for use in physics measurements at the LHC*, ATL-PHYS-PUB-2015-013, 2015, URL: <https://cds.cern.ch/record/2022743> (cit. on p. [101](#)).
- [209] ATLAS Collaboration, *Measurement of differential cross-sections of a single top quark produced in association with a W boson at $\sqrt{s} = 13$ TeV with ATLAS*, *Eur. Phys. J. C* **78** (2018) 186, arXiv: [1712.01602 \[hep-ex\]](#) (cit. on pp. [107](#), [127](#)).
- [210] K. Albertsson et al., *Machine Learning in High Energy Physics Community White Paper*, *J. Phys. Conf. Ser.* **1085** (2018) 022008, arXiv: [1807.02876 \[physics.comp-ph\]](#) (cit. on p. [113](#)).
- [211] D. Guest, K. Cranmer and D. Whiteson, *Deep Learning and its Application to LHC Physics*, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161, arXiv: [1806.11484 \[hep-ex\]](#) (cit. on p. [113](#)).
- [212] *Activation function*, URL: https://en.wikipedia.org/wiki/Activation_function (cit. on p. [116](#)).
- [213] I. Goodfellow et al., *Generative adversarial nets*, (2014) 2672 (cit. on p. [118](#)).
- [214] G. Louppe, M. Kagan and K. Cranmer, *Learning to pivot with adversarial networks*, (2017) 981 (cit. on p. [118](#)).
- [215] F. Chollet et al., *Keras: The Python Deep Learning library*, 2015, URL: <https://keras.io> (cit. on p. [119](#)).
- [216] Theano Development Team, *Theano: A Python framework for fast computation of mathematical expressions*, (2016), arXiv: [1605.02688 \[cs.SC\]](#), URL: <http://arxiv.org/abs/1605.02688> (cit. on p. [119](#)).
- [217] CMS Collaboration, *Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **10** (2018) 117, arXiv: [1805.07399 \[hep-ex\]](#) (cit. on p. [128](#)).
- [218] *MC15WjetsSherpa221LightSamplesPMG*, URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MC15WjetsSherpa221LightSamplesPMG> (cit. on p. [131](#)).
- [219] *MC15ZjetsOtherSamplesPMG*, URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MC15ZjetsOtherSamplesPMG> (cit. on p. [132](#)).

Acronyms

ANN adversarial neural networks. [118](#), [119](#), [121](#), [122](#), [124](#), [125](#), [127](#), [128](#)
ATLAS A Toroidal LHC ApparatuS. [28–34](#), [36–39](#), [43](#), [44](#), [48](#), [51](#), [52](#), [84](#), [85](#), [87](#), [127](#), [169](#)
AUC area under the ROC curve. [68](#), [74](#), [76](#)
BDT boosted decision tree. [65](#), [68–75](#), [77](#), [84](#), [85](#), [87–89](#), [95](#), [102–104](#), [113](#), [119](#), [121](#), [122](#), [124](#), [125](#), [127](#), [134–137](#)
CKM Cabibbo–Kobayashi–Maskawa. [11](#), [12](#), [14](#), [23](#)
CSC resistive plate chamber. [34](#), [36](#), [37](#)
CTP central trigger processors. [38](#)
DR diagram removal. [24](#), [25](#), [52](#), [53](#), [55](#), [85](#), [92](#), [107](#), [119](#), [121](#), [122](#), [169](#)
DS diagram subtraction. [24](#), [25](#), [53](#), [55](#), [69](#), [85](#), [92](#), [107](#), [119](#), [121](#), [122](#), [169](#)
EMC electromagnetic calorimeter. [33–35](#), [38](#), [41](#), [42](#), [46](#), [57](#)
EWSB electroweak symmetry breaking. [10](#), [18](#), [169](#)
FCal forward calorimeter. [33](#), [34](#)
GSC global sequential calibration. [46](#)
HCal hadron calorimeter. [33–35](#), [38](#)
HLT high-level trigger. [38](#), [171](#)
IBL Insertable B-Layer. [32](#), [41](#)
ID inner detector. [31](#), [32](#), [37–39](#), [41–43](#), [169](#)
IP interaction point. [30](#)
ISR/FSR initial- and final-state radiation. [52–55](#), [90](#), [92](#), [107](#), [122](#)
JER jet energy resolution. [46](#), [86](#), [92](#)
JES jet energy scale. [45–47](#), [86](#), [90](#), [92](#), [127](#)
JVF jet-vertex-fraction. [45](#), [46](#), [169](#)
JVT jet-vertex-tagger. [45](#), [46](#), [57](#), [169](#)
KL Kullback-Leibler. [114](#), [169](#)
KS Kolmogorov–Smirnov. [70](#), [74](#), [76](#), [169](#)
L1 level-1. [38](#), [167](#), [168](#), [171](#)
L1Calo L1 calorimeter trigger system. [38](#)
L1Muon L1 muon trigger system. [38](#)

- L1Topo** L1 topological trigger modules. 38
- LAr** liquid-argon. 33, 35, 37
- LHC** Large Hadron Collider. 2, 3, 16, 19–23, 27–30, 38, 51, 127, 128, 169
- LO** leading order. 6, 19–22, 24, 54, 56, 57, 131, 132
- MC** Monte Carlo. 44, 46, 52, 58–63, 69, 71, 83, 85, 86, 97, 98, 100, 101, 103, 105–107, 109–112, 127, 169
- MDT** monitored drift tube. 34, 36, 37
- ME** Matrix Element. 6, 7, 14, 52–55, 84, 90, 92, 121, 122, 169
- MS** muon spectrometer. 34, 36–39, 42, 43, 169
- NLO** next-to-leading order. 6, 23, 25, 53, 54, 56, 84, 131, 132
- NNLL** next-to-next-to-leading log. 6, 20, 21, 23, 53
- NNLO** next-to-next-to-leading order. 6, 20, 21, 23, 53, 54, 106
- NP** nuisance parameter. 81, 83, 84, 87, 90, 92–95, 118, 121, 122, 140, 150, 169
- P2012** Perugia 2012. 52–55, 85
- PDF** parton distribution function. 16, 20, 22, 52–56, 85, 92, 106, 127
- Pixel** pixel detector. 31, 32, 39, 41
- POI** parameter of interest. 81, 83, 84, 87, 90, 169
- PS** parton shower. 18, 25, 46, 52–55, 84, 90, 92, 121, 169
- PU** pile-up. 17, 40, 45, 49, 51–53, 55, 57, 71, 86, 92, 122
- QCD** quantum chromodynamics. 12, 13, 18, 52, 57
- QED** quantum electrodynamics. 1, 9, 11, 12, 15, 16, 54
- QFT** quantum field theory. 5
- ReLU** rectified linear unit. 115, 116, 119
- ROC** receiver operating characteristic. 68, 74, 76, 167
- RPC** resistive plate chamber. 34, 36, 37
- SCT** semiconductor tracker. 31, 32, 39
- SGD** stochastic gradient descent. 117
- SM** Standard Model. 2, 7–12, 14, 18, 19, 23, 80, 84, 127, 169
- TGC** thin gap chamber. 34, 36, 37
- TRT** transition radiation tracker. 31, 33, 39–41
- UE** underlying event. 17, 46, 52, 53, 57

Index

- Wtb* vertex, [12](#)
- p*-value, [82](#)
- b*-tagging
 - MV2c10, [47](#)
 - working point, [48](#)
- diagram removal (DR), [24](#)
- DR2, [25](#)
- diagram subtraction (DS), [24](#)
- electroweak symmetry breaking (EWSB), [10](#)
- inner detector (ID), [31](#)
- jet-vertex-fraction (JVF), [45](#)
- jet-vertex-tagger (JVT), [45](#)
- Kullback-Leibler (KL) divergence, [114](#)
- Kolmogorov–Smirnov (KS) test, [70](#)
- Large Hadron Collider (LHC) Run 2, [29](#)
- LHC experiment
 - A Toroidal LHC ApparatuS (ATLAS), [28](#)
 - ALICE, [28](#)
 - CMS, [28](#)
 - LHCb, [28](#)
 - LHCf, [28](#)
 - MoEDAL, [29](#)
 - TOTEM, [28](#)
- Monte Carlo (MC), [52](#)
- MC simulation
 - GEANT 4, [52](#)
 - HERWIG, [53](#)
 - PYTHIA, [52](#)
- Matrix Element (ME), [6](#)
- muon spectrometer (MS), [34](#)
- nuisance parameter (NP), [81](#)
- parameter of interest (POI), [81](#)
- parton shower (PS), [18](#)
- Standard Model (SM), [7](#)
- RooUnfold, [100](#)
- $t\bar{t}$ production, [19](#)
- absorption length, [34](#)
- activation function, [115](#)
- Asimov data, [84](#)
- asymptotic freedom, [13](#)
- backprop, [117](#)
- batch size, [117](#)
- boosting, [66](#)
 - adaptive boosting, [67](#)
- boson
 - gauge boson, [9](#)
 - Goldstone boson, [10](#)
- branching fraction, [15](#)
- calorimeter, [33](#)
- capacity, [116](#)
- centre-of-mass energy, [16](#)
- colour charge, [12](#)
 - colour confinement, [12](#)
- colourless, [12](#)
- coupling strength, [5](#)
- CP violation, [12](#), [14](#)
- cross-section, [14](#)
 - differential cross-section, [15](#)
 - inclusive cross-section, [14](#)
- dark matter, [14](#)
- decay rate, [15](#)
- decision tree, [65](#)
- electron, [41](#)
 - electron calibration, [42](#)
 - electron identification, [41](#)

- electron isolation, 41
- electron reconstruction, 41
- fake electron, 41
- entropy
 - cross-entropy, 114
 - differential entropy, 113
 - Shannon entropy, 114
- event rate, 15
- fake, 59
- fermion, 7
- Feynman diagram, 7
- five-flavour scheme, 22
- four-flavour scheme, 22
- gauge
 - gauge covariant derivative, 6
 - gauge field, 6
 - gauge invariance, 6
 - gauge theory, 6
- Gini index, 66
- gradient descent, 117
- hadron, 13
- hadronisation, 13
- hard collision/scattering, 16
- Higgs
 - Brout-Englert-Higgs mechanism, 10
 - Higgs field, 10
 - vacuum expectation value, 10
- hypothesis
 - alternative hypothesis, 82
 - hypothesis testing, 82
 - null hypothesis, 82
 - test statistic, 82
- information theory, 113
- jet, 18, 44
 - anti- k_t algorithm, 44
- lifetime, 15
- light quarks, 8
- likelihood
 - binned likelihood, 80
 - log-likelihood, 80
 - maximum likelihood, 80
- loss function, 67, 117
 - absolute loss function, 68
 - Huber loss function, 68
 - logistic loss function, 68
 - squared loss function, 68
- luminosity
 - instantaneous luminosity, 15
 - integrated luminosity, 15
- matrix inversion unfolding, 98
- missing transverse momentum, 16, 48
- multivariate analysis, 65
- muon, 42
 - muon calibration, 44
 - muon identification, 43
 - muon isolation, 44
 - muon reconstruction, 43
- natural units, 14
- neutrino oscillation, 14
- new physics, 14
- optimisation, 117
- overtraining, 68
- parton, 16
- perturbative expansion, 6
- profiling, 83
- pseudo-experiment, 83
- pseudorapidity, 17
- radiation lengths (X_0), 33
- rapidity, 17
- regularised unfolding
 - regularisation function, 99
 - regularisation parameter, 99
- scattering matrix, 6
- self-information, 113
- signal strength, 80
- solenoid, 37
- test sample, 68
- toroid, 37
- transverse momentum, 16
- trigger

high-level trigger (HLT), [38](#)

level-1 (L1) trigger, [38](#)

uncertainty

rate uncertainty, [79](#)

shape uncertainty, [79](#)

statistical uncertainty, [79](#)

systematic uncertainty, [79](#)

unfolding, [97](#)

bin-by-bin unfolding, [98](#)

iterative unfolding, [99](#)

regularised unfolding, [99](#)

unitary gauge, [10](#)

vertex

hard-scatter vertex, [40](#)

weak isospin and hypercharge, [9](#)

Weinberg angle, [9](#), [11](#)

Wolfenstein parametrisation, [12](#)

Acknowledgements

In 2013, I had the fortune to come to Germany and work with a remarkable group for my doctoral research at the University of Bonn. The fruitful journey would not have started without a) support from Bonn-Cologne Graduate School, and b) the opportunity given by my supervisor, Prof. Ian C. Brock. During the journey, Ian persistently encouraged me to pursue my own interests, gave me the freedom to organise my research work, and supported me with valuable insights whenever I am off-track. His passion for accepting new technologies, reforming team management, and his unyielding spirit while playing Kegeln, motivated me to explore high energy physics. He spent most of his time with the group sharing knowledge and being concerned with everybody's problems regarding research and life. Being extremely patient and careful, he created a friendly and productive environment, which made my doctoral life not as awful as described in Quora.com at all.

The next person I must give credits to in the group is Dr. Regina Molles-Valls. We have been working together closely since she joined the group in 2015, when I was working on the qualification task for the particle-flow algorithm. We collaborated from particle flow to tW analysis, from coding to writing documents, from teaching to making Doktorwagens. We had our first big success when she helped me on the application to the Honours Branch of the Bonn-Cologne Graduate School which, without her suggestions, I would have failed at. She followed the status of all the group members without losing track of her own research. Being assertive and friendly, she provided lots of valuable comments to this thesis.

A great group atmosphere was contributed by other group members and former members as well. They were like family to me during the years. Thomas Velz introduced me to the group and spent much time explaining to me the basics of data analysis and the ATLAS experiment. Jan Stillings gave me so many valuable suggestions for my study and academic career. Sebastian Mergelmeyer taught me lots of programming. Ozan Arslan supervised me from the first day, and we became friends immediately afterwards. Without him, I could not have been integrated into the group that quickly and smoothly. His support went far beyond the work; his constant willingness to help me in all aspects is appreciated. I also shared so many great and fun moments with Irina Cioară, who started her PhD the same year I joined the group. I cannot forget the small talks we had during the Terascale schools in Hamburg and München, nor those in the group flat at CERN. We had the greatest moment in Braga which I cherish greatly; it was, of course, the “highlight of our PhDs”! I would also like to thank Anjishnu Bandyopadhyay, a person I could talk to on various topics. Our discussion started with the statistical analysis for the vector-like quark search and continued with IT support and CERN rumours. I am grateful to have worked with Tanja Holm and Federico G. Diaz Capriles, specifically for the fascinating discussions on particle physics, programming, and machine learning, and also to have been able to talk to Professor Ewald Paul occasionally in Christmas parties and in the corridor. In addition, I will not miss the chance to thank all the master and bachelor students who were in the group during my days there, including but not limited to Elena (Zarkh) Amell, Nicolas Boeing, Christian Kirfel,

Peter Falke, Marius Blaut, Alexander Johnston. There would be considerably less fun without the students!

There are many other people at the institute that I will remember. I am grateful for the numerous suggestions, discussions, and critical questions that I received from people of the Desch group, including Klaus Desch (certainly!), Philip Bechtle, Oliver Ricken, Maike Hansen, Lara Schildgen, and Philipp König. I would like to express my gratitude to the local IT-supporters who did remarkably great jobs to keep the machines running: Peter Wienemann, Jan Stillings, Oliver Freyermuth, and all the other members. A special thank you also goes to Martin Schultens and David Hohn for having drinks (and even dancing) during conferences.

Working with the tW analysis group is quite enjoyable. Without the collaboration with these people, the analyses would not have been in good shapes, let alone this thesis. The group was formed in 2015 for the inclusive cross-section analysis with Kevin Finelli, Jin Wang, and Carl Suster. Kevin, Carl and I continued with the differential cross-section analysis. Being knowledgeable and supportive, Kevin was always ready to help me along the way, particularly in programming and physics. Furthermore, I enjoyed working in the ATLAS single top-quark group, where I got the generous support of multiple group and sub-group conveners, especially Carlos Escobar, Carolina Gabaldon, Mark Owen, Frederic Deliot, Reinhard Schwienhorst, and Wolfgang Wagner. In addition, I would like to acknowledge Reinhard Schwienhorst and Wolfgang Wagner (besides Ian) for supporting me in my further academic career and writing recommendation letters.

I was fortunate to study in Germany, 8 000 km away from my hometown, with my parents' and my sister's unconditional love and unlimited supports. They gave me the freedom to choose my career and encouraged me to start my venture; without them, I would never have enjoyed so many opportunities. Special thanks must go to my beloved soulmate, Ruijia Wei, as well, for being the anchor in my life. Being smart and beautiful, she filled my world with joy and peace, and inspired me to follow my dreams. I am fortunate to have her nearby.

Last but not least, I extend gratitude to the whole examining committee: Prof. Ian C. Brock, Prof. Norbert Wermes, PD. Dr. Metsch, Prof. Matthias Hulin for their interest in my dissertation and the solid scientific discussions.

Curriculum Vitae

Personal Details

Name	Rui Zhang
Date of Birth	06 November 1988
Email	zhang@physik.uni-bonn.de / rui.zhang@cern.ch

Education

2013–2019	PhD in Physics, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany.
2010–2013	MSc in Physics, University of Chinese Academy of Sciences, Beijing, China.
2006–2010	BSc in Physics, Tianjin University, Tianjin, China.

Professional Experience

2018	Talk at the 7 th ATLAS Single Top-quark Workshop, Braga, Portugal.
2018	Machine Learning Workshop, Geneva, Switzerland.
2017	Poster at the 10 th International Workshop on Top Quark Physics, Braga, Portugal.
2017	Talk at the 11 th Annual Helmholtz Alliance Workshop, Hamburg, Germany.
2017	Talk at the ATLAS-D Physics Meeting, Dortmund, Germany.
2016	Talk at the 24 th International Conference on Supersymmetry, Melbourne, Australia.
2015–2016	Fieldwork at CERN, Geneva, Switzerland.
2015	Summer Student at Dubna International Advanced School of Theoretical Physics, Dubna, Russia.
2014	Summer Student at Belgian Dutch German School, Kerkrade, Netherlands.

Publications

2017	Measurement of differential cross-sections of a single top-quark produced in association with a W boson at $\sqrt{s} = 13$ TeV with ATLAS.
2016	Measurement of the cross-section for producing a W boson in association with a single top-quark at $\sqrt{s} = 13$ TeV with ATLAS.
2015	Search for the production of single vector-like and excited quarks in the Wt final state at $\sqrt{s} = 13$ TeV with the ATLAS detector.

Awards

2017	Poster prize (2 nd place) in the 10 th International Workshop on Top Quark Physics.
2015	Honours branch (stipend) of Bonn-Cologne Graduate School.

Languages

Chinese	Mother tongue
English	Fluent