

STUDY OF W^+W^- VECTOR BOSON SCATTERING WITH THE ATLAS DETECTOR.

Christophe Roland

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Physics,

Indiana University

March 2023

CERN-THESIS-2023-022
27/01/2023



Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for
the degree of Doctor of Philosophy.

Doctoral Committee

Sabine Lammers, PhD

Harold Evans, PhD

Matthew Shepherd, PhD

Emilie Passemar, PhD

January 27, 2023

Acknowledgments

First and foremost, I would like to thank my supervisor Prof. Sabine Lammers for her continuous support and guidance. Working with her has been very enjoyable throughout my PhD. I would also like to thank all the other members of my committee, Hal Evans, Matthew Shepherd and Emilie Passemar for all of their support and comments.

Working with the Indiana ATLAS group has been a pleasure and has helped me learn a lot. Thanks to all of you for all your help and insight. Special thanks to Sabine Lammers, Philippe Calfayan, Hal Evans, Fred Luehring, Chris Meyer, KyungEon Choi, Rebecca Linck, Greg Myers, Gabriel Palacino, Revital Kopeliansky and Alex Johnson.

Many thanks as well to the members of the ATLAS *WW* run 2 group, I would like to thank in particular Hannes Mildner, Jan Kuechler, Valerie Lang and Kristin Lohwasser for all of their help and advises.

I also had the privilege of working in the TRT software group during my PhD which was very enjoyable and a great learning experience. Thanks to all the members of the TRT group and special thanks to Peter Hansen, Anatoly Romaniuk, Christian Grefe, Andrew Beddall, KyungEon Choi, Sina Aktas, Natalia Korotkova and Aytul Adiguzel. An even more special thank to my dear friend Eleni Skorda, it was really great working with you on the TRT and thank you for your moral support and for all your help in proofreading my thesis.

Finally, I would like to thank all of my friends and my family, with special thanks to all of my siblings for all the happy times we lived together.

Christophe Roland

STUDY OF W^+W^- VECTOR BOSON SCATTERING WITH THE ATLAS DETECTOR.

This thesis presents an analysis of the electroweak production of pairs of opposite-sign W bosons in association with two or three jets using data from proton-proton collisions collected by the ATLAS experiment at a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 139 fb^{-1} .

The analysis is done in the $W^+W^- \rightarrow e^\pm \mu^\mp$ channel, using neural networks to suppress the top and QCD- WW background. A likelihood fit is performed to estimate the signal strength and the corresponding uncertainty.

The cross-section was measured in a fiducial region and was found to be $2.62 \pm 0.50 \text{ fb}$ ($2.20 \pm 0.47 \text{ fb}$ expected) in agreement with the Standard Model prediction. The observed (expected) significance is 7.06σ (6.21σ). This is the first-ever observation and measurement of this process in the ATLAS detector.

Contents

Acknowledgments	iii
Abstract	iv
1 Introduction	1
2 The Standard Model	3
2.1 Introduction	3
2.2 Particles content	3
2.3 Quantum Field Theory	5
2.3.1 Lagrangian formulation	5
2.3.2 Perturbation theory	5
2.3.3 Renormalization	6
2.3.4 Running coupling constant	7
2.3.5 Cross sections	9
2.4 Fundamental interactions	10
2.4.1 Quantum electrodynamics	10
2.4.2 Quantum chromodynamics	11
2.4.3 Electroweak interaction	12
3 Vector boson scattering	16
3.1 Unitarity	16
3.2 Current experimental status	18
3.3 Electroweak production of W^+W^- with two jets	20
4 Event simulation for pp collision	23
4.1 Introduction	23
4.2 Partons distribution functions and factorization	23
4.3 Hadronic jets	25
4.3.1 Initial and final state radiation	28

4.3.2	Matching and merging	29
4.4	Event generation	30
4.4.1	Monte Carlo integration	30
4.4.2	Monte Carlo event generators and parton shower software	31
5	ATLAS and the LHC	34
5.1	The Large Hadron Collider	34
5.1.1	Introduction	34
5.1.2	The detectors	34
5.1.3	The accelerator complex	35
5.1.4	LHC run	36
5.2	The ATLAS detector	37
5.2.1	The ATLAS coordinate system	37
5.2.2	The ATLAS superconducting magnet system	40
5.3	The ATLAS inner detector	40
5.3.1	The pixel detector	41
5.3.2	The semiconductor tracker	41
5.3.3	The transition radiation tracker	42
5.4	The ATLAS calorimeter system	42
5.4.1	The liquid-argon calorimeter	44
5.4.2	The tile calorimeter	45
5.5	The muon spectrometer	45
5.6	The trigger and data acquisition system	45
6	Reconstruction	48
6.1	Introduction	48
6.2	Tracking	48
6.3	Vertex reconstruction	49
6.4	Topological calorimeter clusters	50
6.5	Leptons	50
6.5.1	Electrons	50

6.5.2	Muons	51
6.5.3	Lepton isolation	51
6.6	Jets	52
6.6.1	Jet reconstruction	52
6.6.2	B-tagging	53
6.7	Missing energy	54
7	Data and Monte Carlo samples	55
7.1	Introduction	55
7.2	Data samples	55
7.3	Monte Carlo samples	56
7.3.1	Signal sample	56
7.3.2	Background samples	57
8	Object definition and event selection	59
8.1	Introduction	59
8.2	Object definition	59
8.2.1	Electrons	59
8.2.2	Muons	60
8.3	Jets	60
8.4	Missing transverse energy	61
8.5	Selection	62
8.5.1	Cutflow	64
9	Neural network Analysis	66
9.1	Introduction	66
9.2	Neural networks: definition and theoretical background	66
9.2.1	Definition	66
9.2.2	Classification	68
9.2.3	Training	68
9.2.4	Overfitting	71

9.3	Neural network used for the analysis	72
9.3.1	Introduction	72
9.3.2	Activation functions	72
9.3.3	Overtraining check	72
9.3.4	Neural network inputs	75
9.3.5	Neural network settings	79
9.3.6	Neural network cutflow	80
9.3.7	Validation	82
9.3.8	Control region	86
9.3.9	High NN region	88
9.3.10	Input variables in the signal region	90
10	Uncertainties	92
10.1	Introduction	92
10.2	Experimental uncertainties	92
10.3	Data-driven fake estimation	98
10.4	Theoretical uncertainties	100
10.4.1	Top theoretical uncertainties	100
10.4.2	Signal theoretical uncertainties	101
10.4.3	QCD-WW theoretical uncertainties	101
10.4.4	QCD-EWK WW production interference	103
11	Likelihood Fit	104
11.1	Introduction	104
11.1.1	Probabilites	104
11.1.2	Profile likelihood	106
11.1.3	Profile likelihood ratio	107
11.2	Neural network fit	108
11.2.1	Overview of the likelihood fit	108
11.3	Smoothing of neural network distribution	111
11.3.1	Fake lepton uncertainty	111

11.3.2 Nuisance parameters ranking	111
11.4 Cross-section	117
12 Conclusion	122
Bibliography	123
A List of experimental uncertainties	135
B Additional experimental uncertainty plots	137
Curriculum Vitae	

List of Figures

2.1	Fundamental particles in the Standard Model. Illustration from [2].	4
2.2	The variations of the electromagnetic (α_1), weak (α_2) and strong (α_3) coupling constants with respect to energy are shown here. They get fairly close to each other at around 10^{16} GeV (often called the <i>grand unification scale</i>). Illustration taken from [6]	8
2.3	Representation of the Higgs potential for the simpler case where ϕ is a complex field instead of a complex doublet. Radial oscillations h correspond to the Higgs boson and angular oscillations π correspond to a Goldstone boson.	13
3.1	VBF, VBS and tri-boson cross-section measurements by the ATLAS collaboration[24]. The ratio $\sigma_{\text{exp}}/\sigma_{\text{theo}}$ are shown with the corresponding uncertainties.	19
3.2	Cross section measurement of electroweak production of one or two vector bosons by the CMS collaboration[25]. The ratio $\sigma_{\text{exp}}/\sigma_{\text{theo}}$ are shown with the corresponding uncertainties.	20
4.1	Plots of the parton distribution function multiplied by the momentum fraction x for $Q^2 = 10\text{GeV}^2$ (left) and $Q^2 = 10^4\text{GeV}^2$ (right) for different partons. u_V and d_V are the valence quarks and it can be observed that the gluons and the sea quarks contribute the most at smaller momentum fractions and that they contribute more at higher energy. Plots taken from [29].	25
4.2	Jet seen by the ATLAS detector. The collision point is at the bottom of the picture and a bunch of collimated particles can be seen, forming the jet. Image taken from [31].	27
4.3	Representation of the gluon field for a quark-antiquark pair using lattice QCD computation. A string-like structure can be seen (often called a flux-tube). The colors show the expectation value of the square of the chromoelectric field: $\langle E^2 \rangle$. Illustration taken from [35].	28

4.4	Illustration of proton-proton collision. The red blob represents the hard process. From the hard process, several particles emerge (in red). Some of them create parton showers that produce hadrons (hadronization is represented by light green blobs) and many of these hadrons subsequently decay. multiparton interactions are also represented at the bottom (purple blob). Illustration taken from [40]	32
5.1	Location of the LHC and the four main experiments near Geneva [53].	35
5.2	Illustration of the LHC accelerator complex. Illustration taken from [63].	36
5.3	Average number of interactions per bunch crossing (left) and integrated luminosity (right) for run 2. [64]	37
5.4	The ATLAS detector [54].	38
5.5	The ATLAS coordinate system. Illustration taken from [65].	39
5.6	ATLAS magnetic field in the transverse section (left) and the longitudinal section (right). Illustration from [66]	40
5.7	A 3D view of the ATLAS inner detector. Illustration from [67]	41
5.8	Quadrant view of the ATLAS inner detector. Illustration from [68]	42
5.9	The ATLAS calorimeter system. [54]	43
5.10	Illustration of the structure of the liquid-argon calorimeter. [74]	44
5.11	The ATLAS muon spectrometer. [77]	46
5.12	An illustration of the ATLAS TDAQ [78]. Data flows through the L1 triggers, then through the HLT before being recorded.	47
6.1	The perigee representation of a track in ATLAS. Illustration from [65].	49
6.2	Comparison of clustering algorithms: Anti- k_T (left), Cambridge/Aachen (middle) and k_T (right). Illustration from [92].	53
8.1	Number of jets for signal events in the signal region (but without a cut on the number of jets). There are not many events with more than three jets.	63
8.2	We apply a cut on m_{ll} ($m_{ll} > 80$ GeV) to suppress VBF HWW and ggF HWW (both combined in red) and remove the Higgs peak. The signal (shown in grey) sample was produced without the Higgs resonance.	63

8.3	Cutflow for the signal region.	65
8.4	Signal/backgrounds ratio cutflow for the signal region.	65
9.1	Example of a network with one hidden layer. The lines represent multiplication by some weight $w_{ij}^{(k)}$ and in each node, an activation function is applied. The red dots correspond to the biases.	67
9.2	The rectified linear units function.	70
9.3	Illustration of overfitting for binary classification. The light blue line represents a case of overfitting while the green line shows a good fit.	71
9.4	Comparison between variables used for training the NN output for $n_{\text{jets}} = 2$ and $n_{\text{jets}} = 3$ in the signal region. Significant differences can be seen in the leading and sub-leading jet p_{T} , $\Delta\eta_{jj}$ and centrality distributions.	73
9.5	Training and validation loss in function of the number of epochs. The optimal point is indicated by the vertical line. As expected, the validation loss grows after the minimum while the training loss keeps getting smaller.	74
9.6	Normalized neural network output distribution for $n_{\text{jets}} = 2$ (left) and $n_{\text{jets}} = 3$ (right) in the signal region to check for possible overtraining.	74
9.7	Background rejection vs signal efficiency for $n_{\text{jets}} = 2$ (left) and $n_{\text{jets}} = 3$ (right) in the signal region.	75
9.8	Normalized distribution of input variables for $n_{\text{jets}} = 2$ (above) and $n_{\text{jets}} = 3$ (below) in the signal region.	76
9.9	Input variables correlation matrix for $n_{\text{jets}} = 2$ (left) and $n_{\text{jets}} = 3$ (right) in the signal region for the signal (top) and background (bottom) samples.	77
9.10	Correlation between total m_{lj} and leading jet p_{T} (above) and correlation between $\Delta\eta_{jj}$ and m_{ll} (below) for low NN output (NN output < 0.6) for $n_{\text{jets}} = 2$ (left) and $n_{\text{jets}} = 3$ (right) in the validation region. The error bars correspond to the standard error on the mean and do not include systematic uncertainties.	82
9.11	Correlation between variables used for training the NN output for $n_{\text{jets}} = 2$ in the signal region. The error bars correspond to the standard error on the mean and do not include systematic uncertainties.	83

9.12	Correlation between variables used for training the NN output for $n_{\text{jets}} = 3$ in the signal region. The error bars correspond to the standard error on the mean and do not include systematic uncertainties.	84
9.13	Validation region plots for $n_{\text{jets}} = 2$. We can observe a good agreement with data. All uncertainties are included.	85
9.14	Validation region plots for $n_{\text{jets}} = 3$. We can observe a good agreement with data. All uncertainties are included.	86
9.15	Control region plots for $n_{\text{jets}} = 2$. All uncertainties are included. A good agreement with data can be observed.	87
9.16	Control region plots for $n_{\text{jets}} = 3$. All uncertainties are included. A good agreement with data can be observed.	88
9.17	Observable in the signal region with a neural network > 0.9 cut. All uncertainties are included.	89
9.18	Observables used as input for the neural network in the signal region. All uncertainties are included.	90
9.19	Observables used as input for the neural network in the signal region. All uncertainties are included.	91
10.1	Distributions for the signal in the signal region for the main experimental uncertainties with the systematics/nominal ratio.	95
10.2	Distributions for the top background in the signal region for the main experimental uncertainties with the systematics/nominal ratio.	96
10.3	Distributions for the WW background in the signal region for the main experimental uncertainties with the systematics/nominal ratio.	97
10.4	NN output in the fake-enriched region.	99
10.5	Transfer factor values for different NN output bins.	99
10.6	Distributions for the top background in the signal region for the main theoretical uncertainties with the systematics/nominal ratio.	102
10.7	Ratio plot comparing the events from the signal and the interference. Plot produced by Rabia Omar.	103

11.1	NN output in the signal region before (left) and after (right) the likelihood fit ($n_{\text{jets}} = 2$).	109
11.2	NN output in the signal region before (left) and after (right) the likelihood fit ($n_{\text{jets}} = 3$).	110
11.3	NN output in the control region before (left) and after (right) the likelihood fit. . . .	110
11.4	Smoothing of the systematics. The dashed (resp. solid) line shows the distribution before (resp. after) smoothing. These plots show the smoothing of the neural network output for the top parton shower uncertainty, in the signal region for two jets (left) and three jets (right)	112
11.5	Nuisance parameters correlation matrix.	113
11.6	Nuisance parameter ranked by impact in order of post-fit impact. γ corresponds to the Monte Carlo statistical uncertainty in a particular bin. The upper x-axis corresponds to the uncertainty of the signal strength $\Delta\mu$, the pre-fit uncertainties are shown with the bars filled with white and the post-fit ones by fully colored bars. The lower y-axis shows the nuisance parameter pull $(\hat{\theta} - \theta_0)/\Delta\theta$ where $\hat{\theta}$ is the best-fit value of the nuisance parameter θ , θ_0 is the nominal (pre-fit) value of θ and $\Delta\theta$ is the uncertainty. The pull for the MC statistical uncertainties or any other scale factor is expected to be around one because it is just rescaling the MC prediction.	115
11.7	Signal strength estimation from the likelihood fit (bottom) as well as top and WW-QCD normalization.	116
11.8	Likelihood scan for the signal strength μ	116
11.9	Neural network distribution for two jets (left) and three jets (right). The distribution without an additional generator-level m_{jj} cut is shown in blue and the distribution with an additional $m_{jj} > 500$ GeV generator-level cut in red. We can observe that elements with $NN < 0.6$ are suppressed.	118
11.10	Cut efficiencies for $NN > 0.6$ and generator-level m_{jj} cut for the two jets signal region.	119
11.11	Cut efficiencies for $NN > 0.6$ and generator-level m_{jj} cut for the three jets signal region.	120
11.12	m_{jj} tribosons distribution in the signal region.	120
B.1	Distributions for the W +jets background in the signal region for the main experimental uncertainties with the systematics/nominal ratio.	137

B.2 Distributions for the Z+jets background in the signal region for the main experimental uncertainties with the systematics/nominal ratio. 138

B.3 Distributions for the multibosons background in the signal region for the main experimental uncertainties with the systematics/nominal ratio. 139

List of Tables

7.1	List of single-electron triggers used for the analysis.	55
7.2	List of single-muon triggers used for the analysis.	56
7.3	Number of events in the signal region for the electroweak production of one electron and one muon with opposite sign compared with the number of events where one of the lepton is a tau lepton, for $n_{\text{jets}} = 2$ and $n_{\text{jets}} = 3$	57
7.4	List of MC datasets used in the analysis.	58
8.1	Requirements for nominal and veto electrons.	60
8.2	Requirements for nominal and veto muons.	61
8.3	Signal and control regions definitions.	64
9.1	For each variable removed, the network (here for $n_{\text{jets}} = 2$) is re-trained 20 times and the AUC average is calculated. For each variable, a score is given, and higher scores indicate larger impacts. ‘Default’ means no variable is removed.	78
9.2	For each variable removed, the network (here for $n_{\text{jets}} = 3$) is re-trained 20 times and the AUC average is calculated. For each variable, a score is given, and higher scores indicate larger impacts. ‘Default’ means no variable is removed.	79
9.3	Configuration of the neural network constructed using TMVA in conjunction with Keras.	80
9.4	Neural network cutflow.	81
9.5	Neural network cutflow.	81
10.1	Jet experimental uncertainties.	93
10.2	Muon experimental uncertainties.	93
10.3	Electron experimental uncertainties.	93
10.4	Various experimental uncertainties.	94
11.1	Systematic uncertainties and the corresponding post-fit uncertainty on the signal strength. The statistical and total uncertainties are also shown.	114
11.2	Fiducial region definition.	119

A.1	Jet experimental uncertainties.	135
A.2	Muon experimental uncertainties.	136
A.3	Electron experimental uncertainties.	136
A.4	Various experimental uncertainties.	136

Chapter 1

Introduction

The 20th century saw unprecedented progress in fundamental physics. In particular, the development of quantum physics has revolutionized our understanding of matter and radiation. Many new particles were discovered and fundamental interactions were studied and described in great detail. This progress led to the development of the Standard Model, our best current theory for particle physics.

The impressive predictive power of the Standard Model has been tested, with great precision, using ever more powerful particle colliders. The discovery of the Higgs boson in 2012 was one of the most spectacular confirmations of the validity of the Standard Model.

However, theoretical and experimental evidence indicates that the Standard Model is not the most fundamental theory of particle physics. Collider experiments, such as the ATLAS experiment at the Large Hadron Collider (LHC), are therefore important to test the Standard Model and potentially discover evidence of new particles or interactions, not predicted by the Standard Model.

The ATLAS detector is well suited for the study of weak gauge bosons (particles ‘carrying’ the weak force), as well as their interactions. In this thesis, an analysis of the electroweak production of opposite charge W bosons accompanied by quarks is presented. This process is quite difficult to measure, as it has a very small cross-section and a fairly large irreducible background. This is why neural networks are used for optimal signal and background separation for this analysis.

Chapters 2 and 3 provide the theoretical background for the W^+W^- electroweak production: chapter 2 shows an overview of the Standard Model, chapter 3 then provides a theoretical summary of vector boson scattering, showing also previous measurements in ATLAS and CMS for vector boson scattering and vector boson fusion processes.

The next three chapters are about data gathering, simulation and calibration. Chapter 4 gives information about event simulation in pp colliders (such as the LHC), chapter 5 discusses the LHC and the ATLAS detector and chapter 6 shows how physical objects, such as particles and jets, are reconstructed from the ATLAS raw data.

Chapter 7 through 11 discuss the W^+W^- electroweak production analysis. Chapter 7 shows the data and simulation samples, chapter 8 shows the event selection, chapter 9 presents the neural

networks used for the analysis, chapter 10 shows the systematic uncertainties and how they are estimated and chapter 11 presents the likelihood fit used to extract the signal significance and measure the fiducial cross-section.

Chapter 12 then summarizes and concludes the thesis.

Chapter 2

The Standard Model

2.1 Introduction

The Standard Model of particle physics is a relativistic quantum field theory that includes all the known fundamental particles and their (known) interactions with the exception of gravity. It is the theoretical foundation of all of modern particle physics.

This present chapter starts with an overview of the particles and interactions in the Standard Model. The mathematical and conceptual basis of quantum field theory are then introduced followed by a presentation of the basic equations describing the electromagnetic and strong interactions in the Standard Model. The weak interaction is then discussed as well as the electroweak theory (a unified description of the weak and electromagnetic interactions) and the Higgs mechanics.

2.2 Particles content

A number of fundamental particles are postulated to exist in the Standard Model. As of this writing, all of those particles have been observed experimentally and no particle that would not fit in the Standard Model has been directly observed so far.

The particles in the Standard Model can be put into two categories: *fermions* and *bosons*. Fermions have half-integer spin¹ and have the property that two identical fermions cannot be in the same quantum state, this is the *Pauli exclusion principle*. Bosons have integer spin and there is no limit to how many identical bosons there can be on the same quantum state.

Fundamental fermions in the Standard Model are further categorized into *leptons* and *quarks*. There are three *generations* of leptons and in each generation, there is one charged and one neutral lepton. The three charged leptons are the electron, the muon and the tau. The neutral leptons are called neutrinos (the electron neutrino, the muon neutrino and the tau neutrino). The generations are ordered by the charged lepton mass: the electron (first generation) is much lighter than the muon (second generation) which is itself much lighter than the tau (third generation). It is still an

¹The spin of a particle corresponds to its intrinsic angular momentum.

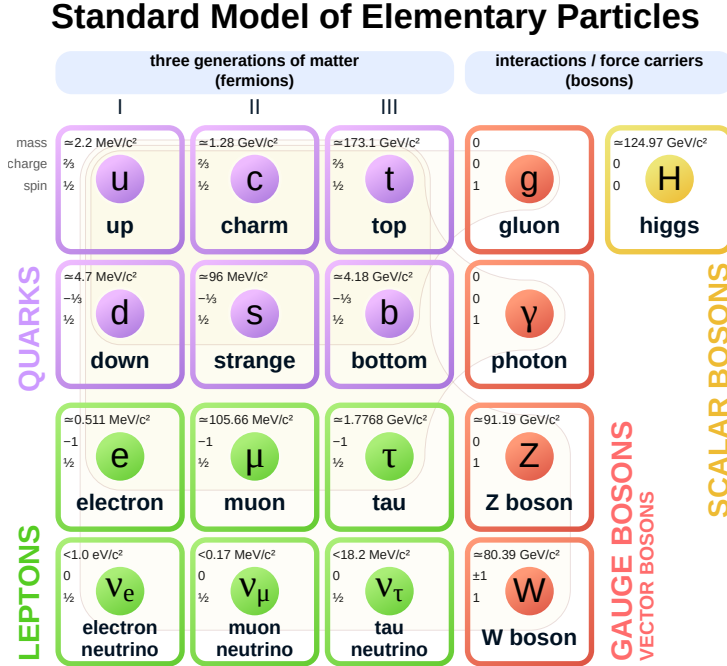


Figure 2.1: Fundamental particles in the Standard Model. Illustration from [2].

unresolved question as to why there are three generations with different masses.

For each (negatively) charged lepton, there also exists a corresponding *antiparticle* with an opposite (positive) electric charge (the positron, the anti-muon and the anti-tau). It is still unknown if neutrinos are also their own antiparticles. The neutrinos were initially postulated to be massless but experimental evidence of neutrino oscillations² show that they have a (very small but non-zero) mass[1].

They are also three generations of quarks, with two quarks in each generation (and the corresponding antiquarks). These two quarks are grouped into what is called an *isospin doublet* ($\begin{pmatrix} u \\ d \end{pmatrix}$) combining a quark with electric charge $+2/3$ with a quark with electric charge $-1/3$. Unlike leptons, quarks carry a *color charge* which is the charge associated with the strong interaction, the charges are by convention called red, green and blue as well as anti-red anti-green and anti-blue. Quarks form composite particles called *hadrons* that do not carry any charge color (they are said to be “white” or “colorless”).

Fundamental bosons in the Standard Model are the spin-1 *gauge bosons* and the (spin-0) Higgs

²Neutrino oscillation is the fact that neutrinos can oscillate between different flavors i.e. an electron neutrino can later be measured as a muon (or a tau) neutrino. This is explained by the mixing of flavor and mass eigenstates.

boson. The gauge bosons are associated with fundamental interactions. The photon is associated with electromagnetism, the eight gluons are associated with the strong interaction and the Z and W bosons with the weak interaction.

Figure 2.1 shows the fundamental particles in the Standard Model, as well as their mass, charge and spin.

2.3 Quantum Field Theory

2.3.1 Lagrangian formulation

It is often convenient (when possible) to use the Lagrangian formalism to define a quantum field theory. The starting point is a mathematical object called a Lagrangian density \mathcal{L} which is a function of the fields ϕ_i and their derivatives $\partial_\mu\phi_i$. The Lagrangian and the values of all the free parameters (variables in the Lagrangian that are not constrained to specific values) are in principle enough to fully specify the theory.

From the Lagrangian, we can define the action S :

$$S[\phi_i] = \int d^4x \mathcal{L}[\phi_i(x), \partial_\mu\phi_i(x)] \quad (2.1)$$

Requiring that $\delta S = 0$ leads to the Euler-Lagrange equations:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu\phi_i)} \right) - \frac{\partial \mathcal{L}}{\partial\phi_i} = 0 \quad (2.2)$$

These equations are the equations of motions (in the Heisenberg picture) of the quantum fields ϕ_i which are (represented by) operators acting on a Hilbert space.

In practice, it is not possible to solve these equations exactly. To get around this problem, various approximation techniques are used, the most often used being *perturbation theory*.

2.3.2 Perturbation theory

The fundamental idea behind perturbation theory is to first solve (exactly) the free theory (i.e. the theory without interaction). The interactions are then added and the quantum amplitude is computed using an asymptotic series, expanding in powers of g , the coupling constant of the

interaction.

It turns out that there is a convenient tool to calculate terms in the asymptotic series. Each term can be represented by a diagram with in-going lines representing the in-going particles, out-going lines representing the out-going particles and internal lines connecting them called “virtual particles”. These representations are called *Feynman diagrams*. A set of rules (called Feynman rules) can be used to write down the term corresponding to a particular diagram.

As an example, here are a few diagrams for the $e^-e^- \rightarrow e^-e^-$ process:

$$\mathcal{M} = \text{[tree-level exchange]} + \text{[one-loop bubble]} + \text{[one-loop vertex correction]} + \dots$$

In this case, each vertex brings a factor of e (the electromagnetic coupling constant). The first term, proportional to e^2 , is the *leading order* term, and the other two, proportional to e^4 , are parts of the next-to-leading order term.

Perturbation theory is a very powerful and useful tool. But it cannot always be used because expanding in powers of the coupling constant only makes sense if the coupling constant is small enough. This is not always the case, for example, the strong interaction between quarks in hadrons cannot be described perturbatively. Other (non-perturbative) techniques have to be used in this case.

Perturbation theory suffers from another problem: calculating diagrams involving loops sometimes seems to give an infinite answer. An important tool to deal with this problem is *renormalization*.

2.3.3 Renormalization

Here is a simple example to illustrate how renormalization works. The one-loop correction to the photon propagator (second diagram in the example shown above) gives a correction to the Coulomb potential (e^2/p^2 in momentum space)[3]:

$$V(p^2) = \frac{e^2}{p^2} \left(1 - e^2\Pi(p^2)\right) \tag{2.3}$$

where $\Pi(p^2)$ is a divergent integral. To deal with this problem, we first define the *renormalized charge* $e_R^2 \equiv k^2 V(k^2)$. Contrary to e (the so-called bare charge) e_R can be measured experimentally by measuring the potential for a certain momentum k . Using equation 2.3 the renormalized charge is:

$$e_R^2 = e^2 - e^4 \Pi(k^2) + \mathcal{O}(e^6) \quad (2.4)$$

Reversing it to express e in terms of e_R :

$$e^2 = e_R^2 + e_R^4 \Pi(k^2) + \mathcal{O}(e_R^6) \quad (2.5)$$

Replacing e^2 in equation 2.3 the potential can be expressed in terms of the renormalized charge:

$$V(p^2) = \frac{e_R^2}{p^2} \left(1 - e_R^2 \left(\Pi(p^2) - \Pi(k^2) \right) \right) + \mathcal{O}(e_R^6) \quad (2.6)$$

This is a completely finite expression because the divergent part in $\Pi(p^2)$ does not depend on p^2 so the difference $\Pi(p^2) - \Pi(k^2)$ is finite. The divergent part of the integral corresponds to the integration over large virtual particle momentum. The cancellation after renormalization shows that whatever is happening at high energy can be absorbed in a redefinition (“renormalization”) of the electric charge and otherwise be ignored.³ The upside is that we do not have to know the correct physics at high energy to be able to do physics at lower energy. The Standard Model is often considered to be an *effective field theory*, a low energy approximation of a more fundamental theory[5].

2.3.4 Running coupling constant

Equation (2.6) gives in the large momentum limit[3]:

$$V(p^2) = \frac{e_R^2}{p^2} \left(1 + \frac{e_R^2}{12\pi^2} \ln \left(\frac{-p^2}{m^2} \right) \right) + \mathcal{O}(e_R^6) \quad (2.7)$$

It is easy to calculate how the potential will be affected using all orders of perturbation theory. Each order will have one more loop than the previous one and will bring a new $\frac{e_R^2}{12\pi^2} \ln \left(\frac{-p^2}{m^2} \right)$ factor

³The above procedure is actually quite dubious from a strictly mathematical point of view (see [4] p. 395) but it works very well in practice.

so the result is a geometric series:

$$V(p^2) = \frac{e_R^2}{p^2} \left(1 + \frac{e_R^2}{12\pi^2} \ln \left(\frac{-p^2}{m^2} \right) + \left(\frac{e_R^2}{12\pi^2} \ln \left(\frac{-p^2}{m^2} \right) \right)^2 + \dots \right) \quad (2.8)$$

Defining $Q^2 \equiv -p^2$ (Q is the momentum transfer) and $e^2(Q^2) \equiv -Q^2 V(-Q^2)$:

$$e^2(Q^2) = \frac{e_R^2}{1 - \frac{e_R^2}{12\pi^2} \ln \left(\frac{Q^2}{m^2} \right)} \quad (2.9)$$

$e^2(Q^2)$ is called a running coupling constant, the strength of the electromagnetic interaction increases with energy. The coupling strengths for the other fundamental interactions also change with energy but for the strong interaction, the strength actually decreases with energy. The values of the coupling constants get fairly close to each other at about 10^{16} GeV (see figure) which suggests that there is a unified description of the weak, strong and electromagnetic interactions (and possibly also a more fundamental theory including gravity as well at higher energy).

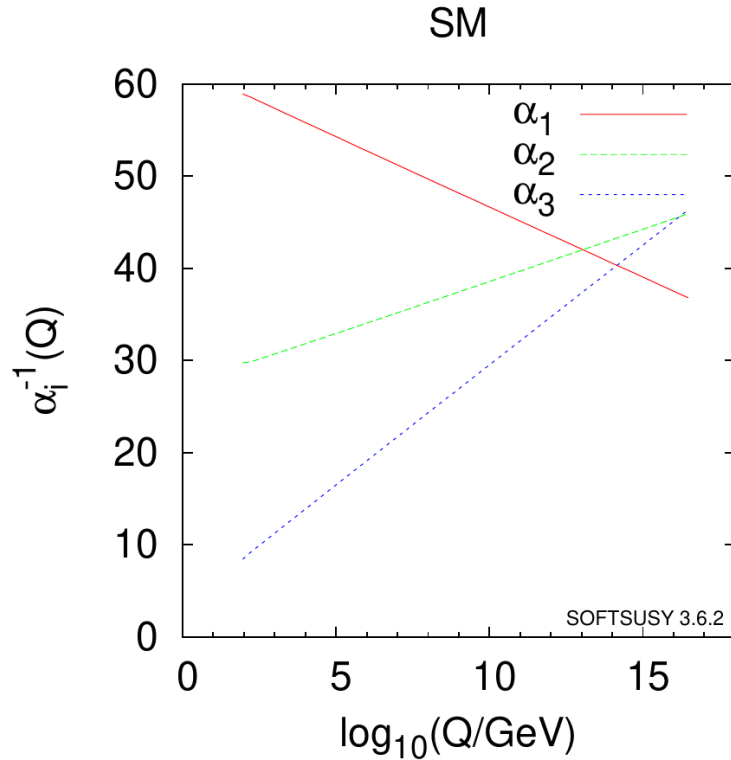


Figure 2.2: The variations of the electromagnetic (α_1), weak (α_2) and strong (α_3) coupling constants with respect to energy are shown here. They get fairly close to each other at around 10^{16} GeV (often called the *grand unification scale*). Illustration taken from [6]

2.3.5 Cross sections

When testing the Standard Model experimentally, collisions between particles are often used, either between two beams or between one beam and a fixed target. It is often useful to measure the *cross-section* for a particular process. The differential cross-section is defined by:

$$d\sigma = \frac{dP}{T\Phi} \quad (2.10)$$

where Φ is the incoming flux, T is the time duration of the experiment and dP is the differential (in kinematic variables) probability of having a particular final state. The total cross-section can be obtained by integrating $\sigma = \int d\sigma$. From the cross-section, one can obtain the event rate:

$$\frac{dN}{dt} = \sigma \mathcal{L} \quad (2.11)$$

where \mathcal{L} is called the *luminosity* and depends on the particle beams.

The differential probability is given by:

$$dP = \frac{|\langle f|S|i\rangle|^2}{\langle f|f\rangle \langle i|i\rangle} d\Pi \quad (2.12)$$

where $d\Pi$ is the differential phase space for the final state and S is a unitary matrix called the *S-matrix* relating the initial state $\langle i|$ and the final state $\langle f|$. If there was no interaction we would have $S = \mathbb{1}$ so we define the transfer matrix T :

$$S = \mathbb{1} + iT \quad (2.13)$$

and the transfer matrix T can be computed from the amplitude \mathcal{M} :

$$T = (2\pi)^4 \delta^4\left(\sum p_i - \sum p_f\right) \mathcal{M} \quad (2.14)$$

where the Dirac delta $\delta^4(\sum p)$ ensures the conservation of energy and momentum. With the S-matrix formalism, we can obtain measurable quantities (like the cross-section) from a more abstract mathematical entity (the quantum mechanical amplitude \mathcal{M}).

2.4 Fundamental interactions

2.4.1 Quantum electrodynamics

Let's start from the Lagrangian for the *Dirac field*:

$$\mathcal{L} = i\bar{\psi}(\gamma^\mu \partial_\mu - m)\psi \quad (2.15)$$

where ψ is a Dirac spinor, a mathematical object represented by four complex numbers, $\bar{\psi} \equiv \psi^\dagger \gamma^0$ and the γ^μ are 4×4 matrices that must obey the Dirac algebra: $\{\gamma^\mu, \gamma^\nu\} \equiv \gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu}$.

We can easily see that it is invariant under the following transformation:

$$\psi \longrightarrow \psi' = e^{i\alpha} \psi \quad (2.16)$$

ψ and ψ' are physically equivalent and this symmetry is called the $U(1)$ symmetry and is an example of a *gauge symmetry*. It is a global symmetry (α is a fixed real parameter) but this symmetry can be made local (replacing α by an arbitrary smooth function $\alpha(x)$) by introducing a new vector field A_μ , and replacing the partial derivative ∂_μ by the *covariant derivative*:

$$D_\mu = \partial_\mu - ieA_\mu \quad (2.17)$$

The trick is then to define the gauge transformation of the A_μ field as:

$$A'_\mu = A_\mu - \frac{1}{e} \partial_\mu \alpha \quad (2.18)$$

with this transformation law, $D_\mu' \psi' = (D_\mu \psi)'$ and gauge invariance is recovered.

Now that there is a new field A_μ a corresponding kinetic term should be added. It is easy to see that $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ is gauge invariant and it can be used to construct a kinetic term. The result is the Lagrangian of quantum electrodynamics:

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi}(\gamma^\mu D_\mu - m)\psi \quad (2.19)$$

We now have an interaction between ψ and A_μ given by $e\gamma^\mu A_\mu \bar{\psi}\psi$ so the theory describes the

interaction between one (or more) spin- $\frac{1}{2}$ particles (ψ) and the photon (A_μ).

There is no mass term for the A_μ field and simply adding one ($\frac{1}{2}m_A^2 A_\mu A^\mu$) would break gauge invariance. The photon indeed appears to be massless, experimentally $m_\gamma < 10^{-18}$ eV [6].

2.4.2 Quantum chromodynamics

Quantum chromodynamics (QCD) is the theory describing the strong interaction. It is a gauge theory based on the $SU(3)$ symmetry, the group of 3×3 unitary matrices with determinant one under multiplication. An element of $SU(3)$ can be written as:

$$\exp\left(i \sum_a \alpha_a \frac{\lambda_a}{2}\right) \quad (2.20)$$

where λ_a are 8 linearly independent matrices, spanning the Lie algebra of $SU(3)$. Their commutations relations define the *structure constant* f^{abc} :

$$[\lambda_a, \lambda_b] = 2i f^{abc} \lambda_c \quad (2.21)$$

and the covariant derivative is given by:

$$D_\mu = \partial_\mu - ig_s \frac{\lambda_a}{2} G_\mu^a \quad (2.22)$$

The QCD Lagrangian is then:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} + \bar{\psi} (i\gamma^\mu D_\mu - m) \psi \quad (2.23)$$

where the field tensor $G_{\mu\nu}^a$ is defined by:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\nu^b G_\mu^c \quad (2.24)$$

The last term is necessary for gauge invariance because the group generators (λ_a) do not commute. It corresponds to interactions between gluons. A similar term for photons does not exist in QED because $U(1)$ is an abelian (i.e. commutative) group.

The QCD charge is called the color charge and both quarks and gluons carry a color charge. Quarks combine together to form colorless hadrons and the description of quarks interactions inside hadrons is very difficult as the QCD is no longer perturbative at low energy because the coupling constant of QCD is larger at lower energy and gets smaller and smaller as energy increases. This also means that quarks and gluons interactions become weaker at higher energy, this phenomenon is called *asymptotic freedom*.

2.4.3 Electroweak interaction

The *electroweak theory* provides a partially [7] unified description of the weak and electromagnetic interactions. It was developed in the late '60s when physicists suspected the existence of massive vectors bosons “carrying” the weak interaction[8]. However, as mentioned earlier simply adding a vector boson mass term in the Lagrangian would break gauge invariance. This problem cannot be simply ignored because the resulting theory would not be renormalizable[3].

One possible way out is to start from a theory in which the gauge symmetry is a symmetry of the equation but not of the solutions. This is called *spontaneous symmetry breaking*. It turns out that the correct theory to describe symmetry breaking for the weak interaction also includes electromagnetism.

The starting point is the following Lagrangian based on the $SU(2) \times U(1)$ symmetry where a new complex doublet ϕ is introduced:

$$\mathcal{L} = -\frac{1}{4}W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + (D_\mu\phi)^\dagger(D_\mu\phi) + \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \quad (2.25)$$

with $W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + gf^{abc}W_\nu^b W_\mu^c$ (with $a = 1, 2, 3$) and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ and the covariant derivative is:

$$D_\mu = \partial_\mu - igW_\mu^a\tau^a - ig'YB_\mu \quad (2.26)$$

where Y is a quantum number called the weak hypercharge and τ^a are the $SU(2)$ generators chosen by convention to be $\tau^a = \sigma^a/2$ where σ^a are the Pauli matrices.

The potential $V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$ has the consequence that ϕ has a non-zero vacuum expectation value $\langle\phi\rangle$ because it is minimized when $\phi^\dagger\phi = \mu^2/(2\lambda)$. The potential is shown (for the

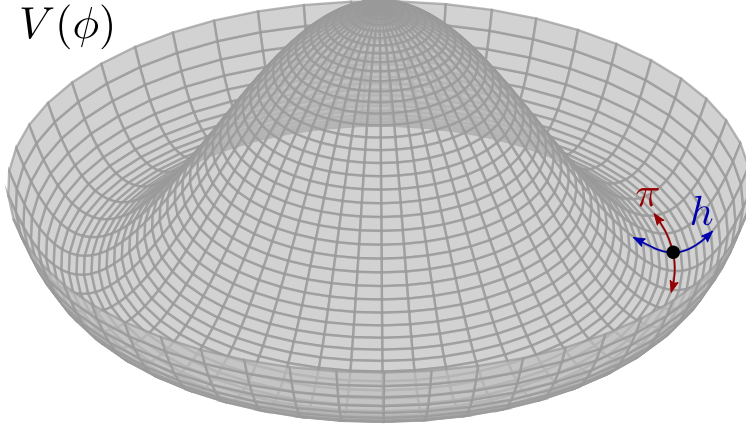


Figure 2.3: Representation of the Higgs potential for the simpler case where ϕ is a complex field instead of a complex doublet. Radial oscillations h correspond to the Higgs boson and angular oscillations π correspond to a Goldstone boson.

simpler case where ϕ is a complex field and not a doublet) in figure 2.3. Around a minimum, ϕ can be decomposed into one field h and three fields π^a . A theorem proved by Goldstone[9] shows that there is a massless particle for every spontaneously broken symmetry and the three π^a (one for each generator of $SU(2)$) are an example of this kind of particle (called *Goldstone bosons*).

By symmetry, we are free to choose any value where the potential is minimized, so we can write:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (2.27)$$

with $v = \mu/\sqrt{\lambda}$ where the gauge symmetry was used to set $\pi = 0$ (this is called the unitary gauge). When plugging ϕ back into the Lagrangian, new mass terms will appear for the gauge bosons. To have the mass eigenstates fields we define:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_w & -\sin \theta_w \\ \sin \theta_w & \cos \theta_w \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \quad (2.28)$$

with $\theta_w \equiv \tan^{-1} \frac{g'}{g}$. Z_μ represents the Z boson and A_μ the photon. The two W bosons are:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2) \quad (2.29)$$

The photon is massless but the W and Z bosons acquire a mass:

$$m_W = \frac{vg}{2} \quad \text{and} \quad m_Z = \frac{m_W}{\cos \theta_w} \quad (2.30)$$

There are four massless vector bosons before symmetry breaking, one massless boson and three massive vector bosons after. Those three massive vector bosons can have a longitudinal polarization (unlike massless vector bosons) so it seems that there are new degrees of freedom. In fact, they come from the three Goldstone bosons, although the unitary gauge “ignores” them, the corresponding degrees of freedom are still there. It is sometimes said that the W and Z bosons have “eaten” the three Goldstone bosons.

There is also a scalar boson h (see equation 2.27) with mass $m_h = \sqrt{2}\mu$, this is the Higgs boson. It is the only fundamental particle with spin-0 in the Standard Model.

After symmetry breaking, we have a new $U(1)$ symmetry associated with electromagnetism which is different from the $U(1)$ symmetry that we started with. The electric charge Q of a given particle is related by its weak hypercharge Y by $Q = Y + T_3$ where T_3 (called the weak isospin) is the eigenvalue of $\tau^3 = \sigma^3/2$ for this particle.

It has been experimentally observed that the weak interaction violates parity[10]. Parity is in fact maximally violated: the W_μ^a bosons (the three $SU(2)$ bosons before symmetry breaking) do not interact with right-handed fermions. Dirac fermions can be split into a “left-handed” and a “right-handed” components $\Psi = \Psi_L + \Psi_R$ where Ψ_L and Ψ_R are given by:

$$\Psi_{R,L} = \frac{1 \pm \gamma^5}{2} \Psi \quad (2.31)$$

where $\gamma_5 \equiv i\gamma_0\gamma_1\gamma_2\gamma_3$. Under a parity transformation $P : (t, \vec{x}) \rightarrow (t, -\vec{x})$ a left-handed spinor becomes a right-handed spinor and a right-handed spinor becomes a left-handed spinor[11]:

$$\begin{aligned} P : \Psi_L(t, \vec{x}) &\rightarrow \Psi_R(t, -\vec{x}) \\ P : \Psi_R(t, \vec{x}) &\rightarrow \Psi_L(t, -\vec{x}) \end{aligned} \quad (2.32)$$

which is why they are called left- and right-handed. The left-handed spinors are grouped into $SU(2)$

doublets. For example, in the first generation the doublets are:

$$E_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L \quad (2.33)$$

and we have similar doublets for the other two generations. The right-handed particles are not affected by the $SU(2)$ symmetry, so they form $SU(2)$ singlets. Using the covariant derivative defined in equation 2.26 the terms involving leptons are:

$$\begin{aligned} \mathcal{L}_{\text{leptons}} &= \bar{E}_L(i\gamma^\mu D_\mu)E_L + \bar{e}_R(i\gamma^\mu D_\mu)e_R \\ &= \bar{E}_L(i\gamma^\mu \partial_\mu - igW_\mu^a \tau^a - ig'Y_L B_\mu)E_L + \bar{e}_R(i\gamma^\mu \partial_\mu - ig'Y_e B_\mu)e_R \end{aligned} \quad (2.34)$$

The right-handed neutrino is not included as there is (so far) no experimental evidence that it exists. For quarks, we have something similar but with two right-handed particles u_R and d_R .

The different coupling between left-handed and right-handed particles creates a problem. A mass term for the fermions in the Lagrangian $-m(\bar{\Psi}_R\Psi_L + \bar{\Psi}_L\Psi_R)$ would break gauge invariance. The solution is simply to postulate an interaction with the Higgs field to have fermions masses. After symmetry breaking, the fermions have a mass $m_f = \lambda_f v/\sqrt{2}$.

Experimental evidence shows that transitions can happen between generations which means that the quark mass and flavor basis do not coincide exactly. By convention, we look at the down-type quarks (d , s and b) and write:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.35)$$

where the left-hand side corresponds to the weak eigenstates and on the right-hand side the mass eigenstates are multiplied by the Cabibbo-Kobayashi-Maskawa matrix[12, 13]. The CKM matrix can also explain the observed violation of CP symmetry in neutral kaon decays[14].

Chapter 3

Vector boson scattering

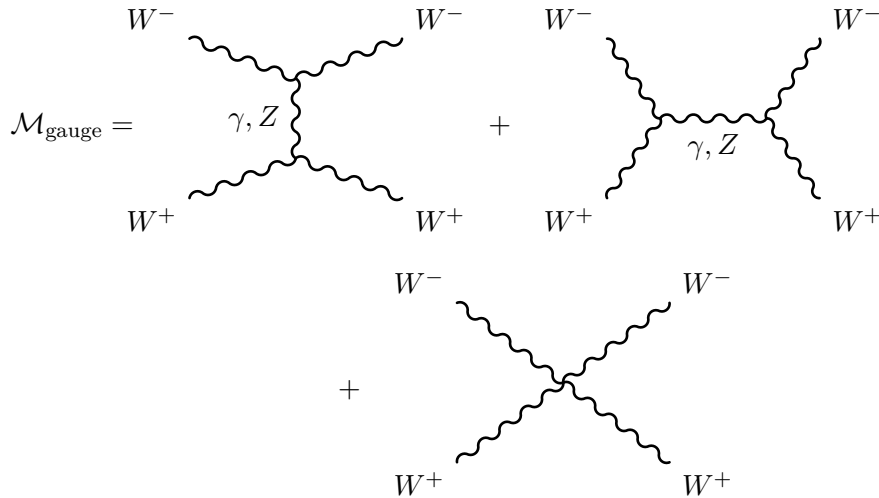
3.1 Unitarity

Unitarity in quantum mechanics is related to the conservation of probability. In particular, for the theory to make any sense, the sum of all the probabilities of the different possible final states (for a given initial state) must obviously be equal to one. This means that the cross-section for a particular process cannot grow arbitrarily large.

Without the Higgs mechanism, the scattering cross-section of two massive vector bosons would grow with energy and violate unitarity. This is due to the longitudinal polarization of the massive bosons which does not exist for massless particles. As an example, consider the scattering of two opposite-sign longitudinally polarized W bosons:

$$W_L^+ W_L^- \longrightarrow W_L^+ W_L^- \quad (3.1)$$

If the Higgs is not included, the Feynman diagrams for the leading order are:



The result is [15]:

$$\mathcal{M}_{\text{gauge}} = \frac{-g^2}{4m_W} u + \mathcal{O}(1) \quad (3.2)$$

where $\mathcal{O}(1)$ includes any unproblematic term (not growing with energy) and $u \equiv (p_1 - p_4)^2 = (p_3 - p_2)^2$

is one of the so-called Mandelstam variables. At high energy $u \approx -2p_1 \cdot p_4$ so $\mathcal{M}_{\text{gauge}}$ grows with energy and unitarity is violated.

To fix the problem, diagrams involving the Higgs boson must be included in the calculation. They are (at leading order):

$$\mathcal{M}_{\text{Higgs}} = \begin{array}{c} W^- \\ \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ W^+ \end{array} \begin{array}{c} \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ W^- \\ \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ W^+ \end{array} \begin{array}{c} h \\ \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ h \end{array} + \begin{array}{c} W^- \\ \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ W^+ \end{array} \begin{array}{c} \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ W^- \\ \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ W^+ \end{array} \begin{array}{c} \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ \text{~~~~~} \\ h \end{array}$$

and we have:

$$\mathcal{M}_{\text{Higgs}} \approx \frac{g^2}{4m_W} u + \mathcal{O}(1) \quad (3.3)$$

The approximation is valid as long as the energy is much larger than the Higgs and W masses. It also means that the Higgs boson cannot be arbitrarily heavy as this cancellation would happen “too late” and unitarity would still be violated. The upper bound for the Higgs boson mass is [16]:

$$m_h^2 \lesssim \frac{8\pi\sqrt{2}}{3G_F} \approx 1\text{TeV} \quad (3.4)$$

It is important to be careful here. If the Higgs boson did not exist, it would not mean (strictly speaking) that unitarity would be violated and that the probabilities of the final states would not sum up to one (a logical impossibility) but that perturbation theory would no longer be reliable at high energy[3]. This could happen (for example) due to a new strong interaction, this is the main idea behind the so-called technicolor models[17].

However, in 2012, a new particle was discovered by the ATLAS[18] and CMS[19] collaborations with properties matching the Standard Model Higgs and with a mass of about 125 GeV. It might be enough to solve the problem of unitarization but in many BSM (Beyond the Standard Model) models the Higgs boson is only partially responsible for unitarization [20, 21].

For example, some models introduce a second Higgs doublet (two-Higgs-doublet models) [22]. There are two CP-even Higgs bosons in these models, one that would be identified with the 125 GeV particle and another heavier one. Their coupling strength with WW would be reduced: $g_{hWW} = \sin(\beta - \alpha)g_{hWW}^{SM}$ and $g_{HWW} = \cos(\beta - \alpha)g_{hWW}^{SM}$ (H is the heavier Higgs) with α being

the mixing angle between the two Higgs bosons and $\tan \beta = \frac{v_1}{v_2}$ is the ratio of the two vacuum expectation values[21]. In this scenario, h only partially unitarizes the vector bosons scattering processes while H completes the unitarization at larger energy ($s > m_H^2$). The WW scattering cross section would therefore grow larger than what is predicted by the Standard Model (at least until $s \approx m_H^2$).

Studying vector boson scattering at the electroweak scale is therefore useful to test the Standard Model and possibly discover hints of new physics.

3.2 Current experimental status

The energy required to have vector boson scattering is quite large and only recently has it become possible to study it experimentally[23]. The Large Hadron Collider (see chapter 5 for more details) is well suited to study vector boson scattering thanks to the large energy and luminosity it provides.

Figure 3.1 shows a recent summary of the results obtained by the ATLAS collaboration for vector boson fusion processes, vector boson scattering and triple boson production. It shows the ratio of the measured cross-section of various processes with the predicted cross-section in the Standard Model. No significant deviation from the Standard Model has been observed. Figure 3.2 shows similar measurements obtained by the CMS collaboration (but not including tri-boson production). Measured cross-sections are also in agreement with Standard Model predictions.

Although no significant deviation from the Standard Model has been observed, figures 3.1 and 3.2 show the measured cross-sections for many of these electroweak tend to be somewhat larger than the theoretical prediction. Future measurements at the LHC will help measure those processes with even greater accuracy and determine whether this is just a statistical effect or a potential sign of new physics.

VBF, VBS, and Triboson Cross Section Measurements Status: February 2022

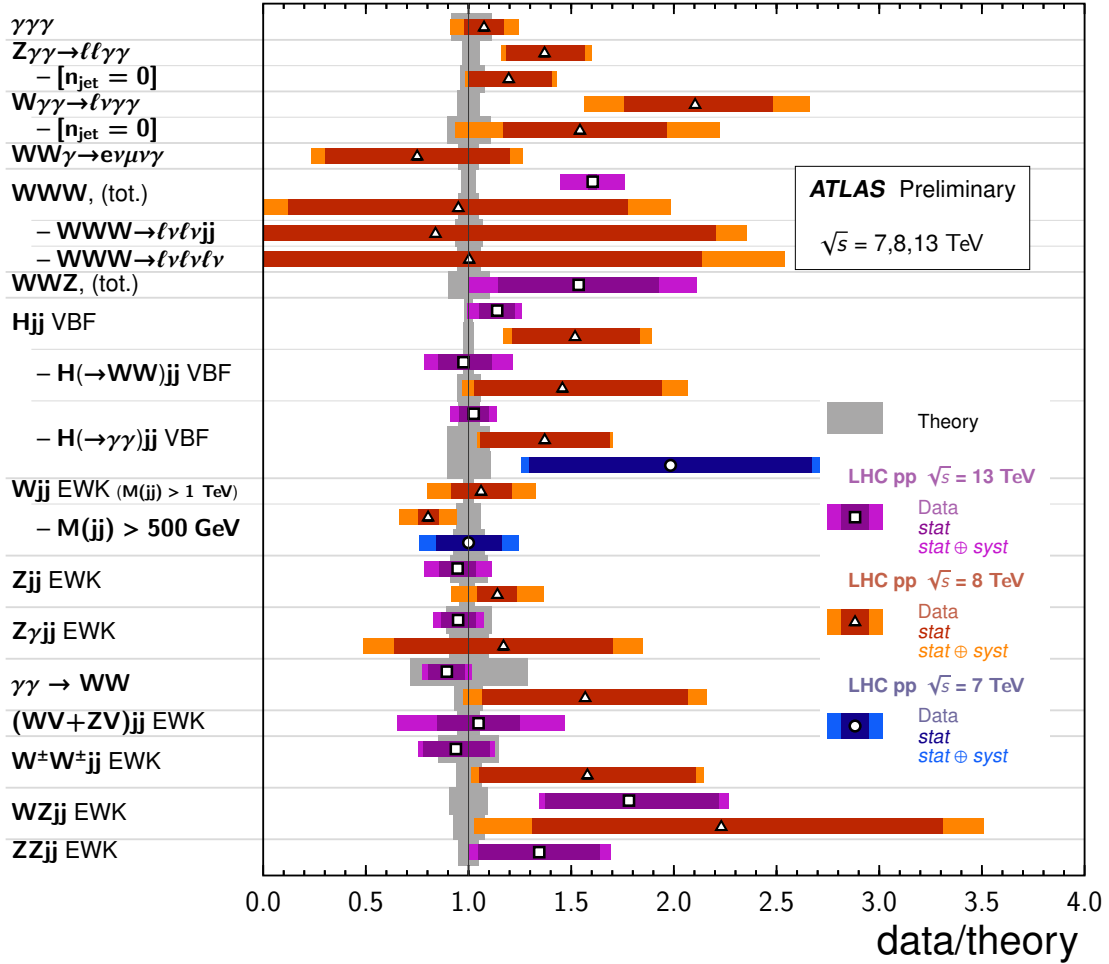


Figure 3.1: VBF, VBS and tri-boson cross-section measurements by the ATLAS collaboration[24]. The ratio $\sigma_{\text{exp}}/\sigma_{\text{theo}}$ are shown with the corresponding uncertainties.

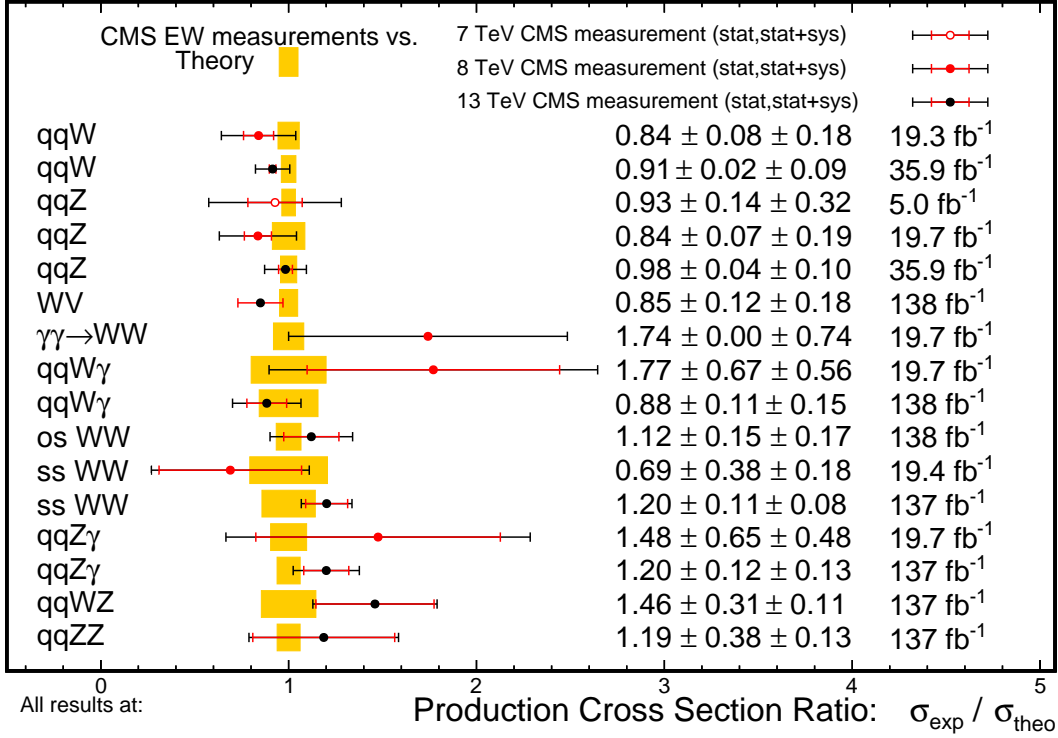
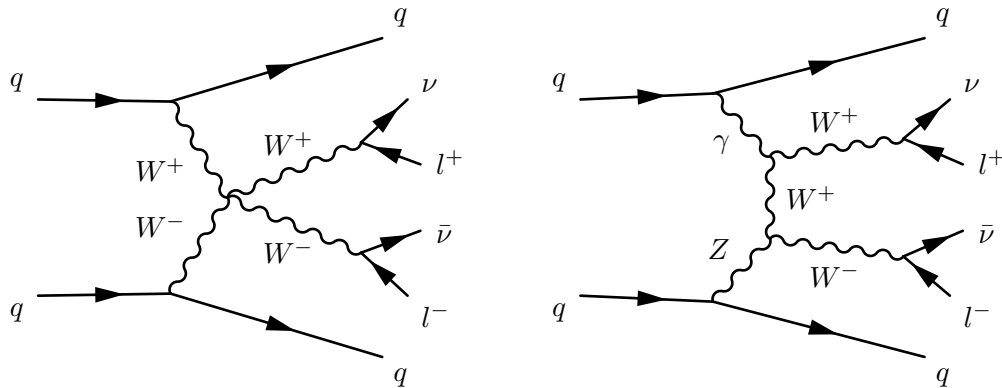


Figure 3.2: Cross section measurement of electroweak production of one or two vector bosons by the CMS collaboration[25]. The ratio $\sigma_{\text{exp}}/\sigma_{\text{theo}}$ are shown with the corresponding uncertainties.

3.3 Electroweak production of W^+W^- with two jets

The LHC is a proton-proton collider so we are looking at processes where vector bosons are radiated from quarks (from the initial protons) as illustrated by the two following examples of Feynman diagrams:

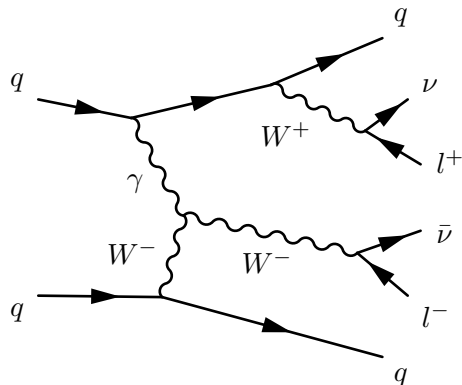


We are interested in the particular case where two leptons and two neutrinos are produced from the decays of the two W bosons. In practice, the neutrinos cannot be detected directly as they interact

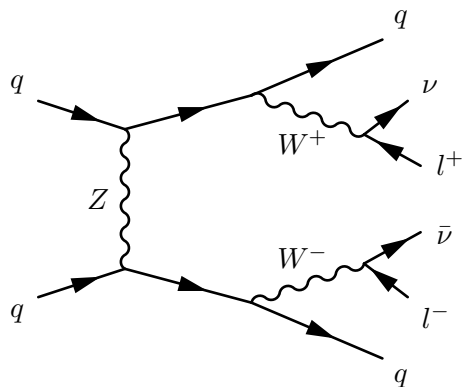
too weakly but their presence can be inferred by the corresponding missing energy (or momentum) in the plane transverse to the beam (the longitudinal momentum of the partons involved is unknown and therefore cannot be used).

Many other Feynman diagrams will also contribute to the same final state but we can split the “electroweak production” from the “QCD” production (but also taking into account the EWK-QCD interference). Electroweak production diagrams (at leading order) are those with a factor of $\alpha_{EW} = 6$ while the QCD production diagrams come with a factor of $\alpha_s = 2$ (and $\alpha_{EW} = 4$).

The electroweak production diagrams include many diagrams involving vector bosons interactions like the two diagrams already shown. They are examples of what is called the VBS (vector boson scattering) topology. There are other diagrams involving vector bosons interactions that do not have the same topology, for example:



Finally, we also have diagrams without interactions between the vector bosons:



When running computer simulations of this kind of process, it is not possible to generate the events corresponding to the VBS topology separately as it would break gauge invariance. When studying vector boson scattering experimentally the signal events are therefore generated using all the Feynman diagrams fitting the electroweak production criteria defined earlier.

The two quarks in the core process are going to form hadronic jets (see chapter 4). The final states as seen by the detector are two (or more) jets, two leptons with opposite charges and missing transverse energy. The two leading jets (i.e. the two jets with the largest transverse momentum p_T) will typically have a fairly large invariant mass m_{jj} and pseudorapidity difference $\Delta\eta_{jj}$ [26], where m_{jj} is defined as:

$$m_{jj} = \sqrt{(E_1 + E_2)^2 - |\vec{p}_1 + \vec{p}_2|^2} \quad (3.5)$$

with E_1 and E_2 being the energy of the two leading jets and \vec{p}_1 and \vec{p}_2 their momentum. $\Delta\eta_{jj}$ is the difference of the pseudorapidity η of the two leading jets defined by:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (3.6)$$

where θ is the angle between the beam axis and the vertical axis at the collision point (see 5 for more details on the coordinate system in ATLAS).

Chapter 4

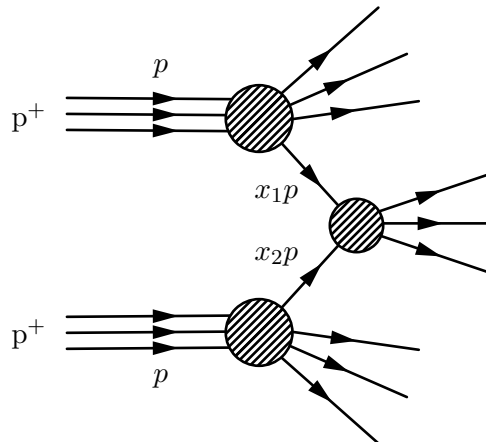
Event simulation for pp collision

4.1 Introduction

Vector boson scattering can be studied in particle accelerators like the Large Hadron Collider (see chapter 5). At the LHC, collisions of two beams of protons are produced at large center-of-mass energy. The results of these collisions are recorded with different detectors and a comparison is made with theoretical expectations. This means that accurate simulations are needed for proton-proton collisions but this presents a difficulty. The internal structure of protons cannot be described perturbatively, because the QCD coupling constant is too large for perturbation theory at the energy scale of quarks interactions inside protons. This chapter will explain how this problem is solved and give an overview of the phenomenology of proton-proton collisions simulation.

4.2 Partons distribution functions and factorization

In a proton-proton collision at high energy, we are often interested in processes with large momentum transfer, that is when there is an interaction between two partons (quarks or gluons) each carrying a significant fraction of the corresponding proton momentum. In the center of mass of the collision, the two incoming protons have the same momentum p so the momentum of the two interacting partons can be written x_1p and x_2p with $0 < x_i < 1$ (x_i are the *momentum fractions*) as illustrated with the following diagram:



The incoming protons are each represented by three lines corresponding to the three quarks. Two of the partons interact (this is the “hard” scattering process) and some particles emerge from the interaction. The partons not involved in the hard scattering interact and produce new hadrons with lower energy. We can factorize out the hard scattering cross-section from the overall cross-section, by virtue of the *factorization theorem* [27, 28]:

$$\sigma = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b \int f_a^1(x_a, \mu_F) f_b^2(x_b, \mu_F) d\sigma_{ab \rightarrow n}(\mu_F, \mu_R) \quad (4.1)$$

where $d\sigma_{ab \rightarrow n}(\mu_F, \mu_R)$ is the differential cross-section of the hard process and $f_k^i(x_k, \mu_F)$ is called a parton distribution function (PDF) and gives the probability of finding the parton k with momentum fraction x_k in the initial hadron i . μ_R is the renormalization scale and μ_F is what is called the factorization scale. The factorization scale separates low energy physics, where perturbation theory is not possible for QCD, from higher energy physics, where perturbation theory can be used.

Ideally, the cross-section should not depend on the factorization and renormalization scales chosen but it does in practice because we cannot compute to all orders of α_s when using perturbation theory. The associated uncertainties are often estimated by varying the factorization and renormalization scales by a factor of two.

The parton distribution function (PDF) for a proton includes the three quarks making up the proton (called *valence quarks*) but also gluons as well as antiquarks and heavier quarks from the second or third generation, called *sea quarks*. It varies with the energy of collision: the probability of having a gluon or a quark from the sea as an incoming parton for the hard process increases with energy as illustrated in figure 4.1 where the parton distribution function is shown at two different energy scales. The qualitative difference between valence and sea quarks can also be seen.

Various PDFs exist, created by different collaborations. These collaborations provide a whole set of PDFs, including a nominal PDF and many other PDFs to take into account the PDF uncertainties. These uncertainties are the experimental uncertainties (for the data used for the PDF fit), various theoretical uncertainties and parametrization uncertainties, missing higher order terms and uncertainties on the strong coupling constant α_S .

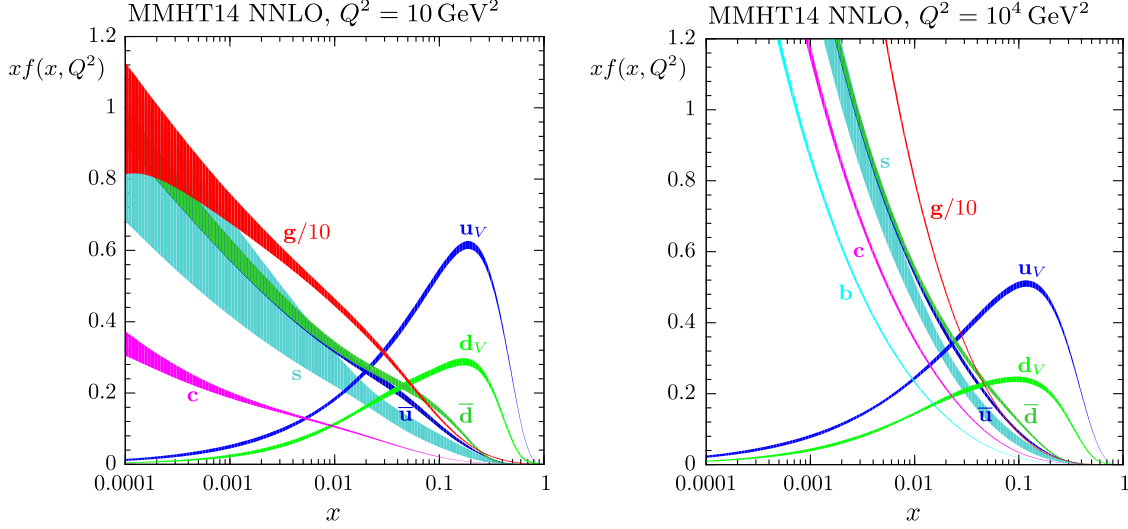


Figure 4.1: Plots of the parton distribution function multiplied by the momentum fraction x for $Q^2 = 10\text{GeV}^2$ (left) and $Q^2 = 10^4\text{GeV}^2$ (right) for different partons. u_V and d_V are the valence quarks and it can be observed that the gluons and the sea quarks contribute the most at smaller momentum fractions and that they contribute more at higher energy. Plots taken from [29].

4.3 Hadronic jets

When there is a quark or a gluon in the final state of the hard process, an *hadronic jet* is created. A hadronic jet is a collection of particles (mostly pions) forming a cone-like structure. It is produced because a high-energy parton radiate gluons that themselves create other gluons as well as quark-antiquarks pairs leading to a cascade of (anti)quarks and gluons. This is called a *parton shower*.

Every time a new parton is radiated, we can compute the differential cross-section $d\sigma_{n+1}$ from $d\sigma_n$ (going from n to $n + 1$ partons) with the equation [30]:

$$d\sigma_{n+1} \approx d\sigma_n \frac{\alpha_S}{2\pi} \frac{d\theta^2}{\theta^2} dz d\phi P_{ji}(z, \phi) \quad (4.2)$$

where α_S is the strong coupling constant, θ the opening angle between the emitted parton and the initial parton, z the energy fraction carried by the emitted parton, ϕ the azimuthal angle, and P_{ij}

are the *splitting functions* [28]:

$$P_{qq}(z) = C_F \frac{1+z^2}{1-z} \quad (4.3)$$

$$P_{gq}(z) = C_F \frac{1+(1-z)^2}{z} \quad (4.4)$$

$$P_{gg}(z) = C_A \frac{z^4+1+(1-z)^4}{z(1-z)} \quad (4.5)$$

$$P_{qg}(z) = T_R(z^2+(1-z)^2) \quad (4.6)$$

where $C_F = \frac{N_c^2-1}{2N_c}$, $C_A = N_c$ (N_c is the number of colors) and $T_R = \frac{1}{2}$ (a color factor fixed by convention, a different convention can be used by using a different definition of α_S). The splitting functions are shown here averaged over ϕ . P_{ji} corresponds to the $i \rightarrow j k$ splitting, for example: P_{gq} corresponds to $q \rightarrow gq$. It is also important to note that equation 4.2 is an approximation, where only the collinear, soft contributions are included.

A simulation of a parton shower proceeds iteratively, starting from an initial parton and generating the additional partons using equation 4.2. As the parton shower grows, the constituent particles become less energetic and the process of *hadronization* eventually starts. Because of QCD confinement at lower energy, quarks and gluons combine to form hadrons which are the jets that can be seen in a particle detector. Although the original parton cannot be seen directly, its energy and momentum are very close to the total energy and momentum of the jet. An example of a jet as seen by the ATLAS detector is shown in figure 4.2.

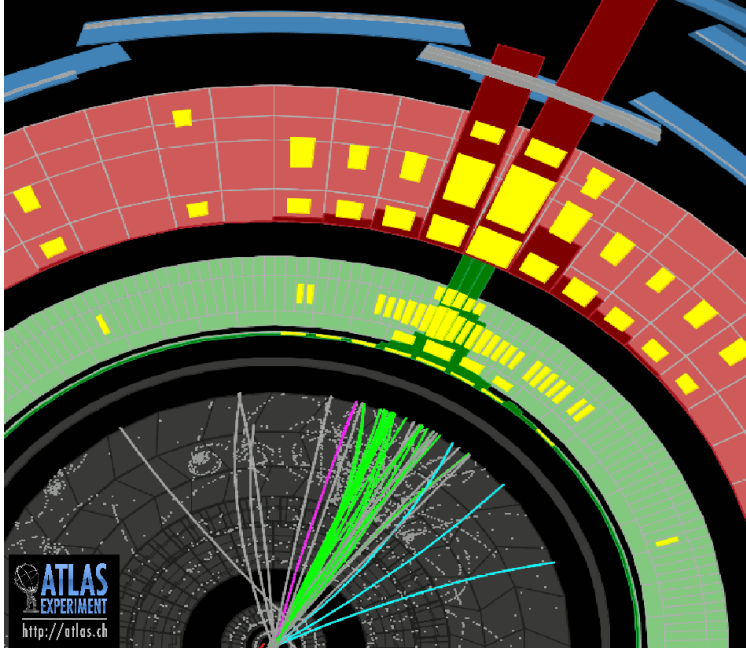


Figure 4.2: Jet seen by the ATLAS detector. The collision point is at the bottom of the picture and a bunch of collimated particles can be seen, forming the jet. Image taken from [31].

There are different models for hadronization simulation. One popular approach is the Lund string model [32]. The model is based on the hypothesis that the QCD field lines between quarks form a string-like structure due to the gluon field self-interaction. This hypothesis is supported by lattice QCD computation (see figure 4.3) and agreement with data when using this model. With enough energy, the strings can be stretched and broken to produce two strings, with a quark-antiquark pair created at the breaking point. This process is simulated iteratively until there is no splitting anymore.

Another popular model is cluster hadronization [33]. It is based on the concept of pre-confinement [34]: partons in a hadron shower tend to cluster into colorless groups, with an invariant mass distribution that depends only on the energy scale and the QCD scale. Those clusters (called *proto-hadrons*) are identified and are then decayed into the final-state hadrons.

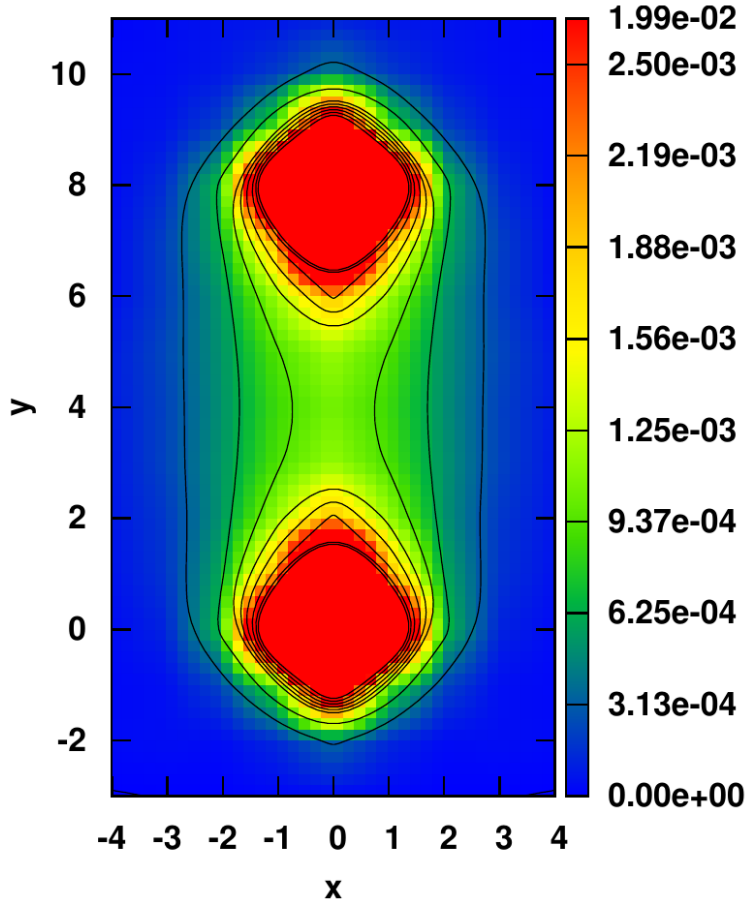
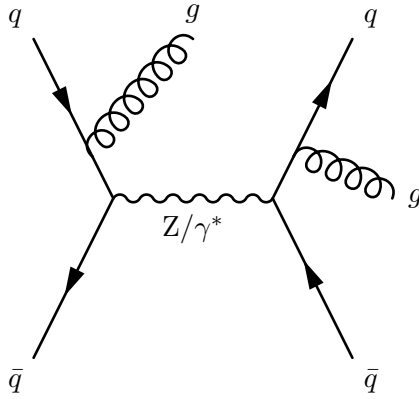


Figure 4.3: Representation of the gluon field for a quark-antiquark pair using lattice QCD computation. A string-like structure can be seen (often called a flux-tube). The colors show the expectation value of the square of the chromoelectric field: $\langle E^2 \rangle$. Illustration taken from [35].

4.3.1 Initial and final state radiation

It is useful to separate what is computed at fixed order of perturbation theory using Feynman diagrams, called the *core process* from the parton shower and hadronization computation. Hadronic jets can be produced by final state partons in the core process, but they can also be produced by additional partons emissions either from an initial or final state parton in the core process. They correspond to *initial and final state radiations* as illustrated with the following diagram:



where additional gluon emissions to the core process are shown. Those additional partons can give rise to hadronic parton showers.

These gluon emissions can be computed using fixed-order perturbative QCD and be considered to be part of the core process. It would not be a good idea to always trust a parton shower simulation software to compute all gluon emissions. This is because parton shower simulations rely on the soft and collinear approximation (as explained before), so hard, wide-angle gluons emissions and parton splittings are better handled by using a perturbative computation. This is important if the precision needed is NLO (next-to-leading order) or better in QCD, where real emissions need to be included in the core process.

4.3.2 Matching and merging

This creates a problem: how to make sure that there is no double counting between the perturbative QCD computation and the parton shower? Two techniques are often used: *matching* and *merging*.

The idea behind matching is to subtract the parton shower contribution at a fixed order. This results in a modified fixed-order computation, which is then used to simulate the parton shower. Because the extra contribution was subtracted, there is no double counting.

For merging, the phase space is divided into a soft and a hard region (defined by the *merging scale*) to exploit the fact that parton shower algorithms are more accurate for soft/collinear parton splitting/gluon emission. Fixed order computations are done for each relevant parton multiplicity (i.e. number of parton in the final state before showering) in the hard region. The parton shower is then run, but if an extra jet is produced above the merging scale, the event is rejected in order to avoid double counting.

4.4 Event generation

4.4.1 Monte Carlo integration

The computation of the cross-section (equation 4.1) involves an integration over the phase space that can be computationally very intensive. This kind of integration can be done numerically using a *Monte Carlo method* [36] which can be very useful to compute multidimensional integrals. We want to solve an integral over some region $\Omega \in \mathbb{R}^n$:

$$I = \int_{\Omega} f(\mathbf{x}) d\mathbf{x} \quad (4.7)$$

The first step is to uniformly generate N sample points $\mathbf{x}_i \in \Omega$. The integral is approximated by [37]:

$$I \approx \frac{V}{N} \sum_{i=1}^N f(\mathbf{x}_i) \quad (4.8)$$

where $V = \int_{\Omega} d\mathbf{x}$ is the n -dimensional volume of Ω . This approximation converges toward the true value I (as N grows toward infinity) by the law of large numbers. The expected error is:

$$\delta I = \frac{V\sigma(f)}{\sqrt{N}} \quad (4.9)$$

where

$$\sigma^2(f) = \int_{\Omega} (f(\mathbf{x}) - I)^2 d\mathbf{x} \quad (4.10)$$

The fact that the error grows as $1/\sqrt{N}$ and does not depend on the number of dimensions (unlike other numerical integration algorithms like Simpson's rule), makes Monte Carlo integration particularly useful for multidimensional integrals.

Several techniques exist that can be used to further reduce the error. One approach that is often used is *importance sampling* [38]. The idea is to replace the uniform distribution over Ω with another (non-uniform) distribution $p(\mathbf{x})$. We can also restrict ourselves to normalized distributions: $\int_{\Omega} p(\mathbf{x}) d\mathbf{x} = 1$. Equation 4.8 is then replaced by:

$$I \approx \frac{1}{N} \sum_{i=1}^N \frac{f(\mathbf{x}_i)}{p(\mathbf{x}_i)} \quad (4.11)$$

where the sampling points \mathbf{x}_i have been chosen using the probability distribution $p(\mathbf{x})$. Note that if we set $p(\mathbf{x}) = 1/V$ (the uniform distribution) we recover equation 4.8.

The expected error is:

$$\delta I = \frac{\sigma(f/p)}{\sqrt{N}} \quad (4.12)$$

If $p(\mathbf{x})$ is proportional to $f(\mathbf{x})$, the expected error is zero. But this would mean (to satisfy the normality condition $\int_{\Omega} p(\mathbf{x})d\mathbf{x} = 1$) that $p(\mathbf{x}) = f(\mathbf{x})/I$ and I is what we want to compute. On the other hand, if we can find a normalized function $p(\mathbf{x})$ that has a shape similar to $f(\mathbf{x})$, this means that the expected error can be reduced significantly. This is the idea behind adaptative Monte Carlo methods like the VEGAS algorithm where an approximation of the ideal probability function (i.e. $f(\mathbf{x})/I$) is created iteratively [37, 39].

4.4.2 Monte Carlo event generators and parton shower software

The simulation of proton-proton collision is done by computing the cross-section, using Monte Carlo techniques for integration. The result of this integration is then used to generate events. Discrete events need to be generated (as opposed to smooth distributions) because the interaction between the particles and the detector, as well as the detector response, need to be simulated to be able to make comparisons with real data. This is why programs used to perform this kind of simulation are called *Monte Carlo event generators*.

Monte Carlo event generations can be summarized in a few steps:

1. The core process is simulated, using the parton distribution functions and the factorization theorem.
2. The parton shower is simulated, using matching and/or merging if necessary.
3. The hadronization and subsequent hadronic decays are then simulated.
4. Partons that are not involved in the hard process interact with each other and can give rise to additional jets, this phenomenon is called *multiparton interactions* (MPI) and also needs to be simulated.

Figure 4.4 shows an illustration of proton-proton collision.

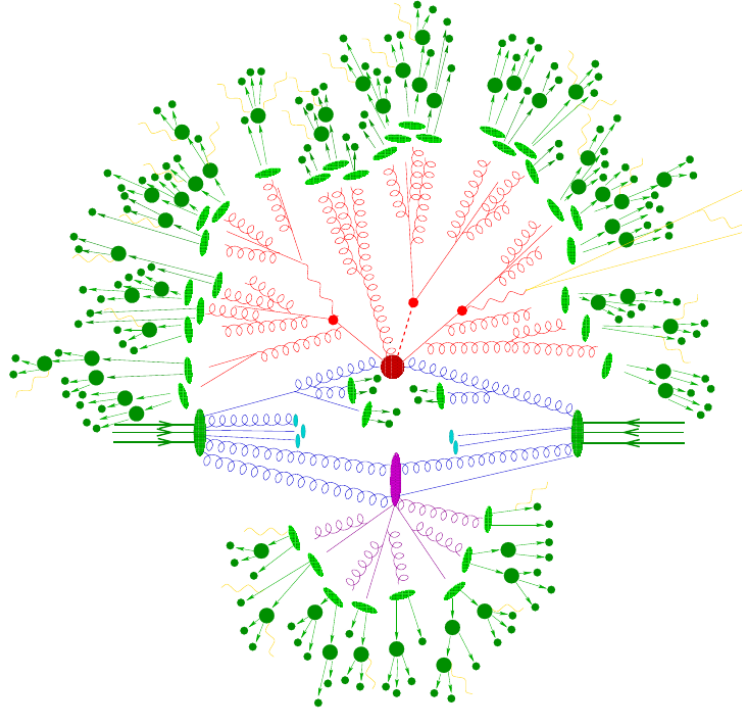


Figure 4.4: Illustration of proton-proton collision. The red blob represents the hard process. From the hard process, several particles emerge (in red). Some of them create parton showers that produce hadrons (hadronization is represented by light green blobs) and many of these hadrons subsequently decay. Multiparton interactions are also represented at the bottom (purple blob). Illustration taken from [40]

Various software exists for Monte Carlo events production. In general, different programs are used for the simulation of the core process and the parton shower and hadronization simulation. Here is a short review of some of the most popular Monte Carlo event generators:

- MADGRAPH5_aMC@NLO [41] can produce events at leading or next-to-leading order. Processes are defined by the user, specifying the initial state (particles in the initial state, center-of-mass energy, ...), the particles in the final state and the electroweak and/or QCD order required. The process can be defined either within the Standard Model or within a predefined BSM model. Processes can also be simulated, at leading order, for user-defined Lagrangians. The parton shower and hadronization can be performed by a separate program.
- SHERPA [42] (Simulation of High-Energy Reactions of PArticles) is another multi-purpose Monte Carlo generator. Processes are defined by the user (similarly to MADGRAPH5_aMC@NLO) either within the Standard Model or within a predefined BSM model. There is a special emphasis on multijet merging and SHERPA can also perform the parton shower (and hadronization)

without the need for an external tool.

- POWHEGBOX [43] is a Monte Carlo event generator implementing the POWHEG (Positive Weight Hardest Emission Generator) method [44, 45] for matching NLO QCD computations with parton shower simulation. Many pre-defined processes exist for NLO QCD, as well as a few NNLO processes using a method to match NNLO QCD to parton showers called MINNLO_{PS} [46]. It can be interfaced with various parton shower programs.
- PYTHIA [47] is yet another event generator with a strong focus on parton shower and hadronization. The Lund model is used for hadronization. Some internally defined processes can be used, but PYTHIA is most often used in conjunction with an external matrix element generator such as MADGRAPH5_aMC@NLO or POWHEGBOX.
- Herwig [48, 49] is an event generator with, like PYTHIA, an emphasis on parton shower and hadronization. The cluster hadronization model is used for hadronization. Although there is a built-in matrix elements generator, it is often used (again like PYTHIA) as a parton shower/hadronization program, using inputs from e.g. MADGRAPH5_aMC@NLO or POWHEGBOX.

Chapter 5

ATLAS and the LHC

5.1 The Large Hadron Collider

5.1.1 Introduction

The Large Hadron Collider (LHC) [50] is the most powerful particle collider ever built. The LHC is operated by the European Council for Nuclear Research, more commonly known as CERN (*Conseil européen pour la recherche nucléaire*). For the past few decades, the focus of research at CERN has been mostly high-energy particle physics rather than nuclear physics, which is why the laboratory is sometimes referred to as the *European Laboratory for Particle Physics* [51].

The LHC was designed to create proton-proton collisions with a center-of-mass energy of up to 14 TeV and a peak luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. It was also designed so that it can be used as a heavy ion collider, by colliding $^{208}\text{Pb}^{82+}$ with a total center-of-mass energy of 1.15 PeV, however, the average center-of-mass energy per colliding nucleon pair is 5.52 TeV/nucleon [52], and a luminosity of $10^{27}\text{cm}^{-2}\text{s}^{-1}$.

The LHC is a circular collider with a circumference of about 27 km. It was built near Geneva, at the border between France and Switzerland, and 100 m underground. Two beams of protons circulate in opposite directions inside a vacuum tube and cross at four different locations. There are four large detectors, one for each collision point, as well as a few smaller ones. Figure 5.1 shows an illustration of the LHC tunnel as well as the locations of the four main experiments.

5.1.2 The detectors

The two largest detectors are ATLAS (A Toroidal LHC Apparatus) [54] and CMS (Compact Muon Solenoid) [55]. They are both general-purpose detectors and are designed for precision tests of the Standard Model and potential discoveries of new BSM phenomena and new particles. More information about ATLAS will be given in section 5.2. The other two main experiments are LHCb (Large Hadron Collider beauty experiment) [56] which is an experiment dedicated to heavy flavor physics, looking for CP violations and rare decays of b - and c -quarks; and ALICE (A Large Ion

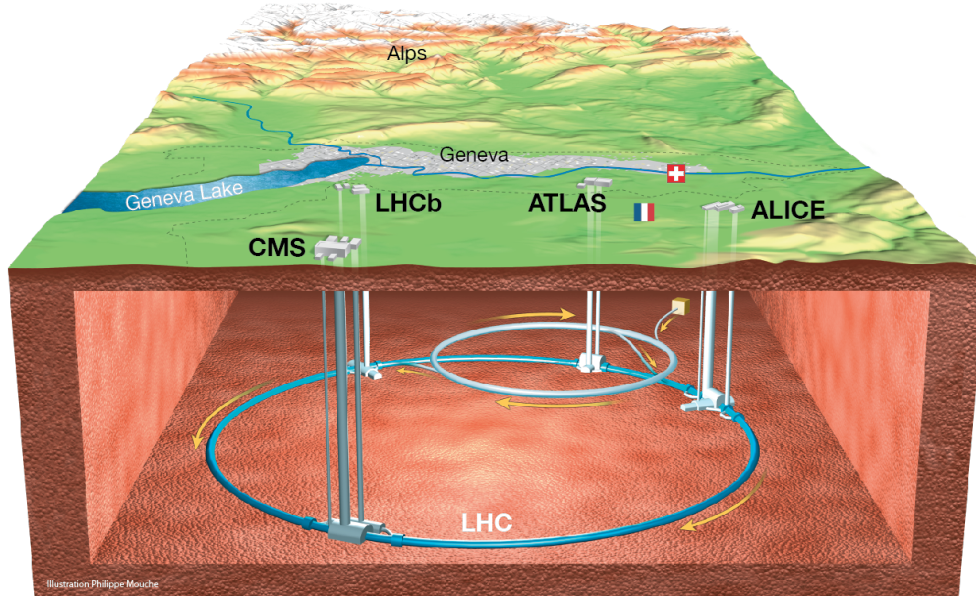


Figure 5.1: Location of the LHC and the four main experiments near Geneva [53].

Collider Experiment) [57] an experiment dedicated to the study of the strong interaction using data from heavy ions collisions.

There are other smaller experiments: LHCf (LHC-forward) [58] is constituted of two small detectors on each side of ATLAS, 140 m away from the collision point. It is designed to study particles produced at small angles in order to improve cosmic rays models. TOTEM (TOTAl Elastic and diffractive cross section Measurement) [59] is made of detectors of both sides of CMS, 500 m away from the collision point. It is designed to detect protons coming from the collision at very small angles. MoEDAL (Monopole and Exotics Detector at the LHC) [60] is located near LHCb and is designed to detect and study monopoles (hypothetical particles carrying a magnetic charge). Finally, FASER (ForwArd Search ExpeRiment) [61] was recently installed 480 m from the ATLAS collision point and was designed to detect and study neutrinos as well as other weakly interacting particles (e.g. potential dark matter particles).

5.1.3 The accelerator complex

The protons in the LHC are accelerated in several steps. The first step is to ionize hydrogen atoms and accelerate the resulting protons. The protons are accelerated to an energy of 50 MeV by Linac2,

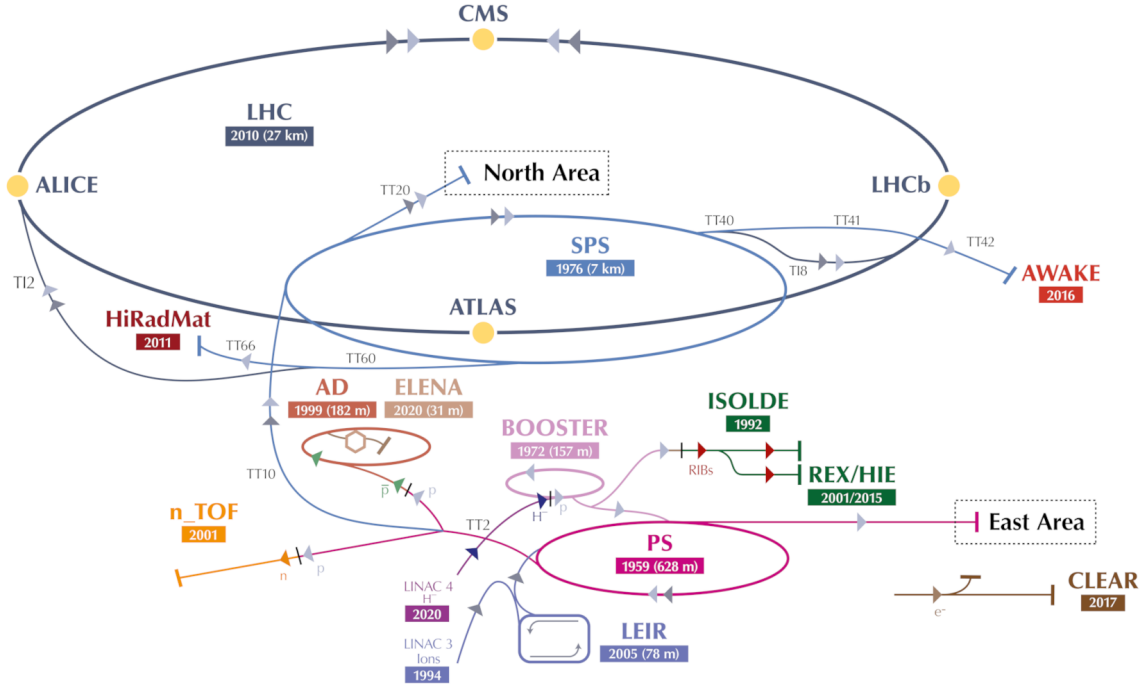


Figure 5.2: Illustration of the LHC accelerator complex. Illustration taken from [63].

until recently when Linac2 was replaced by Linac4, accelerating protons¹ to an energy of 160 MeV [60]. The protons are then injected into the booster where their energy is increased to 1.4 GeV (2 GeV with Linac4). The next step is the Proton Synchrotron (PS), where the protons are further accelerated to an energy of 25 GeV. The proton beams are then injected into the Super Proton Synchrotron (SPS) and the energy is increased to 450 GeV. Finally, the protons are injected into the LHC and are accelerated to 6.5 TeV, giving rise to collisions with an energy of 13 TeV.

The two beams are made out of ‘bunches’ of protons that are spaced by 25 ns [62]. When the bunches cross at a collision point, there are several proton-proton interactions (up to about 70).

The LHC accelerator complex is shown in figure 5.2.

5.1.4 LHC run

The LHC was running in 2010 and 2011 at a center-of-mass energy of 7 TeV, and in 2012 at a center-of-mass energy of 8 TeV. This was the so-called *run 1* of the LHC. The collider was then shut down for a few years, this was the first *long shutdown* (LS1) of the LHC, and the collider and

¹More precisely, Linac4 accelerates negative hydrogen ions H^- , the two electrons are removed when injected to the booster.

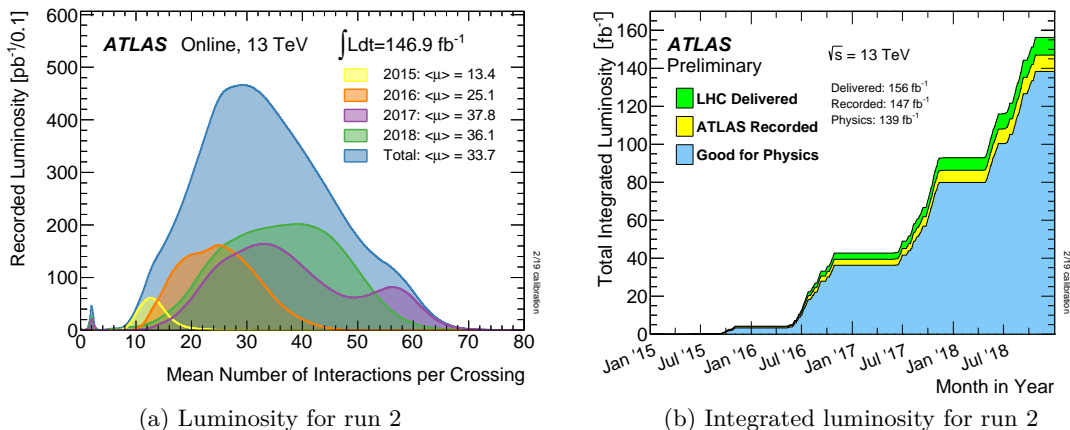


Figure 5.3: Average number of interactions per bunch crossing (left) and integrated luminosity (right) for run 2. [64]

detectors were upgraded during that time. The LHC started again in 2015 with higher center-of-mass energy (13 TeV) and higher luminosity. This was run 2 which stopped in 2018. There was another long shutdown until 2022 when the LHC restarted with a center-of-mass energy of 13.6 TeV (run 3). Between long shutdowns, the collider runs almost every day except during winter when the collider is shut down for a few months.

The total amount of data generated by the collider can be estimated by integrating the luminosity $L = \int \mathcal{L} dt$. Figure 5.3 shows the average number of interactions per bunch crossing and integrated luminosity for ATLAS during run 2. The fact that there are so many pp interactions per bunch crossing means that there is a significant background to interesting events for physics analysis. This background is called *pile-up* and makes reconstruction and calibration more difficult.

5.2 The ATLAS detector

ATLAS is 44 m long, 25 m tall and weighs about 7000 tons. It consists of several layers of detectors surrounding the collision point. Figure 5.4 shows an illustration of the detector.

5.2.1 The ATLAS coordinate system

The ATLAS coordinate system is centered at the point of collision. The z axis is tangent to the beamline and points toward Geneva and defines the A-side ($z > 0$) and the C-side ($z < 0$) of the detector. The vertical axis is the y axis, and the x axis is horizontal and points toward the center of

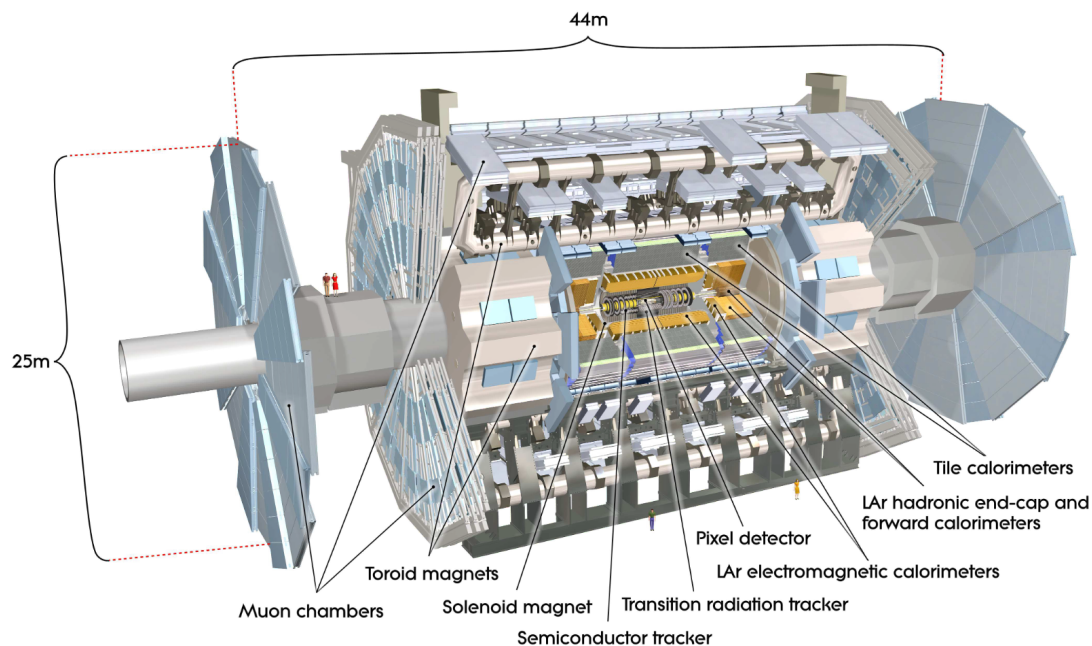


Figure 5.4: The ATLAS detector [54].

the collider.

Due to the cylindrical symmetry of the detector, it is often useful to use cylindrical coordinates rather than cartesian coordinates. The z axis is the same as defined before, the azimuthal coordinate ϕ is the angle with the positive side of the x axis in the $x - y$ plane and the polar angle θ is the angle with the positive side of the z axis. Figure 5.5 shows the ATLAS coordinate system.

The *pseudorapidity* η defined by:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (5.1)$$

is often used instead of θ . The upside is that when expressed in terms of the momentum p of a particle:

$$\eta = -\frac{1}{2} \ln \left[\tan^2 \left(\frac{\theta}{2} \right) \right] = \frac{1}{2} \ln \left[\frac{1 + \cos \theta}{1 - \cos \theta} \right] = \frac{1}{2} \ln \left[\frac{p + p_z}{p - p_z} \right] \quad (5.2)$$

the pseudorapidity η is a good approximation of the *rapidity* y defined by:

$$y = \frac{1}{2} \ln \left[\frac{E + p_z}{E - p_z} \right] \quad (5.3)$$

for large momentum, i.e. when $p \gg m$. What makes rapidity interesting to use is that the rapidity

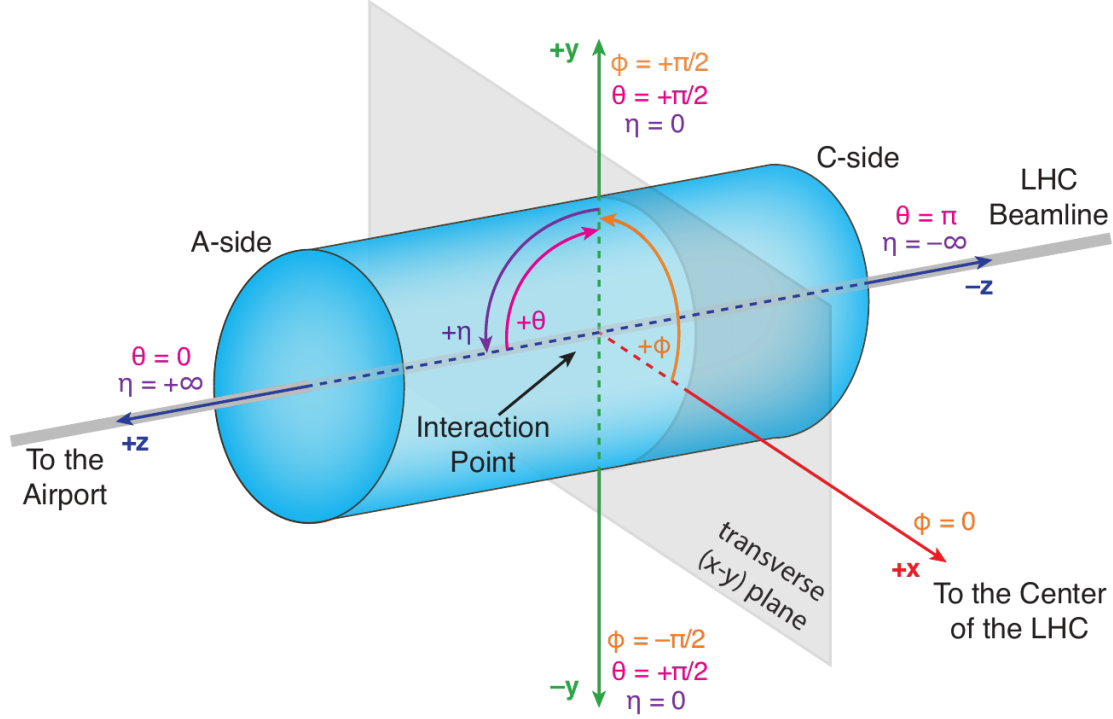


Figure 5.5: The ATLAS coordinate system. Illustration taken from [65].

difference Δy is Lorentz invariant under a boost along the z axis. The ϕ difference $\Delta\phi$ between two particles is also Lorentz invariant under a boost along the z axis, so it is useful to define the angular separation between two particles (or jets) ΔR :

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (5.4)$$

It is often useful to look at the projection of the momentum vector in the transverse ($x - y$) plane, called the *transverse momentum* with norm $|\vec{p}_T| = p \sin \theta = \sqrt{p_x^2 + p_y^2}$. The transverse energy can also be defined: $E_T = \sqrt{p_T^2 + m^2}$. The total transverse momentum before the collision is zero by symmetry, therefore by the law of conservation of momentum, the *missing transverse momentum* defined by:

$$\vec{p}_T^{\text{miss}} = - \sum_{\text{visible particles}} \vec{p}_T \quad (5.5)$$

provides an estimation of the total transverse momentum of invisible particles, i.e. particles that cannot be detected, like neutrinos or (potentially) dark matter particles.

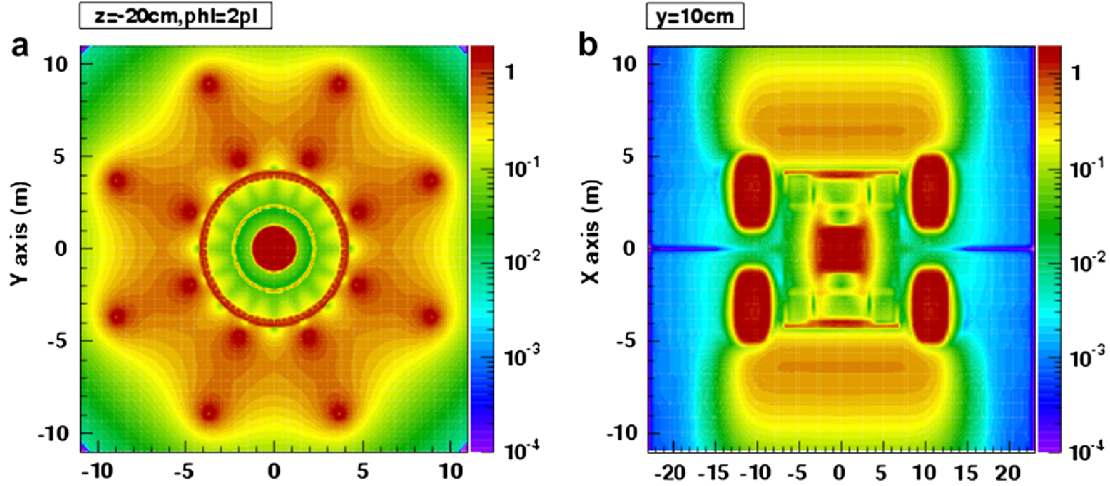


Figure 5.6: ATLAS magnetic field in the transverse section (left) and the longitudinal section (right). Illustration from [66]

5.2.2 The ATLAS superconducting magnet system

An intense magnetic field is produced inside ATLAS by a central solenoid and three toroids: one surrounding the central solenoid called the barrel toroid, and two others on each side, the end cap toroids. The ATLAS magnet system provides a magnetic field of about 2T in the inner detector (see section 5.3) and about 1T in the muon spectrometer (see section 5.5). It has a length of 26 m and a diameter of 22 m, and stores an energy of about 1.6 GJ [66]. Figure 5.6 shows the strength of the magnetic field in different regions of the detector.

With the strong magnetic field generated by the superconducting magnet system, the momentum and charge of the particles can be deduced, as charged particles in a magnetic field follow a curved trajectory. The trajectory is given by solving the equation of motion $F = q\vec{v} \times \vec{B}$ (Lorentz force law).

5.3 The ATLAS inner detector

The inner detector is the innermost part of the ATLAS detector. It is designed to track charged particles with great precision, and measure their momentum. It is made out of three components: the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). Figure 5.7 shows a 3D illustration of the inner detector and figure 5.8 shows a schematic view.

There are three different regions in the inner detector: a barrel region where the sub-detectors

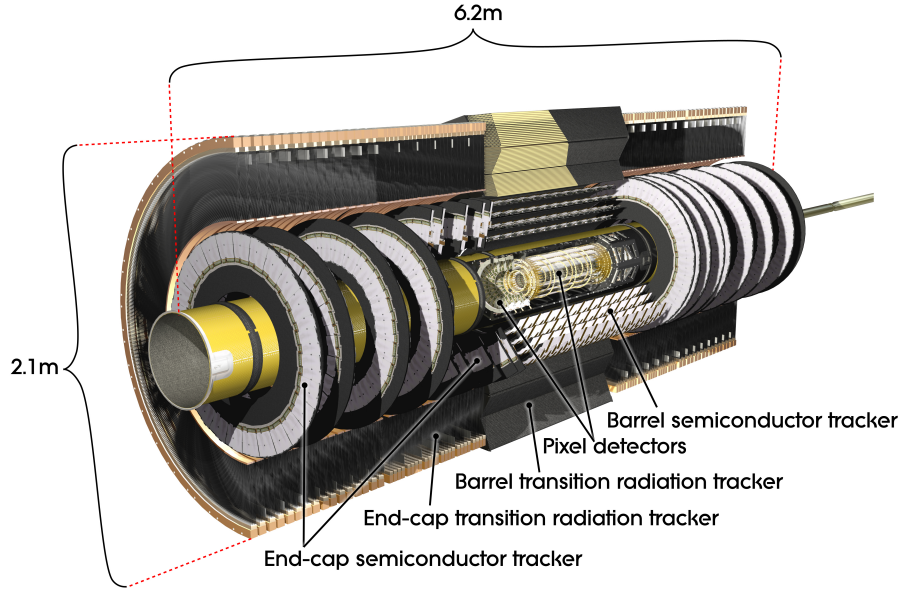


Figure 5.7: A 3D view of the ATLAS inner detector. Illustration from [67]

have cylindrical symmetry, and two endcap regions (A-side and C-side) where the sub-detectors are disks with the beamline going through the center.

5.3.1 The pixel detector

The pixel detector [69] is composed of four layers of silicon pixels in the barrel regions and three disks in each endcap. When a charged particle goes through a silicon pixel, electrons from the valence band are taken to the conduction band, creating electron-hole pairs. The electrons and holes then drift toward the electrodes and create a signal. Silicon was chosen as it has a favorable energy band gap (of 1.12 eV) that can be tuned with doping [70].

There are 92 million silicon pixels in total, with a size of $50 \times 400 \mu\text{m}^2$ except for the innermost layer which has pixels with a size of $50 \times 250 \mu\text{m}^2$. The innermost layer is called the insertable B-layer (IBL) and was added in 2014 [71].

5.3.2 The semiconductor tracker

The semiconductor tracker (SCT) [72] is composed of four layers in the barrel and nine layers in each end cap. The geometry of the detector is designed so that particles with $|\eta| < 2.5$ cross at least four layers. Like the pixel detector, silicon sensors are used, with over 6 million micro-strips

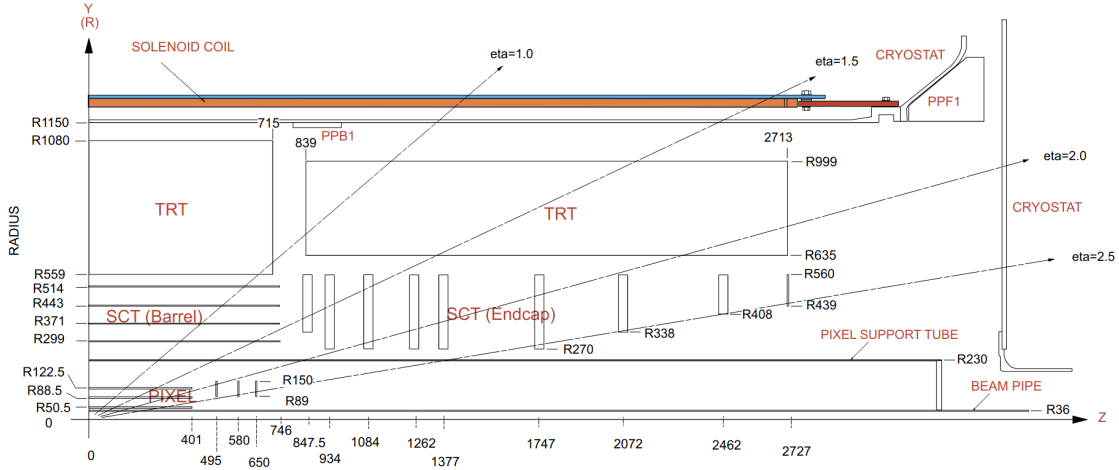


Figure 5.8: Quadrant view of the ATLAS inner detector. Illustration from [68]

sensors packed into 4088 modules. In each SCT module, there are two silicon layers, one on each side but rotated by 40 mrad (about 2.3°) so that SCT can measure the z coordinate as well.

5.3.3 The transition radiation tracker

The transition radiation tracker (TRT) [73] is made of about 300000 drift tubes called straws, some filled with a xenon-based gas mixture, and others with an argon-based gas mixture. There is gold-plated tungsten wire at the center of the straw, and the wall is kept at a voltage of -1.5 kV, creating a strong electric field. Ionization clusters are produced when a particle is passing through the gas mixture, and those clusters drift toward the central anode wire, creating an electrical signal.

The TRT can also perform particle identification, by distinguishing, for example, electrons and pions. This is because the detector is designed to trigger transition radiation, which is the radiation emitted from charged particles moving through inhomogeneous media. The transition radiation is more intense for larger γ -factor, and less massive particles tend to have a larger γ -factor (because $\gamma = E/m$), which enables particle identification.

5.4 The ATLAS calorimeter system

The ATLAS calorimeter system is composed of two parts: the liquid-argon calorimeter (LAr) [74] and the tile calorimeter (TileCal) [75]. It was designed to measure the energy of electrons, photons and hadrons, by using alternative layers of passive material to create particle showers and active

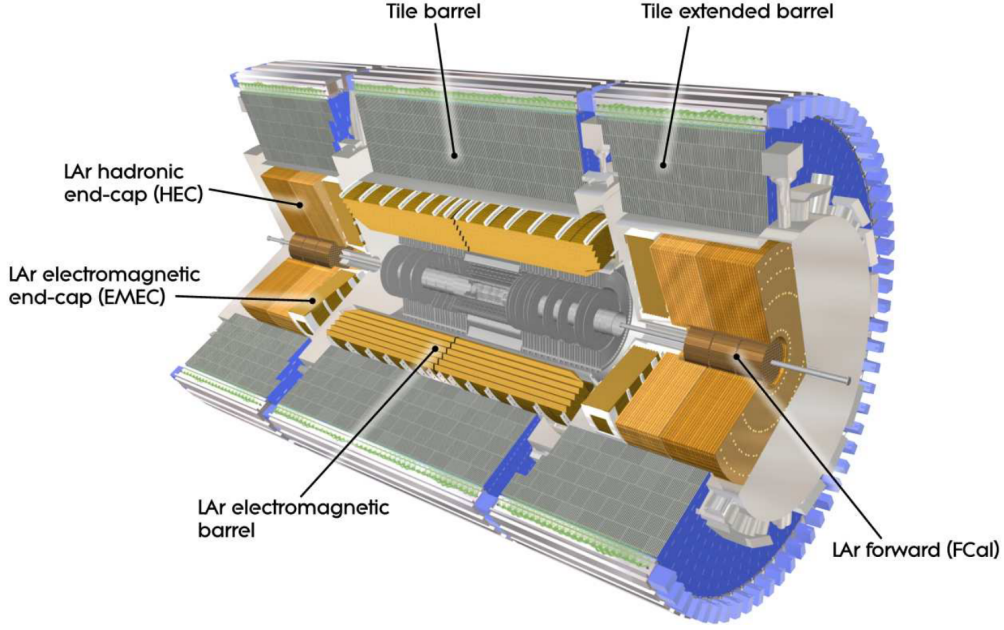


Figure 5.9: The ATLAS calorimeter system. [54]

material for detection. Figure 5.9 shows an illustration of the ATLAS calorimeter system.

The components of the ATLAS calorimeter system are also designed to be large enough so that, with high probability, all of the energy of the particles is deposited inside the detector. For the electromagnetic part of the calorimeter, one looks at the *radiation length* X_0 for the material used, defined by the average distance traveled by an electron before having its initial energy reduced by a factor of $1/e$. For the hadronic part, the relevant quantity is the *interaction length* λ defined by the average distance traveled by a hadron before having an inelastic nuclear interaction.

The LAr is composed of the electromagnetic barrel calorimeter (EMB, $|\eta| < 1.475$), the electromagnetic endcap calorimeter (EMEC, $1.375 < |\eta| < 3.2$), the hadronic endcap calorimeter (HEC, $1.5 < |\eta| < 3.2$) and the forward calorimeter (FCal, $3.1 < |\eta| < 4.9$) (see figure 5.9). The EMB and the EMEC form the ATLAS electromagnetic calorimeter, while the HEC, the FCal and the TileCal (see section 5.4.2) form the ATLAS hadronic calorimeter. The thickness of the electromagnetic calorimeter is greater than $22 X_0$ in the barrel and $24 X_0$ in the endcaps, for the hadronic calorimeter the thickness is 9.7λ in the barrel and 10λ in the endcaps.

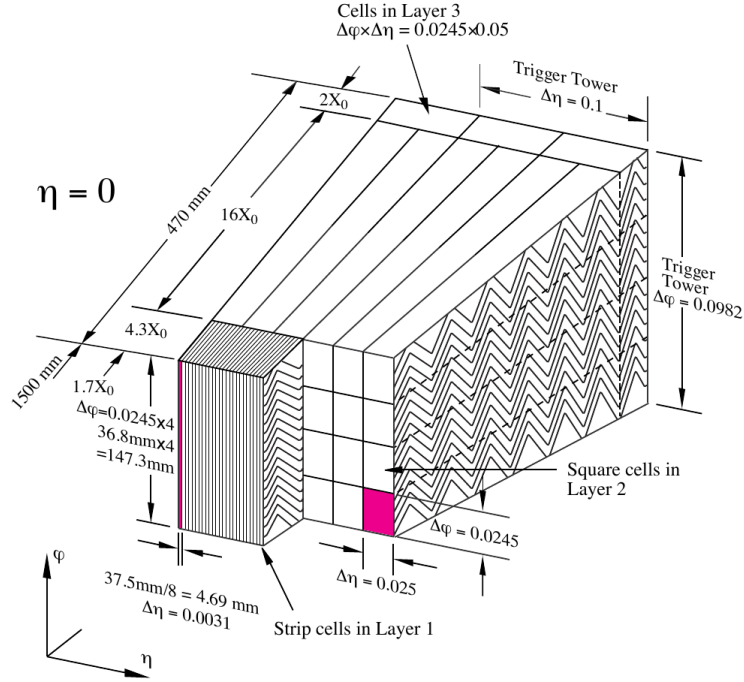


Figure 5.10: Illustration of the structure of the liquid-argon calorimeter. [74]

5.4.1 The liquid-argon calorimeter

The active layers in the LAr are filled with liquid argon and are kept at a temperature of -184°C . A high voltage is applied between the absorbing layers (the passive material) and readout electrodes in the liquid argon so that when a charged particle travels through the liquid argon it creates an ionization signal.

The use of liquid argon provides a good energy resolution for a reasonable cost as well as good resistance to radiation. However, the response time can be somewhat large in part because of the capacitance and inductance created by the cables and connections between the layers. The solution is to use an ‘accordion’ geometry [76] as shown in figure 5.10, which also has the advantage of removing dead space to get a full ϕ coverage. The accordion geometry is used for the electromagnetic calorimeter (EMB and EMEC), and lead is used for the passive layers, to create electromagnetic showers.

In the FCal copper is used for the first layer and tungsten for the other two layers and in the HEC, copper is used for all passive layers. The first layer is used for electromagnetic measurement and the other two for hadronic measurement.

5.4.2 The tile calorimeter

The TileCal is composed of the tile barrel covering the $|\eta| < 1$ region and the extended barrels (one on each side) covering the $0.8 < |\eta| < 1.7$ region. Steel is used as an absorber (passive layers) and plastic scintillating tiles are used for the active layers. When charged particles travel through the active layers, scintillating light is produced and transmitted through fibers. The light then reaches photomultiplier tubes, giving rise to an electronic signal.

5.5 The muon spectrometer

The muon spectrometer is the outermost part of the ATLAS detector and is designed to identify muons and measure their momentum. It is composed of the monitored drift tubes (MDTs), the cathode strip chambers (CSCs), the resistive plate chambers (RPCs) and the thin gap chambers (TGCs).

The MDTs are used for tracking. They are an ensemble of 354,240 drift tubes, filled with an argon-based gas mixture, and covering most of the $|\eta| < 2.7$ region. The CSCs are also used for tracking and cover the $2 < |\eta| < 2.7$ region. They are multi-wire proportional chambers: there are anode wires between cathode strips, with the space filled by an Ar (80%) and CO₂ (20%) gas mixture. There are two sets of strips that are orthogonal to each other to have more accurate tracking.

The RPCs are used for triggering and cover the $|\eta| < 1.05$ region. They are made of anode and cathode plates (they are wireless chambers) with an organic gas mixture. The TGCs are also used for triggering and cover the $1.05 < |\eta| < 2.4$ region. They are multi-wire proportional chambers operating with a CO₂ and n-pentane gas mixture.

5.6 The trigger and data acquisition system

The large collision rate at the LHC is needed for the study of rare processes, but it also makes it impossible to record everything. Fortunately, the vast majority ($> 0.99\%$) of collisions in the LHC are not really useful for physics, with just parton interactions creating low-energy jets.

The ATLAS trigger and data acquisition system (TDAQ) [78] is designed to filter out these superfluous events in real-time and record all the potentially interesting ones. There are two levels

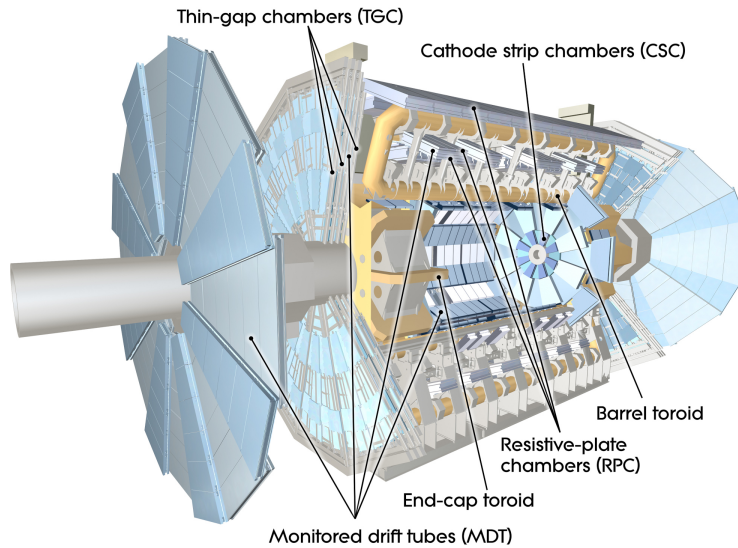


Figure 5.11: The ATLAS muon spectrometer. [77]

for the triggering system. The first is called the L1 (level 1) trigger and reduces the data rate from 40 MHz to about 100kHz. The second one is the high-level trigger (HLT) reducing the data further from 100kHz to approximately 1kHz.

In the L1 triggering system, information from the calorimeters and the muon spectrometer is fed to the L1Calo trigger and the L1Muon trigger. After processing, information about identified objects goes through the L1Topo (level-1 topological) and to the central trigger processor. The decision to keep or throw away an event is done in less than $2.5 \mu\text{s}$.

The HLT uses information from the whole detector. Event selection is done in less than $200 \mu\text{s}$ using about 40,000 CPU cores. Events are then recorded based on trigger chains, which are a combination of L1 trigger items and HLT algorithms. Different trigger chains exist that are useful for different purposes, there are trigger chains for (to give a few examples) leptons (electrons and muons), jets, photons, and missing energy. A combination of the available trigger chains is chosen for each analysis.

The events are recorded with different streams, corresponding to different trigger menus (lists of trigger chains). Among these streams are physics streams for analysis, but also, (for example) streams for calibration and monitoring.

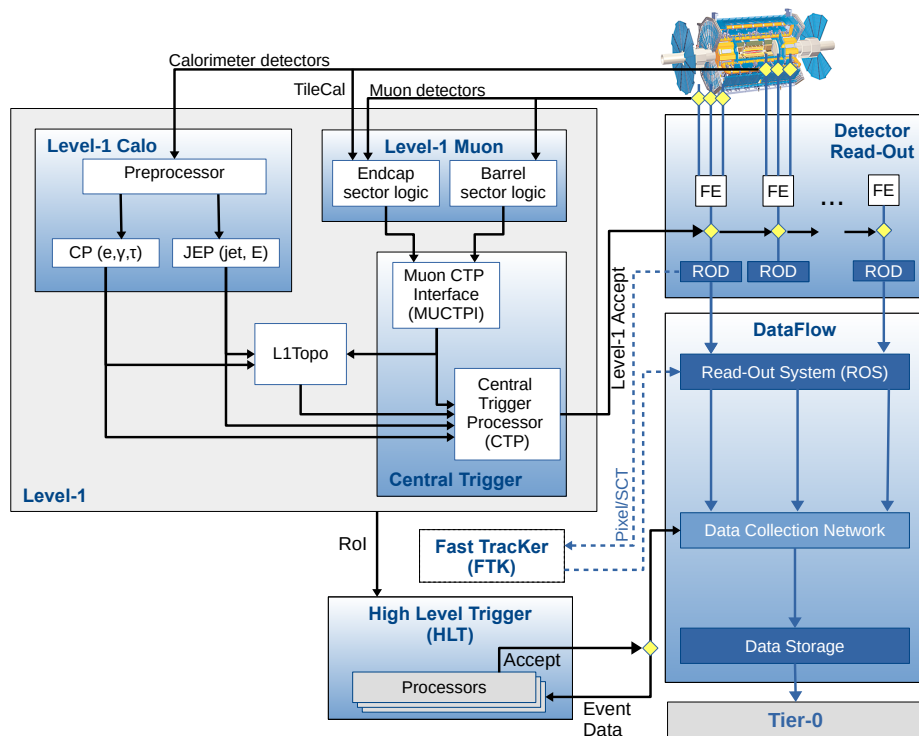


Figure 5.12: An illustration of the ATLAS TDAQ [78]. Data flows through the L1 triggers, then through the HLT before being recorded.

Chapter 6

Reconstruction

6.1 Introduction

Raw data from the detector are used to reconstruct various physics objects such as electrons muons, jets or missing transverse energy. These reconstructed and calibrated objects are then used by physics analysis. This chapter discusses how reconstruction is done in ATLAS.

6.2 Tracking

Tracking consists of identifying and measuring charged particle tracks using the inner detector. Two different approaches are used for tracking, the inside-out and the outside-in tracking [79].

Inside-out tracking is performed following several steps:

1. The tracking starts with the creation of 3D space points from clusters in the pixel detector and the SCT. The clusters consist of connected cells in the pixel detector and the SCT, with energy deposits above a certain threshold [80].
2. The next step is the selection of space point triplets for seeding. For every seeding triplet, a track-finding algorithm is run, using a combinatorial Kalman filter technique [81].
3. An ambiguity-solving module is then used to remove fake tracks and track duplicates. A score is given to every track, depending on the quality of the track, overlap with other tracks and missing expected hits (called holes). Track candidates with a score that is too small are deleted.
4. The last step is to extend the track to the region covered by the TRT to improve tracking and momentum resolution.

On the other hand, the outside-in approach starts by finding and fitting tracks in the TRT. A tool is used to prevent TRT hits that were used for inside-out tracking to be used for outside-in tracking. The track is then extended backward toward the SCT, this is called backtracking.

Tracks can be represented with five parameters using the so-called *perigee representation* [82], illustrated in figure 6.1. The parameters are $(d_0, z_0, \phi, \theta, q/p)$ and are defined at the perigee point,

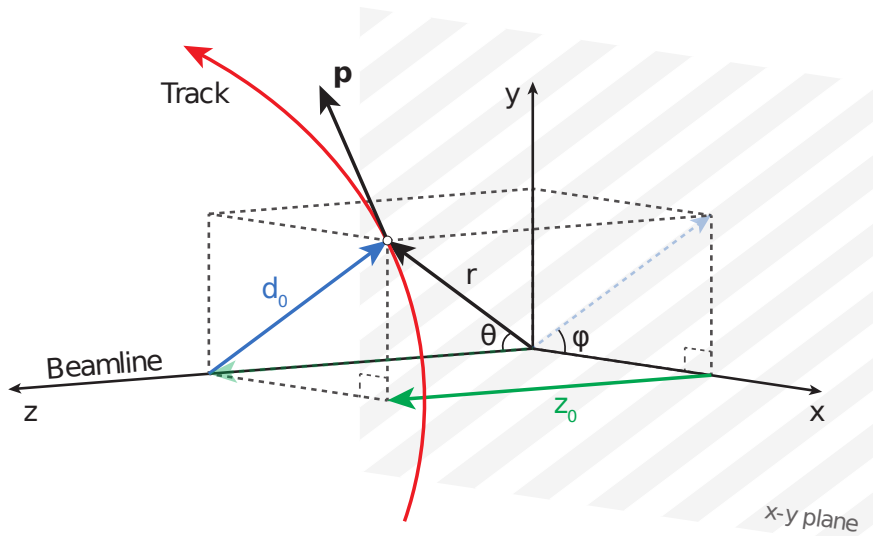


Figure 6.1: The perigee representation of a track in ATLAS. Illustration from [65].

the point in the track that is the closest to the beamline. d_0 is the distance between the perigee and the beamline and is called the transverse impact parameter. z_0 is the longitudinal impact parameter, the distance between the perigee and the origin along the z -axis. ϕ and θ are the azimuthal and polar angles. Finally, q/p is the ratio of the electric charge q over the momentum p . Larger values of q/p correspond to smaller track curvature.

6.3 Vertex reconstruction

A *vertex* is a point of interaction between two protons. There are several steps for vertex reconstruction [83]:

1. A set of tracks is selected, based on a few quality requirements.
2. A seed position is determined for the first vertex.
3. An iterative fit is used to determine the position of the vertex at each iteration and tracks that are less compatible with this vertex are given a smaller weight.
4. Once the fit has converged, incompatible tracks are removed from the vertex and are reused to find the next vertex.
5. The procedure continues until no new vertex can be found, or until there is not any track left.

6.4 Topological calorimeter clusters

Clusters of topologically connected calorimeter cell signals are called topo-clusters [84] and are used for the reconstruction of the hadronic part of the final states (such as jets and tau decaying hadronically) and the reconstruction of softer particles (electrons and photons).

The first step is to identify seeding cells as starting points for the algorithm. The identification is based on cell significance, defined as the ratio of the cell energy over the expected noise. Seeding cells are those with a cell significance above four. All neighboring cells with cell significance above two are iteratively added to the cluster. The last step is to add all neighboring cells with a positive signal¹.

A cluster-splitting algorithm is then run to avoid having a single cluster produced by two or more particles. If a cluster has at least two local maxima, indicating the presence of two or more particles in close proximity, cells are associated iteratively with the corresponding maxima until the splitting is complete. A local maximum corresponds to a cell with $E_{\text{cell}}^{\text{EM}} > 500\text{GeV}$ and having at least four neighbors, as well as a signal larger than any neighboring cell.

6.5 Leptons

6.5.1 Electrons

Electrons are identified when energy deposits in the electromagnetic calorimeter (topo-clusters) can be matched with an inner detector track [85]. After the initial matching, tracks are refitted to take bremsstrahlung radiation into account. If the track points to the primary vertex, an electron is identified. If no matching track could be found, the particle is identified as a photon instead of an electron. If there is a track that does not match the primary vertex, a converted photon is identified, i.e. a photon that created a e^+e^- pair before the EM calorimeter.

A single electron (or photon) can create multiple topo-clusters for example from bremsstrahlung radiation. For this reason, ‘superclusters’ are defined, starting from seed cluster candidates and adding satellite clusters. For electrons, satellite clusters are searched in a window of $\Delta\eta \times \Delta\phi = 0.125 \times 0.300$ around the center of the seed cluster.

Electron identification is performed using a likelihood-based method [84]. Different thresholds

¹Negative cell signals can exist due to electronic noise and fluctuations from pile-up

for the likelihood discriminant are defined, corresponding to different efficiencies. They are called: Loose, Medium and Tight, and correspond (on average and for typical electroweak processes) to an efficiency of 93%, 88% and 80% respectively.

6.5.2 Muons

Four muons types are defined for reconstruction [86]:

- Combined muons: An ID and an MS track are refitted together to create a combined track.
- Segment-tagged (ST) muons: an ID track is extended to the MS with at least one track segment in the MDT or CSC chambers.
- Calorimeter-tagged (CT) muons: a track in the ID is matched with an energy deposit in the calorimeter. The idea is to be able to reconstruct muons in regions where the MS can't be used, at the cost of lower purity.
- Extrapolated (ME) muons: the muon trajectory is reconstructed using the MS only, with a requirement of pointing toward the interaction point. This allows muon reconstruction outside the ID acceptance.

Various working points exist for muon identification [87]. These are the Loose, Medium, and Tight working points, each one corresponding to a tighter selection than the previous one. Two other working points are defined: the High- p_T working point optimized for muons with $p_T < 100\text{GeV}$ and the Low- p_T working point for low p_T muons.

6.5.3 Lepton isolation

We are often interested in *prompt* leptons, defined as leptons produced from the main collision, for example from the decay of a heavy particle like a Z or W boson. Prompt leptons tend to be isolated unlike non-prompt leptons coming from secondary processes.

This is why isolation requirements exist for leptons in ATLAS. The idea is to look inside a cone around the lepton and apply a cut on the total transverse momentum of the other tracks there as well as cutting on the calorimeter energy deposited around the lepton cluster. Various isolation working points are defined for electrons and muons and those relevant to the analysis described in this work are discussed in chapter 8.

6.6 Jets

6.6.1 Jet reconstruction

The jet reconstruction algorithm used for this analysis is called the particle flow algorithm [88]. The first step is the reconstruction of *particle flow* (PFlow) objects that are built from the matching of ID tracks with topo-clusters, representing (ideally) individual particles. This combination of information from the ID and the calorimeters is advantageous as the ID gives better tracking for charged particles and better momentum resolution for low p_T particles, while the calorimeter offers a better energy resolution at high p_T and can also reconstruct neutral particles (unlike the ID). To avoid double counting the energy when combining tracks and topo-clusters, the energy of the particle, as measured in the ID, is subtracted from the calorimeter measurement. The main difficulty is to avoid removing energy deposits from other particles. The particle flow algorithm has a better resolution and is more robust against pile-up, compared to algorithms that are using only topo-clusters as inputs [88].

Once particle flow objects are reconstructed, an algorithm is used to combine them to form jets. Several algorithms exist for this task. The most popular one, which is also the one used for this analysis, is called the anti- k_T algorithm [89]. The following ‘distances’ are defined:

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (6.1)$$

$$d_{iB} = p_{T,i}^{2p} \quad (6.2)$$

where p is either -1, 0 or 1, $p_{T,i}$ is the transverse momentum of the i^{th} object, $\Delta R_{ij}^2 = \Delta\phi_{ij}^2 + \Delta\eta_{ij}^2$ and R is a fixed parameter, called the jet radius parameter. Different values of p are used for different algorithms: $p = -1$ corresponds to the anti- k_T algorithm, $p = 1$ to the k_T algorithm [90] and $p = 0$ to the Cambridge/Aachen algorithm [91]. d_{ij} is the distance between object i and j , while d_{iB} is the distance between object i and the beam B .

The algorithm can be decomposed as follow:

1. All the d_{ij} and d_{iB} are computed.
2. The smallest of all the d_{ij} and d_{iB} is identified.
3. If the minimum found is a d_{ij} , a new object is formed (called a *pseudojet*) by combining

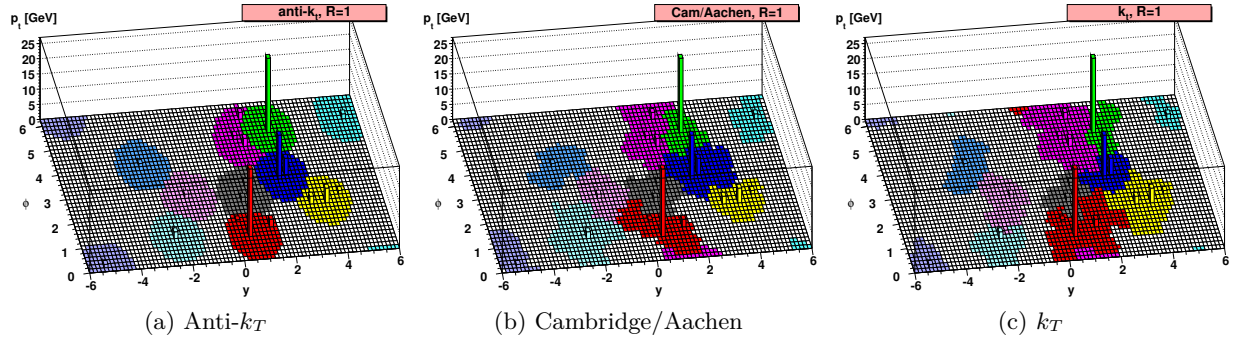


Figure 6.2: Comparison of clustering algorithms: Anti- k_T (left), Cambridge/Aachen (middle) and k_T (right). Illustration from [92].

objects i and j .

4. If the minimum found is a d_{iB} , the object i is identified as a jet and removed from the list.
5. Iterate until no object (particle or pseudojet) is left.

It is noteworthy that with the anti- k_T algorithm, soft particles tend to combine with hard particles rather than among themselves. An isolated hard particle tends to gather all soft particles within a circle of radius R , and the resulting jet will be conical. This is not true for the k_T and Cambridge/Aachen algorithms, as illustrated in figure 6.2

All three algorithms mentioned in this section are infrared and collinear safe: an additional soft emission or collinear splitting would leave the set of hard jets in a particular event unchanged. The anti- k_T algorithm has the additional property of being soft-resilient, which means that the shape of the jet is not influenced by soft radiation.

6.6.2 B-tagging

It is important, for many physics analyses, to identify jets produced by b quarks. Several properties of the b quark are useful for tagging: its relatively long lifetime means that b hadrons decay products tracks have vertices with large impact parameters, its relatively large mass means that the jets tend to be wider and with a larger invariant mass (compared to light-flavor jets) and its decay properties have the consequence that the presence of muon in the jet is enhanced in b -jets [93].

B-tagging in ATLAS is done in two steps. In the first step, low-level algorithms are used, based either on the impact parameters of the tracks or the reconstruction of displaced vertices. The output of low-level algorithms is then used as inputs for high-level algorithms, which are based on

multivariate techniques and classify jets as b -jets, c -jets or light-flavor jets. Several working points exist corresponding to different efficiencies.

6.7 Missing energy

Missing transverse energy reconstruction is important in order to have information about (weakly interacting) particles going through ATLAS without being detected, like neutrinos.

In order to have the most accurate reconstruction for E_T^{miss} , we should not only take into account ‘hard’ objects (reconstructed leptons, photons and jets) but ‘soft’ signals as well, corresponding to charged particles tracks from the primary vertex that are not associated with any hard object. The reconstructed missing transverse energy is given by [94]:

$$\vec{E}_T^{\text{miss}} = - \sum_{i \in \{\text{hard objects}\}} \vec{p}_{T,i} - \sum_{j \in \{\text{soft signals}\}} \vec{p}_{T,j} \quad (6.3)$$

$$= - \underbrace{\sum \vec{p}_T^e - \sum \vec{p}_T^\mu - \sum \vec{p}_T^\tau - \sum \vec{p}_T^\gamma - \sum \vec{p}_T^{\text{jets}}}_{\text{hard term}} - \underbrace{\sum \vec{p}_T^{\text{soft}}}_{\text{soft term}} \quad (6.4)$$

Chapter 7

Data and Monte Carlo samples

7.1 Introduction

This chapter gives an overview of the data and simulation datasets used for the EWK WWjj analysis. The next section describes data collected by ATLAS, with information about triggers and data quality requirements. The last section gives information about signal and background samples settings and configurations for production.

7.2 Data samples

The data used for this analysis was collected by the ATLAS detector during run-II of the LHC (from 2015 to 2018). This corresponds to a total integrated luminosity of 139 fb^{-1} . Single electron and muon triggers are used, with electrons (muons) required to pass at least one electron (muon) trigger. The list of single electron triggers [95] and single muon triggers [96] is shown in table 7.1 and 7.2 respectively.

Years	Trigger
2015	HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium HLT_e120_lhloose
2016-2018	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0

Table 7.1: List of single-electron triggers used for the analysis.

By convention, triggers in ATLAS are written as:

[Trigger level]_[object multiplicity][object type][minimum E_T value in GeV]

For example, HLT_e24_lhmedium_L1EM20VH is an HLT trigger for single electrons with a E_T threshold of 24GeV, using the lhmedium working point for electron identification. Different working

Years	Trigger
2015	HLT_mu20_iloose_L1MU15 HLT_mu50
2016-2018	HLT_mu26_ivarmedium HLT_mu50

Table 7.2: List of single-muon triggers used for the analysis.

points correspond to different efficiencies. The available working points for electron identification are `lhvloose`, `lhloose`, `lhmedium` and `lhtight`, corresponding to looser or tighter selection criteria. `L1EM20VH` refers to an L1 trigger requiring $p_T > 20\text{GeV}$, an E_T -dependent veto on the energy deposit in the hadronic calorimeter (behind the EM calorimeter signal left by the electron candidate), with the E_T threshold depending on η (to compensate for different amounts of passive material in different η slices).

In table 7.1, `nod0` means that the impact parameter d_0 is not used for electron identification and `ivarloose` means that variable-sized cone isolation requirement is applied.

In table 7.2, `L1MU15` refers to an L1 muon trigger requiring $p_T > 15\text{GeV}$. `iloose` and `ivarmedium` refer to isolation requirements for muon candidates, based on the scalar sum of tracks p_T in a cone around the muon track. `ivarmedium` has replaced `iloose` with an optimized cone sized to reduce the effect of pile-up [97].

7.3 Monte Carlo samples

7.3.1 Signal sample

The signal sample used in this analysis is a pure electroweak sample which includes diagrams of $\mathcal{O}(\alpha_{EW}^6)$ in perturbation theory. The sample includes events with one electron and one muon in the final state, as well as events with a tau lepton, these tau-lepton events amount to around 11% of signal events in the signal region (for the definition of the signal region, see chapter 8), as shown in table 7.3. The sample is simulated with the POWHEGBOX v2 [43] generator at NLO QCD with the NNPDF2.3 LO[98] PDF set. The sample is showered with PYTHIA 8.244 [47] using the CTEQ6L1 PDF set [99] with the AZNLO tune [100]. The AZNLO tune is a particular PYTHIA parametrization

used for POWHEG processes involving electroweak bosons.

Electroweak W^+W^-jj production involves diagrams with an exchange of a low-virtuality photon in the t-channel which leads to singularities. This is characterized by outgoing partons with low transverse momentum. In order to damp this kind of configuration, a *born suppression factor* $F(\Phi_n)$ is defined [101]:

$$F(\Phi_n) = \left(\frac{p_{Tj_1}^2}{p_{Tj_1}^2 + \Lambda^2} \right)^2 \left(\frac{p_{Tj_2}^2}{p_{Tj_2}^2 + \Lambda^2} \right)^2 \quad (7.1)$$

where $p_{Tj_{1,2}}$ are the born-level (i.e. the core process, before showering) leading and sub-leading partons p_T and $\Lambda = 10 \text{ GeV}$. NLO electroweak correction for this process [102] are not included.

	Two jets	Three jets
EWK WWjj (no τ)	116.48	40.46
EWK WWjj (with τ)	14.54	5.11
Ratio (with τ /all)	11.10%	11.21%

Table 7.3: Number of events in the signal region for the electroweak production of one electron and one muon with opposite sign compared with the number of events where one of the lepton is a tau lepton, for $n_{\text{jets}} = 2$ and $n_{\text{jets}} = 3$.

7.3.2 Background samples

The largest background for this analysis is the top background, composed of $t\bar{t}$ and single top (Wt) events. Both samples ($t\bar{t}$ and Wt) samples were produced using the POWHEGBOX v2 [43] generator at NLO QCD with the NNPDF3.0NLO [98] PDF set. The sample is showered with PYTHIA 8.230 [47] using the A14 tune [103] and the NNPDF2.3 LO[104] set of PDFs. The mass of the top quark is set to 172.5 GeV . The h_{damp} parameter, one of the parameters controlling the matrix element/parton shower matching in POWHEGBOX is set to $1.5 m_{\text{top}}$. Bottom and charm hadrons decays are performed by EVTGEN 1.6.0 [105].

The other major background is the QCD W^+W^- production. QCD $qq \rightarrow W^+W^-$ and $gg \rightarrow W^+W^-$ production are simulated with SHERPA 2.2.2 [42] with up to one additional parton emission at NLO. The sample is showered with the internal SHERPA parton shower, with the matching and merging based on Catani-Seymour dipole factorization [106, 107] using the MEPS@NLO prescription [108–110].

Other processes contributing to the total background are W +jets, Z +jets and multibosons (WZ , ZZ , WWW , WWZ and ZZZ). W +jets production is simulated with the MADGRAPH5_aMC@NLO [41] generator with up to four partons at LO QCD. The sample is showered with PYTHIA 8.210 [47] using the A14 tune [103] and the NNPDF2.3 LO[104] set of PDFs. Z +jets production is simulated with SHERPA 2.2.1 [42] with up to two partons at LO and four partons at NLO. The sample is showered with the internal SHERPA parton shower, using the MEPS@NLO prescription [108–110] and the NNPDF2.3 LO[104] PDF set. The WZ and ZZ samples are simulated with the POWHEGBOX v2 [43] generator at NLO QCD. The samples are showered with PYTHIA 8.210 [47] using CTEQ6L1 [99] with the AZNLO tune [100]. The samples used to simulate the triboson processes (WWW , WWZ and ZZZ) are generated with the SHERPA 2.2.2 [42] with an NLO accuracy for the inclusive process and at LO for up to two additional parton emissions. The sample is showered with the internal SHERPA parton shower, with the matching and merging based on Catani-Seymour dipole factorization [106, 107] using the MEPS@NLO prescription [108–110].

Table 7.4 shows a summary of the matrix element generators and parton showers software used for signal and background simulation, as well as the matrix element QCD order $\mathcal{O}(\alpha_S)$.

Process	ME Generator	Parton shower	ME $\mathcal{O}(\alpha_S)$
EWK $W^+W^- + 2$ jets	POWHEGBOX	PYTHIA	NLO
QCD W^+W^-	SHERPA	SHERPA	NLO
$t\bar{t}$	POWHEGBOX	PYTHIA	NLO
single-top Wt	POWHEGBOX	PYTHIA	NLO
Z +jets	SHERPA	SHERPA	0,1,2jets@NLO +3,4jets@LO
W +jets	MADGRAPH5_aMC@NLO	PYTHIA	0-4jets@LO
WZ and ZZ	POWHEGBOX	PYTHIA	NLO
WWW , WWZ and ZZZ	SHERPA	SHERPA	NLO

Table 7.4: List of MC datasets used in the analysis.

Chapter 8

Object definition and event selection

8.1 Introduction

This chapter presents the reconstructed objects (electrons, muons, jets and missing transverse energy) used for this analysis. The signal and top control regions are also defined and discussed. Finally, cutflows showing the impact of various selection cuts on signal and background yields are shown at the end of the chapter. Finally, we present cutflow histograms that illustrate the impact of the various selection requirements on the signal and background event yields.

8.2 Object definition

8.2.1 Electrons

Electrons are reconstructed from the matching of tracks with energy deposits in the electromagnetic calorimeter. Reconstructed electrons used for this analysis are defined within $|\eta| < 2.47$ and excluding the $1.37 < |\eta| < 2.52$ region, which is the transition region between the barrel and the endcaps.

The transverse impact significance, defined as $|d_0|/\sigma(d_0)$ (where d_0 is the transverse impact parameter and $\sigma(d_0)$ the uncertainty on d_0), is required to be smaller than five. There is also a requirement on the longitudinal impact parameter: $|z_0 \sin(\theta)| < 0.5$ mm.

Two kinds of electrons are defined: *nominal* electrons and *veto* electrons. Nominal electrons are used, after further selections, in the signal and control region, while veto electrons are used for removing events with an extra electron.

There is an isolation requirement for nominal electrons, using tracks and calorimeter information around a cone surrounding the electron. The **Gradient** working point is used, which was designed to give an efficiency of 90% at $p_T = 25$ GeV and 99% at $p_T = 60$ GeV [85]. The minimal quality requirement for electron identification is the **TightLH** working point.

For veto electrons, the isolation working is **(FC)Loose** and the quality requirement is **TightLH** if $10 \text{ GeV} < p_T < 25 \text{ GeV}$ and **MediumLH** (if $p_T > 25 \text{ GeV}$).

Only electrons with $p_T > 10$ GeV are considered. Table 8.1 summarize the requirements for electrons used in the analysis.

	Requirements
p_T	$p_T > 10$ GeV
η	$ \eta < 2.47$ and excluding $1.37 < \eta < 2.52$
Impact parameters	$ d_0 /\sigma(d_0) < 5$ and $ z_0 \sin(\theta) < 0.5$ mm
Isolation	Gradient WP (nominal) (FC) Loose (veto)
Identification	TightLH WP (nominal) TightLH (if 10 GeV $< p_T < 25$ GeV) or MediumLH (if $p_T > 25$ GeV) (veto)

Table 8.1: Requirements for nominal and veto electrons.

8.2.2 Muons

Muons are reconstructed from the matching of inner detector tracks and information from the muon spectrometer and are defined within $|\eta| < 2.5$. The requirements on the impact parameters are $|d_0|/\sigma(d_0) < 3$ and $|z_0 \sin(\theta)| < 0.5$ mm.

As for electrons, we define nominal muons and veto muons.

The `Tight_FixedRad` working point is used for nominal muons isolation. It is similar to the `Tight` working point (defined in [86]) but with different requirements for muons with $p_T > 50$ GeV to increase the background rejection. The minimal quality requirement for nominal muons identification is the `Tight` working point.

Only muons with $p_T > 10$ GeV are considered. Table 8.2 summarize the requirements for muons used in the analysis.

8.3 Jets

Jet reconstruction starts from particle flow objects, which are objects combining tracking and calorimetric information [88]. Jets are then reconstructed using the anti- k_T algorithm [89] with

	Requirements
p_T	$p_T > 10 \text{ GeV}$
η	$ \eta < 2.5$
Impact parameters	$ d_0 /\sigma(d_0) < 3$ and $ z_0 \sin(\theta) < 0.5 \text{ mm}$
Isolation	Tight_FixedRad WP (nominal) Loose (veto)
Identification	Tight (nominal) Loose WP (veto)

Table 8.2: Requirements for nominal and veto muons.

the distance parameter set to $R = 0.4$. The jet energy scale and jet energy resolution are then calibrated. Jets are also required to be within $|\eta| < 4.5$.

In order to suppress jets coming from pile-up, the jet-vertex-tagger (JVT) algorithm [111] is used for jets with $p_T < 60 \text{ GeV}$ and $|\eta| < 2.5$. For jets in the forward region (i.e. $2.5 < |\eta| < 4.5$) where ID tracking information cannot be used, another algorithm is used, called the forward jet-vertex-tagger (FJVT) algorithm.

To identify b -jets, the DL1r b -tagging algorithm [93] is used for jets with $p_T > 20 \text{ GeV}$ and $|\eta| < 2.5$ (as ID tracking information is used for b -tagging) at the 85% efficiency working point. The DL1r algorithm uses a recurrent neural network for the low-level tagger and a deep neural network for the high-level tagger. The 85% efficiency working point was chosen because it is the highest efficiency DL1r working point available in ATLAS, and we want to suppress the top background as much as possible.

8.4 Missing transverse energy

Missing transverse energy is reconstructed using the momentum of calibrated electrons, muons and jets and a track-based soft term as was explained in section 6.7.

The missing transverse energy significance is also used for this analysis and is defined by:

$$E_T^{\text{miss}} \text{ significance} = \frac{E_T^{\text{miss}}}{\sqrt{H_T}} \quad (8.1)$$

where H_T is the scalar sum of the transverse momentum of reconstructed leptons and jets:

$$H_T = \sum p_T^e + \sum p_T^\mu + \sum p_T^{\text{jets}} \quad (8.2)$$

8.5 Selection

The signal region is defined with relatively loose requirements in order to let the neural network do most of the work (see chapter 9).

Events are required to include one electron and one muon (both nominal) and are required to have $p_T > 27$ GeV. A lepton veto is applied to remove events with additional electrons or muons that are passing the requirement to be a veto electron or a veto muon, as defined in the previous section.

A b-jet veto is also applied to suppress the top background. Drell-Yan events, which are events where a virtual photon or Z boson is produced and then decays to two opposite-charged leptons, are also suppressed by requiring the missing energy to be above 15 GeV (as there is no neutrino in the final state for this process).

The centrality of the leptons with respect to the two leading jets is defined by:

$$\text{centrality} = \min [\min(\eta_{lep0}, \eta_{lep1}) - \min(\eta_{jet0}, \eta_{jet1}), \max(\eta_{jet0}, \eta_{jet1}) - \max(\eta_{lep0}, \eta_{lep1})] \quad (8.3)$$

where η_{lep0} , η_{lep1} , η_{jet0} and η_{jet1} are the η coordinates of the leading and sub-leading lepton and jet. Centrality is required to be greater than 0.5, as it turned out to improve the signal/background ratio, as well as the performance of the neural network defined in chapter 9 when applied before training.

Only events containing either two or three jets with $p_T > 25$ GeV are considered. Events with more than three jets are rejected from the signal region, but they represent a fairly small fraction of all the signal events as shown in figure 8.1.

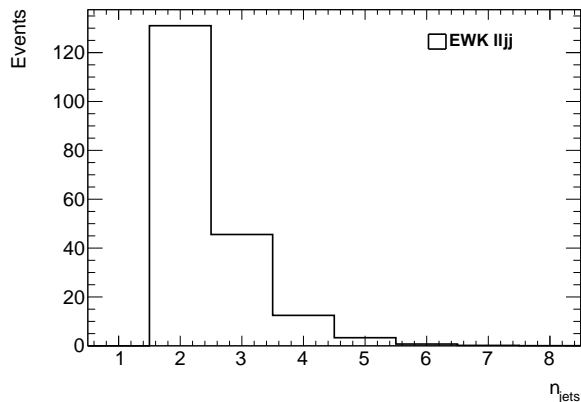


Figure 8.1: Number of jets for signal events in the signal region (but without a cut on the number of jets). There are not many events with more than three jets.

The VBF HWW process is suppressed by cutting on the invariant mass of the two leptons: $m_{e\mu} > 80 \text{ GeV}$, as depicted in Figure 8.2.

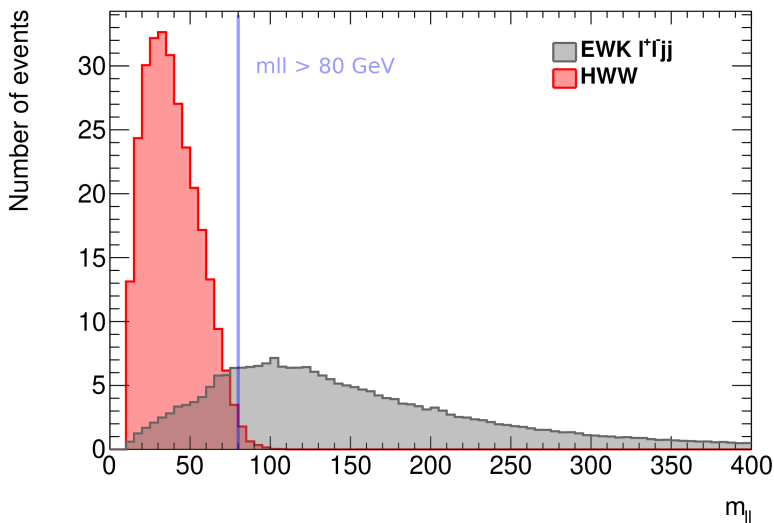


Figure 8.2: We apply a cut on m_{ll} ($m_{ll} > 80 \text{ GeV}$) to suppress VBF HWW and ggF HWW (both combined in red) and remove the Higgs peak. The signal (shown in grey) sample was produced without the Higgs resonance.

A control region (CR) is defined by requiring one of the two leading jets to be a b-jet. This region thus mostly contains $t\bar{t}$ events, which allows the control of their modeling, and constrains the related systematic uncertainties when performing the likelihood fit in chapter 11. Although QCD-WW is also a significant background, there is no corresponding dedicated control region as it is very difficult to define one that is sufficiently QCD-WW enriched.

Table 8.3 summarises the selection in the signal and control region.

Region	Requirements
Inclusive	<p>One electron and one muon with opposite charges</p> <p>No additional lepton with $p_T > 10$ GeV, (FC)Loose isolation, and TightLH (if $10 \text{ GeV} < p_T < 25 \text{ GeV}$) or MediumLH (if $p_T > 25 \text{ GeV}$) (el.)/Loose (mu.) identification</p> <p>Leptons p_T: $p_T > 27 \text{ GeV}$</p> <p>Lepton η: $\eta < 2.47$ and not $1.37 < \eta < 1.52$ (electron), $\eta < 2.5$ (muon)</p> <p>Lepton identification: TightLH</p> <p>Lepton isolation: Tight_FixedRad</p> <p>Two or three jets with $p_T > 25 \text{ GeV}$ and $\eta < 4.5$</p> <p>$p_T^{miss} > 15 \text{ GeV}$</p> <p>centrality = $\min \left[\min(\eta^{lep0}, \eta^{lep1}) - \min(\eta^{jet0}, \eta^{jet1}), \max(\eta^{jet0}, \eta^{jet1}) - \max(\eta^{lep0}, \eta^{lep1}) \right] > 0.5$</p> <p>$m_{e\mu} > 80 \text{ GeV}$</p>
Signal Region	b-jet veto (DL1r b-tagging with the 85% working point)
Control Region	Exactly one b-jet among the two leading jets

Table 8.3: Signal and control regions definitions.

8.5.1 Cutflow

A simple cutflow is shown in figure 8.3 for the signal region. We can see that the top background remains the largest one even after the b-veto cut is applied because of its large cross-section. The last two bins correspond to the signal region, split by $n_{\text{jets}} = 2$ and $n_{\text{jets}} = 3$. Figure 8.4 shows the corresponding signal/background ratio for the same cuts and regions.

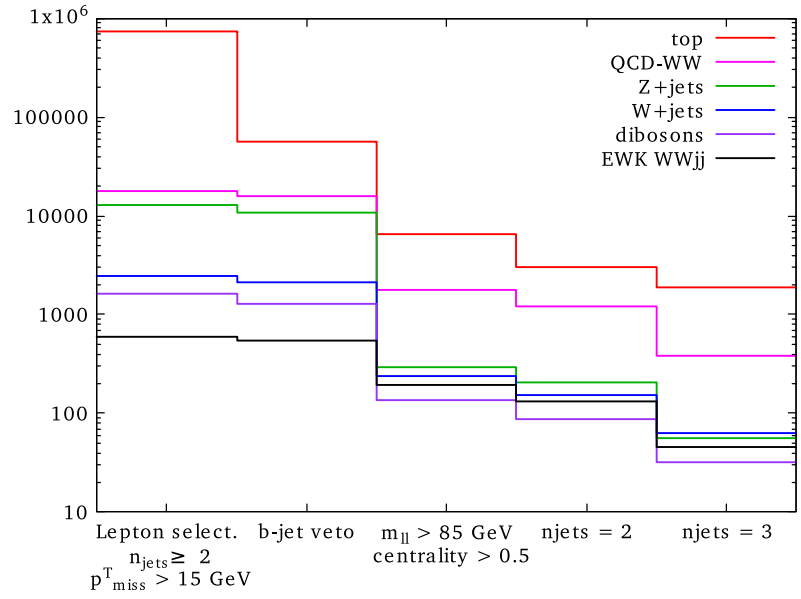


Figure 8.3: Cutflow for the signal region.

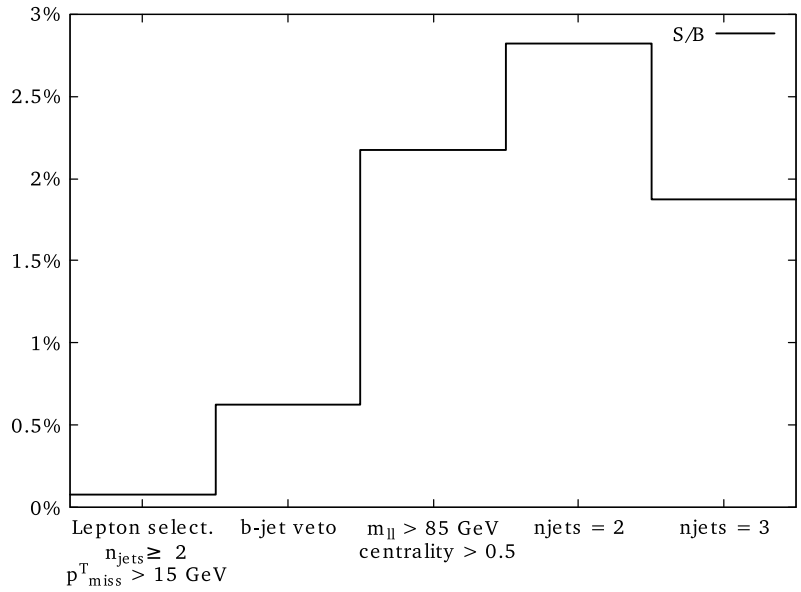


Figure 8.4: Signal/backgrounds ratio cutflow for the signal region.

Chapter 9

Neural network Analysis

9.1 Introduction

The cross-section for EWK W^+W^- production is very small, and the two main backgrounds (top and QCD W^+W^-) are quite difficult to suppress. It was therefore decided to use a neural network to optimize the signal/background separation. The idea is to trust the neural network to do most of the work, which is why only fairly loose cuts are used to define the signal region (see chapter 8 for more details on the selection).

This chapter starts with a brief introduction to neural networks on a theoretical level. The neural networks used for the analysis are then explained (input features chosen, hyperparameters, training, ...), and finally checks of overtraining, validations and performances are described.

9.2 Neural networks: definition and theoretical background

9.2.1 Definition

An *artificial neural network* is a collection of connected artificial *neurons* usually organized in layers.

An artificial neuron is a mathematical function $f: \mathbb{R}^n \times \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ of the form[112]:

$$y(\mathbf{x}, \mathbf{w}) = \sigma \left(\sum_{i=1}^n w_i x_i + w_0 \right) = \sigma \left(\sum_{i=0}^n w_i x_i \right) \quad (9.1)$$

where σ is a (usually monotonic) nonlinear function called the *activation function* and w_i are real numbers called *weights* or sometimes *synaptic weights*. w_0 is called the *bias* and we define $x_0 = 1$ to have simpler notations. Historically, artificial neurons are designed to be similar to biological neurons (e.g. in a human brain) in the way they process information from their inputs. However, recent progress in neuroscience has shown that biological neurons are often much more complex than artificial neurons [113, 114].

A *feed-forward neural network* is an artificial neural network without any closed cycle i.e. there can only be a forward propagation of information through the neural network. For example, a

two-layer feed-forward neural network is of the form:

$$y_k(\mathbf{x}, \mathbf{w}) = \sigma_2 \left(\sum_{j=0}^M w_{kj}^{(2)} \sigma_1 \left(\sum_{i=0}^D w_{ji}^{(1)} x_i \right) \right) \quad (9.2)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_D)$ is called the *input* and $\mathbf{y} = (y_1, y_2, \dots, y_K)$ is called the *output* of the neural network. \mathbf{x} corresponds to the input layer and \mathbf{y} to the output layer. In this example, there is also an *hidden layer* corresponding to the vector \mathbf{z} defined by:

$$z_j = \sigma_1 \left(\sum_{i=0}^D w_{ji}^{(1)} x_i \right) \quad (9.3)$$

In this case, there is only one hidden layer but there can be as many as we want (including zero). Figure 9.1 shows an illustration of such a network with $D = 4$, $M = 5$ and $K = 3$, i.e. an input with four numbers, five neurons in the hidden layer and an output with three numbers.

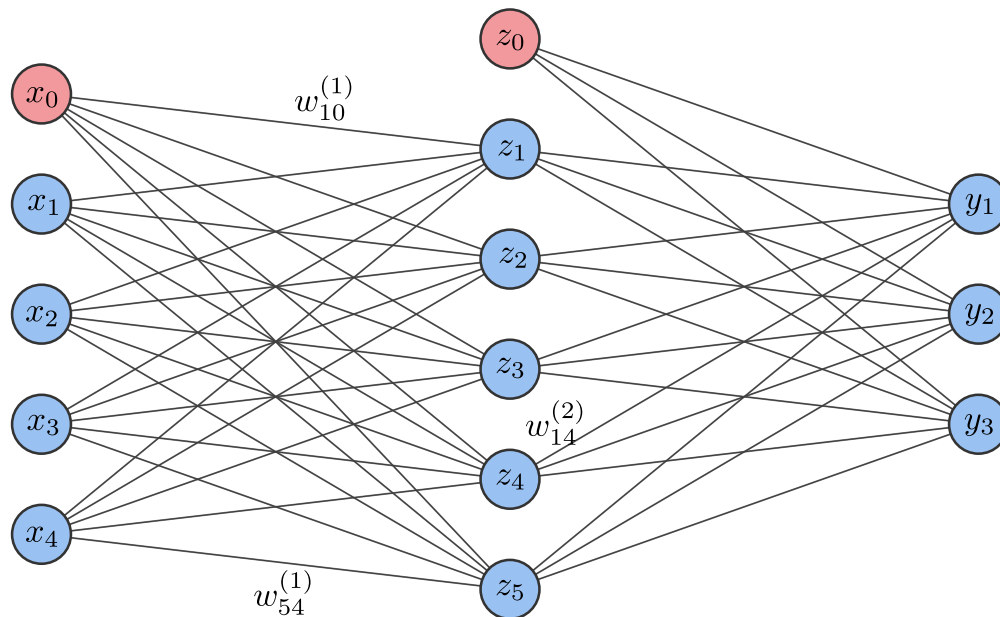


Figure 9.1: Example of a network with one hidden layer. The lines represent multiplication by some weight $w_{ij}^{(k)}$ and in each node, an activation function is applied. The red dots correspond to the biases.

It can be shown[115] that a network with at least two layers can approximate any continuous function with compact support to arbitrary precision if it has a sufficient number of neurons in the hidden layer(s). This is called the *universal approximation theorem*. This is why neural networks

are useful for a variety of tasks: regression, classification, modeling, simulation, ...

9.2.2 Classification

For our purposes here we only need to worry about classification problems. The idea is that we can use a network to classify the inputs into K different classes C_k given by the output of the network (a vector with K elements). Usually, the softmax function defined by:

$$\mathbf{y}_k = \sigma(\mathbf{a})_k = \frac{e^{a_k}}{\sum_{i=1}^K e^{a_i}} \quad (9.4)$$

is used on the output layer¹ for classification tasks. The result is a vector $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K)$ where $\forall i \in \{1, 2, \dots, K\}, 0 < \mathbf{y}_i < 1$ and $\sum_{i=1}^K \mathbf{y}_i = 1$. The interpretation of the output value is that the closer y_k is to one, the more likely it is that the input belongs to class C_k . It would be tempting to interpret the output as the probability that the input belongs to class C_k but it is not always the case in practice [116].

9.2.3 Training

In order to use a neural network for classification, the network must first be trained using data for which the correct categories are known. During the training, the weights are adjusted to minimize the *loss function*. The purpose of the loss function is to quantify how well the network can perform the classification. One very popular choice for the loss function is the *categorical cross-entropy* defined by:

$$L(\mathbf{w}) = \sum_{i=1}^N L_i(\mathbf{w}) = \sum_{i=1}^N \sum_{j=1}^K p_i(C_j) \ln(\hat{p}_i(C_j)) \quad (9.5)$$

where $p_i(C_j)$ is the probability that the i th input belongs to class C_j and $\hat{p}_i(C_j)$ is the i th output of the network corresponding to class C_j . Because we know during training what is the correct class, we have $p_i(C_j) \in \{0, 1\}$.

If we restrict ourselves to binary classifications (i.e. with only two categories), the softmax function is simply:

$$\sigma(\mathbf{x}_s, \mathbf{x}_b)_j = \frac{e^{x_j}}{e^{x_s} + e^{x_b}} \quad (9.6)$$

¹In our example (equation 9.2) $a_k = \sum_{i=0}^M w_{kj}^{(2)} \sigma_1 \left(\sum_{i=0}^D w_{ji}^{(1)} x_i \right)$

where the two classes are C_s and C_b (“signal” and “background”). What we are interested in and what we will call the output of the neural network is:

$$y = \sigma(\mathbf{x}_s, \mathbf{x}_b)_s = 1 - \sigma(\mathbf{x}_s, \mathbf{x}_b)_b \in [0, 1] \quad (9.7)$$

In this case, the loss function is the *binary cross-entropy*:

$$L(\mathbf{w}) = - \sum_{i=1}^N (t_i \ln(y_i) + (1 - t_i) \ln(1 - y_i)) \quad (9.8)$$

where t_i is 1 for C_s and 0 for C_b , and $y_i \in [0, 1]$ is the output of the neural network. The binary cross-entropy function is just the categorical cross-entropy function for cases where there are only two classes.

To minimize the loss functions, the weights are updated iteratively:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \eta \nabla L_{i(k)}(\mathbf{w}_k) \quad (9.9)$$

where $\eta > 0$ is called the *learning rate*. A different data point $i(k)$ is picked randomly at each iteration. This algorithm is called the *stochastic gradient descent*. It is also possible to use the gradient for the full training dataset $\nabla L(\mathbf{w}_k)$ at each iteration (*batch gradient descent*) or split the training dataset into batches and compute the gradient for each batch (*mini-batch gradient descent*). The full training dataset is often used many times, and every time it is used fully, it corresponds to what is called an *epoch*. Sometimes, thousands of epochs are needed to train a network, sometimes one epoch is enough [117].

To compute the gradient, the most commonly used technique is called *backpropagation*[118]. The algorithm can be summarized in two steps:

1. A forward pass, when the output of all neurons (including the final output) is computed.
2. A backward pass, when the gradient is computed with the chain rule using the intermediate values computed on the first step. This is done by starting at the output and propagating the errors (the partial derivative of the loss function with respect to the output of the neuron) to compute the gradient.

One significant issue with gradient descent is the *vanishing gradient problem* [119]. The gradient

in equation 9.9 can become very small, especially with networks with many hidden layers, which makes training very difficult. Various solutions exist to mitigate this problem. One of these possible solutions is an appropriate choice of the activation function, for example, the rectified linear units function defined by [120, 121]:

$$f(x) = \max(0, x) \tag{9.10}$$

is often a better choice than, for example, the sigmoid function $\frac{1}{1+\exp(-x)}$ to deal with the vanishing gradient problem. This might sound surprising as the gradient is zero for negative inputs, on the other hand, the gradient does not converge to zero for larger values. Having an output of zero for some neurons, (which is called *sparsity*) can be advantageous and lead to more efficient networks [122].

Figure 9.2 shows the rectified linear units function. It is often recommended to use it when using feedforward neural networks [123]. Another obvious advantage of this function is that is very fast to compute.

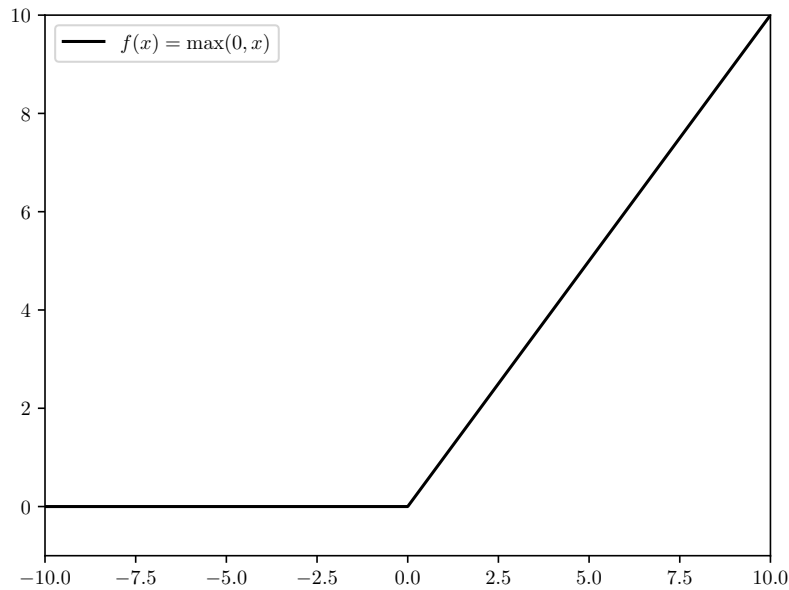


Figure 9.2: The rectified linear units function.

9.2.4 Overfitting

When a network is trained, it is also important to avoid *overfitting*. This happens when the network is fitting the training data too closely and ‘learns’ the noise in the dataset. An illustration is shown in figure 9.3.

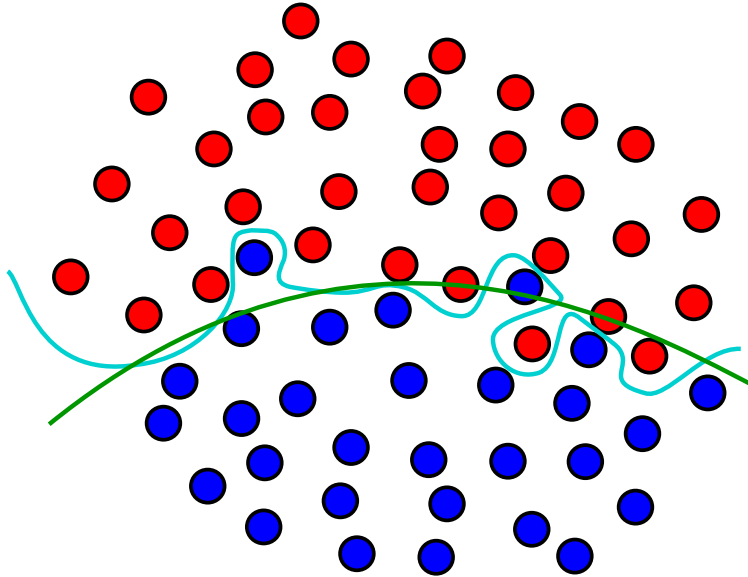


Figure 9.3: Illustration of overfitting for binary classification. The light blue line represents a case of overfitting while the green line shows a good fit.

Several techniques can be used to reduce (or even eliminate) overfitting. We will only discuss one of them, the one that is used for our analysis. This technique is called *dropout regularization* and is one of the most popular regularization methods used today.

The idea is to randomly drop neurons (as well as their connections) during training [124]. This is, to some approximation, equivalent to combining exponentially many different networks (all the different networks created by dropping neurons) and averaging the results. This not only reduces overfitting but also increases network performance. The probability to drop a neuron at each step can be adjusted, this is called the dropout rate. It can in principle be different for different layers.

9.3 Neural network used for the analysis

9.3.1 Introduction

For this analysis, the training and evaluation are done using TMVA [125] and Keras [126]. TMVA is a ROOT [127] environment with support for training and evaluation of multivariate classification and regression. ROOT is a data analysis framework particularly useful for high-energy physics. Keras is a python library providing an interface for training and running artificial neural networks. It can be used together with ROOT/TMVA using the PyKeras TMVA interface.

Two neural networks are used for this analysis: one for $n_{\text{jets}} = 2$ and one for $n_{\text{jets}} = 3$ (see also chapter 8 for information about event selection). This allows us to easily add input information about the third jet (for three jets event) and (implicitly) about the number of jets. Another motivation for having two different networks is that the kinematics are different between two and three jets events as shown in figure 9.4.

Both are trained to separate the signal from the two main backgrounds, the top ($t\bar{t}$ and single top) and QCD W^+W^- backgrounds.

9.3.2 Activation functions

The activation function used on the last layer is the softmax function:

$$y = \sigma(\mathbf{x}_s, \mathbf{x}_b)_s = \frac{e^{x_s}}{e^{x_s} + e^{x_b}} \quad (9.11)$$

and for the other layers, the activation function is the rectified linear function $f(x) = \max(0, x)$.

9.3.3 Overtraining check

The signal and background samples are both split into a *training sample* and a *testing sample* of equal size. The training sample is used for computing the gradient and updating the weights (using equation 9.9). The loss function used for training is the binary cross-entropy function and is also computed for the testing sample. If the loss function in the training sample gets significantly smaller than in the testing sample, this is a sign of possible overtraining. The trick is to keep the weights for which the loss function is minimal in the testing sample (often called the validation loss). In

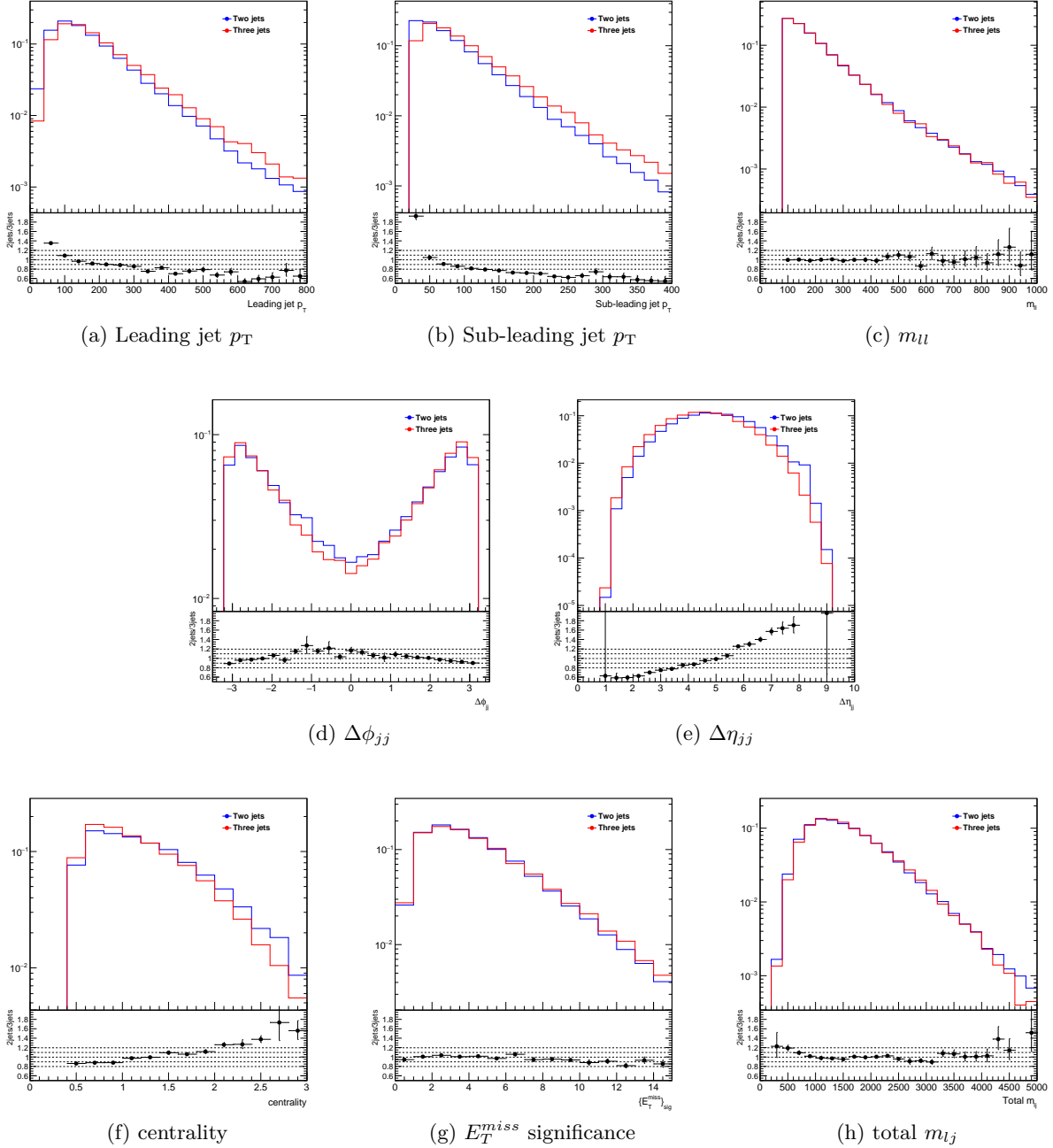


Figure 9.4: Comparison between variables used for training the NN output for $n_{\text{jets}} = 2$ and $n_{\text{jets}} = 3$ in the signal region. Significant differences can be seen in the leading and sub-leading jet p_T , $\Delta\eta_{jj}$ and centrality distributions.

principle, the training loss will keep decreasing, while the validation loss will first decrease, and then increase again. This is illustrated in figure 9.5 for the two-jets and three-jets networks.

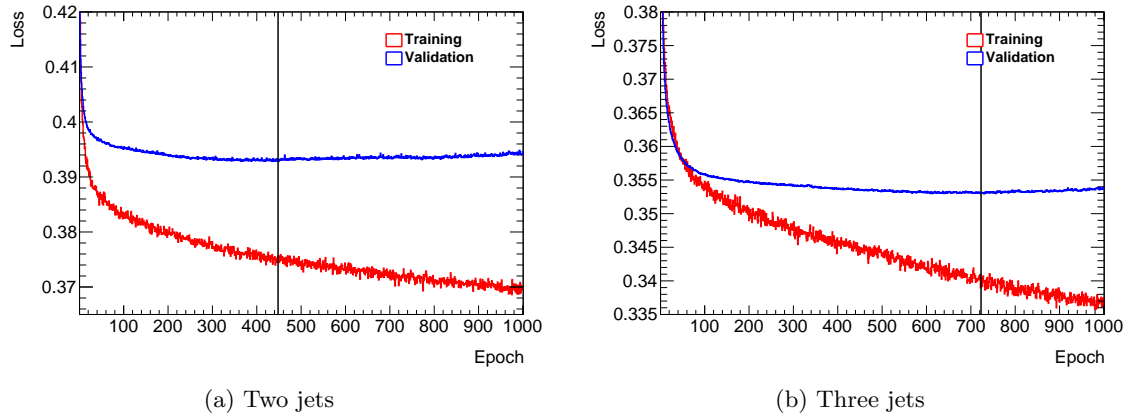


Figure 9.5: Training and validation loss in function of the number of epochs. The optimal point is indicated by the vertical line. As expected, the validation loss grows after the minimum while the training loss keeps getting smaller.

The dropout method is also used with a dropout rate of 0.1 to prevent overtraining. Nonetheless, we can see that the validation loss is a few percent smaller than the training loss, which could indicate a small overtraining. One simple check is to compare the neural network output distributions in both the training and testing dataset, and check if they are close to each other. This is shown in figure 9.6 comparing the neural network output distributions for signal and background. We can see that there is no significant difference between the training and testing dataset.

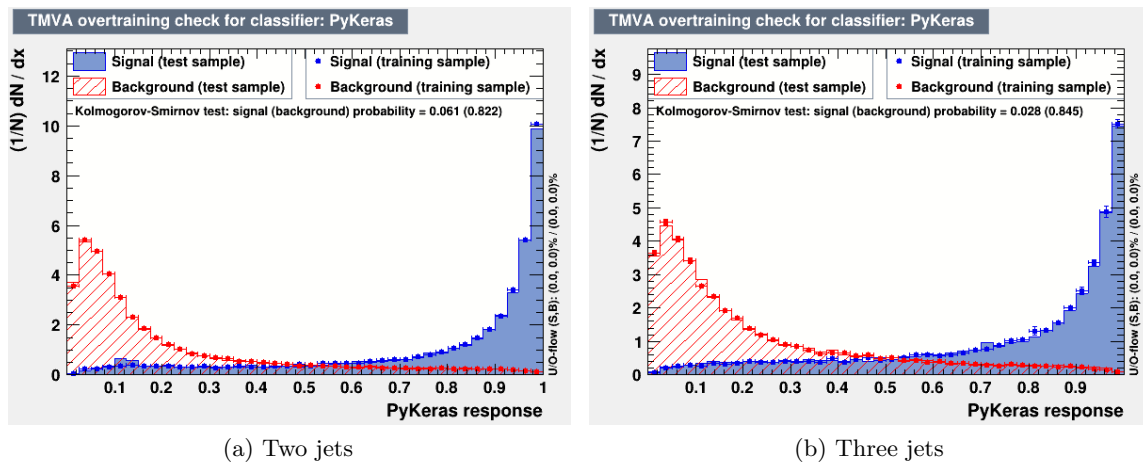


Figure 9.6: Normalized neural network output distribution for $n_{\text{jets}} = 2$ (left) and $n_{\text{jets}} = 3$ (right) in the signal region to check for possible overtraining.

9.3.4 Neural network inputs

The input variables to the two-jets neural network are the leading and sub-leading jet p_T , the invariant mass of the two leptons m_{ll} , the η difference between the two leading jets $\Delta\eta_{jj}$, the φ difference between the two jets $\Delta\varphi_{jj}$, centrality (defined in chapter 8), the missing transverse energy E_T^{miss} significance and total m_{lj} (the invariant mass of the two leading jets and the two leptons). The same variables are used for the three-jets network, with the addition of the third leading jet p_T and third jet centrality. The latter is defined by:

$$\left| y_{jet2} - \frac{y_{jet0} + y_{jet1}}{2(y_{jet0} - y_{jet1})} \right| \quad (9.12)$$

where y_{jet0} , y_{jet1} and y_{jet2} are the leading, subleading and third jet rapidities.

The choice of input variables was made by trying different sets of variables and looking, after re-training for each set, at the area under the ROC (Receiver Operating Characteristic) curve. The ROC curve shows the relation between background rejection and signal efficiency. It is created by varying a cut on the neural network distribution. Figure 9.7 shows the ROC curve for the two neural networks.

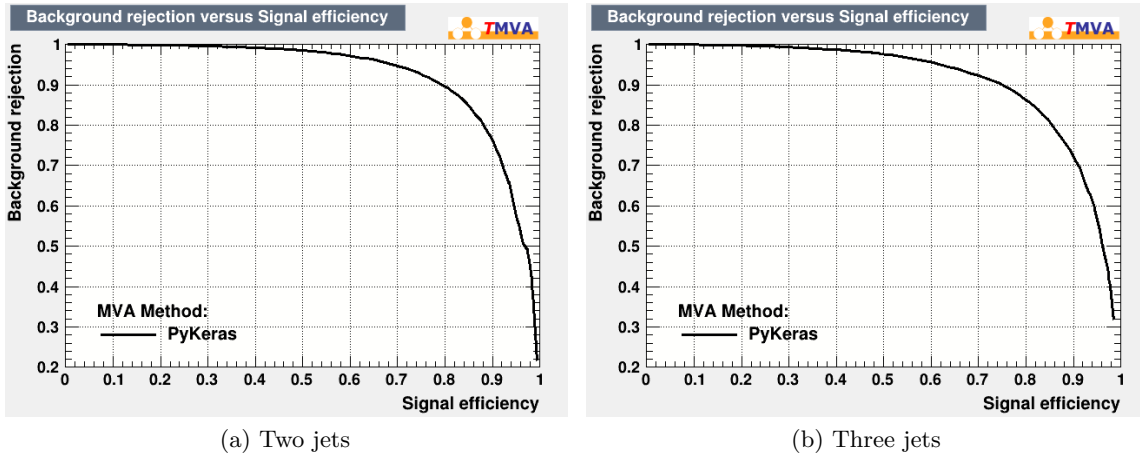
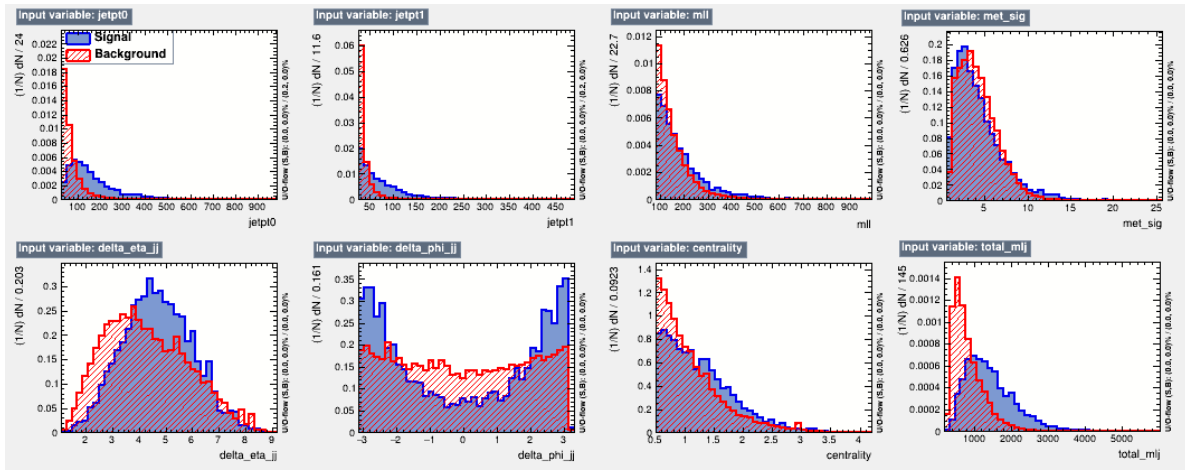


Figure 9.7: Background rejection vs signal efficiency for $n_{jets} = 2$ (left) and $n_{jets} = 3$ (right) in the signal region.

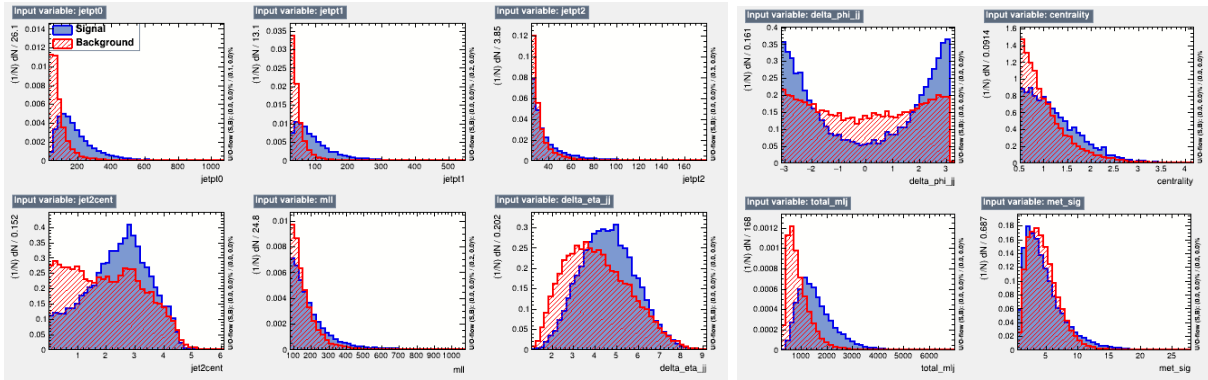
It is useful to look at the normalized distributions of the input variables for both networks and compare signal and background. Figure 9.8 shows this comparison. A few observations can be made:

- The leading and sub-leading jet p_T (and to some extent the third jet p_T) tends to be higher for signal events compare to background events.
- The separation between the two jets tends to be larger for signal events as well (as can be seen by looking at $\Delta\eta_{jj}$ and $\Delta\varphi_{jj}$).
- Centrality is also on average larger for signal events.
- The same is true for the total invariant mass of the two leading jets and the two leptons.

All of this corresponds to what is expected for vector boson scattering events. Jets-related observables are particularly useful to distinguish signal from background.



(a) Two jets



(b) Three jets

Figure 9.8: Normalized distribution of input variables for $n_{\text{jets}} = 2$ (above) and $n_{\text{jets}} = 3$ (below) in the signal region.

Also of interest is to look at the correlation between the input variables used by the neural networks. This is a useful check as highly correlated variables could indicate that one or more input

variables can be removed. Figure 9.9 shows the input correlations for each network, and for both signal and background samples. Correlations are overall fairly small, with only a few examples of correlations of about $\approx 60\%$.

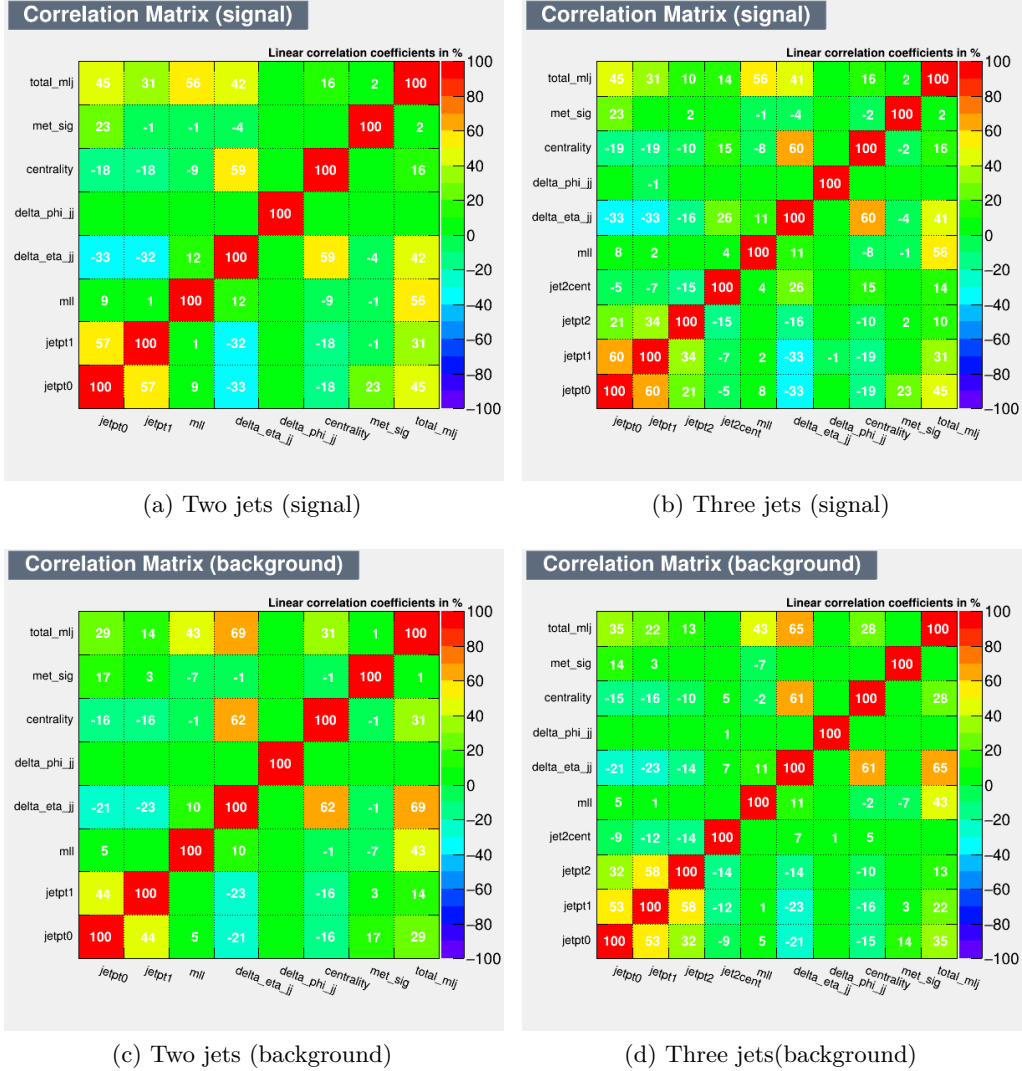


Figure 9.9: Input variables correlation matrix for $n_{\text{jets}} = 2$ (left) and $n_{\text{jets}} = 3$ (right) in the signal region for the signal (top) and background (bottom) samples.

The initial choice of variables was made in a truth-level analysis (see [65] for more details) and was later refined, removing some variables with small impact ($\cos \theta^*$ and m_{jj}), splitting into two networks (for two and three jets) and adding third leading jet p_T and third jet centrality as input to the three-jets network.

It is useful to look at what input variables are contributing the most to the separation power of

the network. A score is attributed to each input variable, based on how the area under the ROC change when removing this variable from the network inputs. The score is defined as:

$$\text{score} = 1000 \times (AUC_{\text{nom}} - AUC) \quad (9.13)$$

where AUC_{nom} is the nominal area under the ROC curve, and AUC is the area under the ROC curve when the corresponding variable is removed.

Neural network training is not deterministic, so for each input variable the training is done twenty times and the average is used to compute the score. This avoids the possible issue of having a misleading score due to some fluctuation. This is also done with a different learning rate than the one used for the analysis (0.1 instead of 0.001) and with a batch size much larger than the nominal one (1024 instead of 32) in order to speed up the process. Tables 9.1 and 9.2 show the score for each input variables.

Variable	AUC	Score
Default	0.9266	0
Leading jet p_T	0.91745	9.15
Sub-leading jet p_T	0.91755	9.05
Total m_{lj}	0.9232	3.4
m_{ll}	0.9245	2.1
$\Delta\eta_{jj}$	0.92465	1.95
centrality	0.9251	1.5
$\Delta\phi_{jj}$	0.92595	0.65
E_T^{miss} significance	0.9264	0.2

Table 9.1: For each variable removed, the network (here for $n_{\text{jets}} = 2$) is re-trained 20 times and the AUC average is calculated. For each variable, a score is given, and higher scores indicate larger impacts. ‘Default’ means no variable is removed.

Variable	AUC	Score
Default	0.9071	0
Sub-leading jet p_T	0.897	10.01
Leading jet p_T	0.89815	8.95
Third jet centrality	0.9002	6.9
$\Delta\eta_{jj}$	0.9051	2.0
centrality	0.90575	1.35
$\Delta\phi_{jj}$	0.906	1.01
m_U	0.906	1.01
E_T^{miss} significance	0.90705	0.05
Third jet p_T	0.907	0.01
Total m_{lj}	0.907	0.01

Table 9.2: For each variable removed, the network (here for $n_{\text{jets}} = 3$) is re-trained 20 times and the AUC average is calculated. For each variable, a score is given, and higher scores indicate larger impacts. ‘Default’ means no variable is removed.

9.3.5 Neural network settings

The two-jets and three-jets networks are trained similarly. A maximum number of 1000 epochs is set, where only the epoch corresponding to the smallest value of the validation loss is saved. There are two hidden layers, one with 108 neurons and another one with 60 neurons. The learning rate is set to 0.001 and the batch size to 32. Those settings were set based on the performance of the network (by looking at the area under the roc curve), on how fast the network can be trained and on generally accepted best practices for machine learning. Those settings were first tuned for the truth-only analysis [65] and later refined.

The parameters that are set before the training (number of neurons, learning rate, dropout rate, ...) are called *hyperparameters*. It should be mentioned that ideally hyperparameters optimization should be done using a separate dataset to avoid any bias (so there should be a training dataset, a testing dataset and a validation dataset). TMVA developers chose to merge the testing and validation dataset to increase statistics as the bias introduced is in practice often negligible [125].

Table 9.3 shows a summary of the various settings used in the neural networks.

Samples	signal (EWK W^+W^-) and background (top, QCD W^+W^-) Split: 50% training, 50% testing and validation
Regularization	Dropout with rate = 0.1
Keras Model	Feed-forward neural network with two hidden layers Number of neurons in the first hidden layer = 108 Number of neurons in the second hidden layer = 60
Activation function	Rectified linear unit (all layers except output) and softmax function on the output layer.
Loss Function	Binary cross entropy
Optimizer	Mini batch gradient descent with learning rate = 0.001
Variable Transformation(s)	Gaussianization
Number of Epochs	≤ 1000 (maximum)
Batch Size	32

Table 9.3: Configuration of the neural network constructed using TMVA in conjunction with Keras.

9.3.6 Neural network cutflow

No cut is applied to the neural network distribution to define the signal region. The neural network distribution is used to perform a likelihood fit to extract the signal strength and compute the significance (see chapter 11 for more details). Nonetheless, it can be interesting to look at the fraction (and total number) of signal and background events. Table 9.4 shows the fractions of Monte Carlo events that are signal events or part of one of the backgrounds. ‘ WW ’ refers to the QCD production of two opposite signs W bosons and ‘multibosons’ corresponds to WZ , ZZ , WWW , WWZ and ZZZ productions. The numbers shown here result from the combination of the two-jets and three-jets networks. Table 9.5 shows the same information but shows the number of predicted events instead of the percentage.

NN cut	EWK osWWjj	top	WW	W+jets	Z+jets	multibosons
0.0	2.44 %	67.37 %	21.97 %	2.97 %	3.60 %	1.66 %
0.1	4.14 %	68.49 %	20.75 %	2.48 %	2.41 %	1.73 %
0.2	6.31 %	67.91 %	19.89 %	2.42 %	1.71 %	1.77 %
0.3	8.30 %	66.26 %	19.41 %	2.65 %	1.62 %	1.75 %
0.4	10.52 %	64.89 %	18.96 %	2.60 %	1.31 %	1.72 %
0.5	12.98 %	63.18 %	18.22 %	2.78 %	1.19 %	1.66 %
0.6	16.04 %	60.88 %	17.41 %	3.06 %	1.06 %	1.55 %
0.7	20.36 %	58.44 %	16.32 %	2.60 %	0.89 %	1.39 %
0.8	27.27 %	53.18 %	14.54 %	2.65 %	1.05 %	1.31 %
0.9	41.35 %	43.25 %	11.44 %	2.30 %	0.63 %	1.04 %
0.95	56.53 %	32.08 %	8.52 %	1.39 %	0.63 %	0.85 %

Table 9.4: Neural network cutflow.

NN cut	EWK osWWjj	top	WW	Wjets	Zjets	multibosons
0.0	176.581	4881.75	1591.77	214.976	260.664	120.012
0.1	172.89	2858.29	865.89	103.30	100.69	72.33
0.2	165.74	1784.74	522.61	63.50	44.98	46.59
0.3	159.72	1274.88	373.50	51.00	31.18	33.64
0.4	153.67	948.06	277.10	37.97	19.16	25.14
0.5	147.13	716.08	206.46	31.46	13.51	18.79
0.6	139.10	528.04	150.97	26.52	9.22	13.47
0.7	128.81	369.69	103.24	16.48	5.64	8.78
0.8	113.97	222.30	60.80	11.06	4.39	5.49
0.9	87.40	91.43	24.18	4.85	1.33	2.19
0.95	61.28	34.77	9.24	1.51	0.69	0.93

Table 9.5: Neural network cutflow.

The top and QCD WW backgrounds largely dominate, while the other backgrounds add up to only a few percent. That is why the networks are only trained with the top and QCD WW backgrounds.

9.3.7 Validation

Careful validation needs to be done in order to trust the neural network output. In particular, looking at low values of the neural network is useful as it can be safely compared to real data, because this is where the background completely dominates.

A validation region was defined to be the same as the signal region but with an extra cut on the neural network output: NN output < 0.6 . Several checks were made. One of them is to compare the correlations between input variables with the same correlations in data. Figure 9.10 shows a few examples of these correlation plots. We can see an overall good agreement.

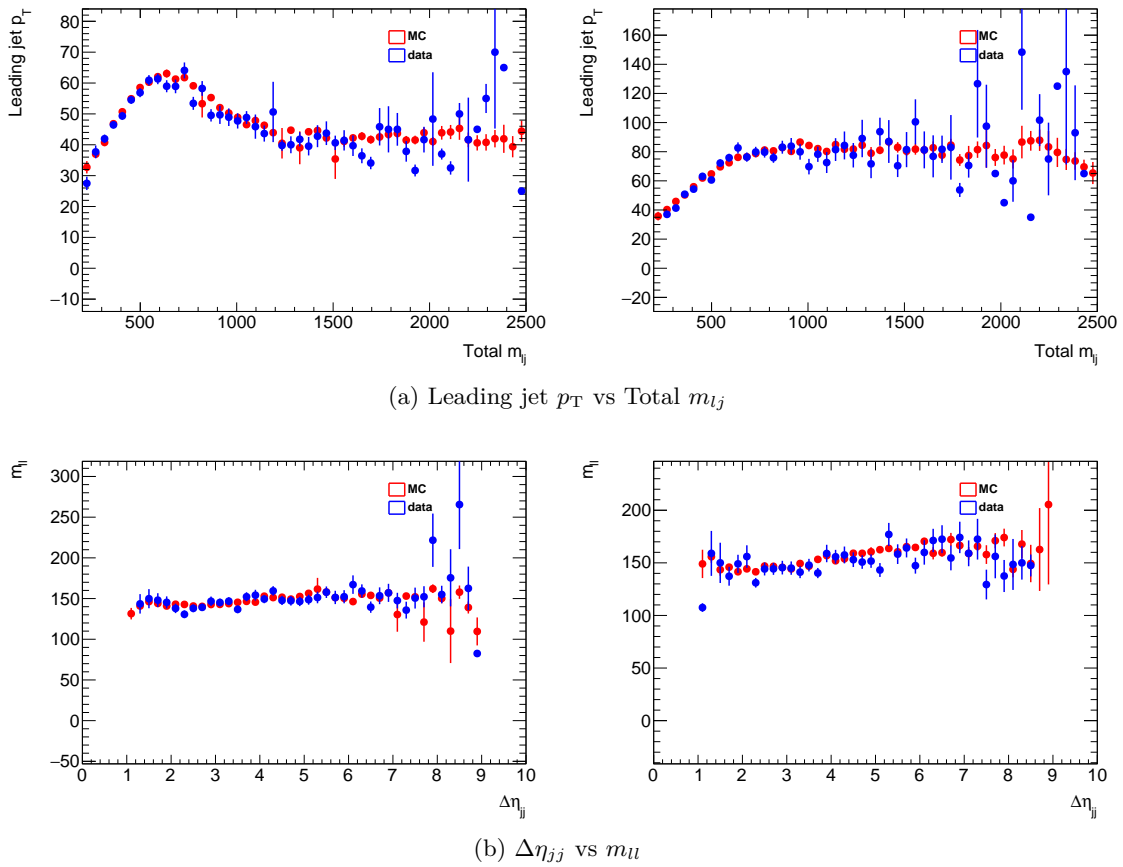


Figure 9.10: Correlation between total m_{lj} and leading jet p_T (above) and correlation between $\Delta\eta_{jj}$ and m_{ll} (below) for low NN output (NN output < 0.6) for $n_{\text{jets}} = 2$ (left) and $n_{\text{jets}} = 3$ (right) in the validation region. The error bars correspond to the standard error on the mean and do not include systematic uncertainties.

Another important check is to look at the correlations between the input variables and the neural network output. Figures 9.11 and 9.12 show these correlations for the two-jets and three-jets

regions, with data in the validation region and Monte Carlo events in the full signal region.

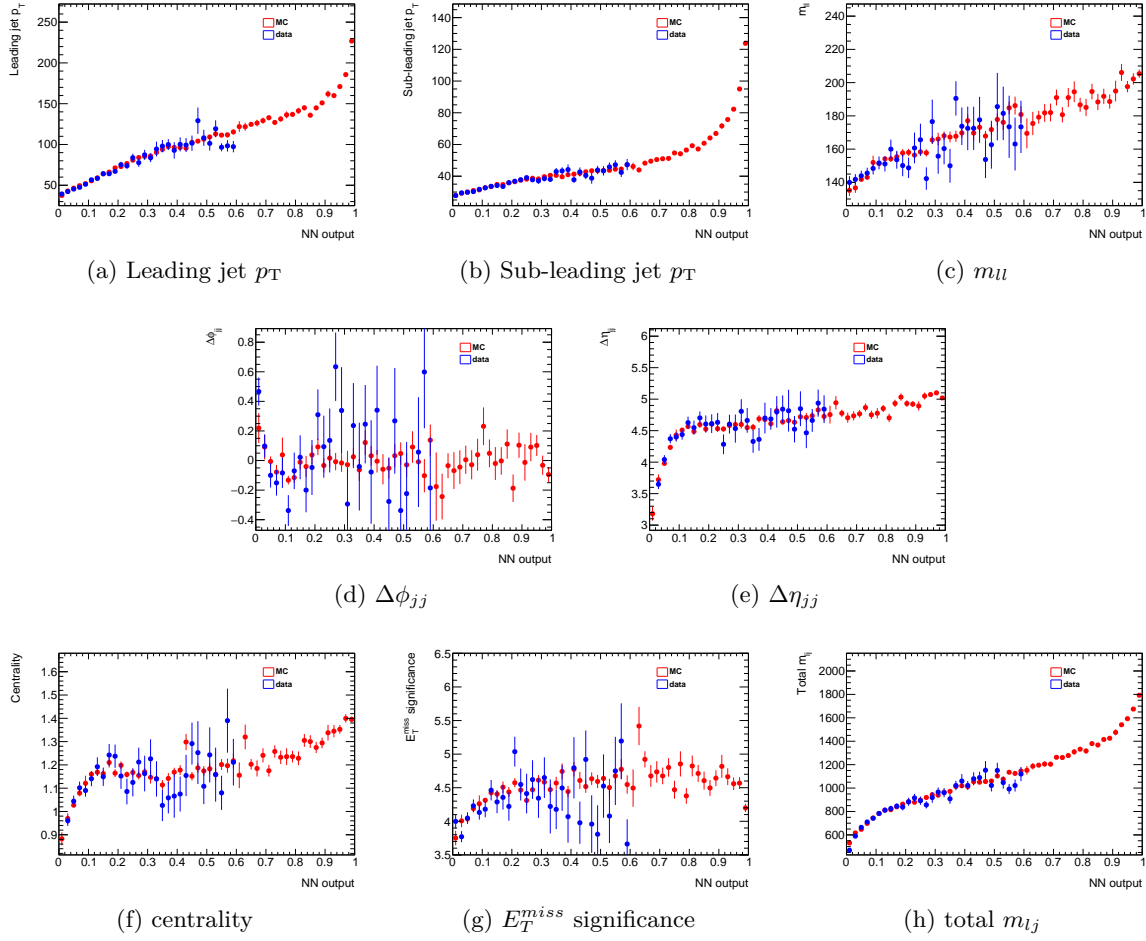


Figure 9.11: Correlation between variables used for training the NN output for $n_{\text{jets}} = 2$ in the signal region. The error bars correspond to the standard error on the mean and do not include systematic uncertainties.

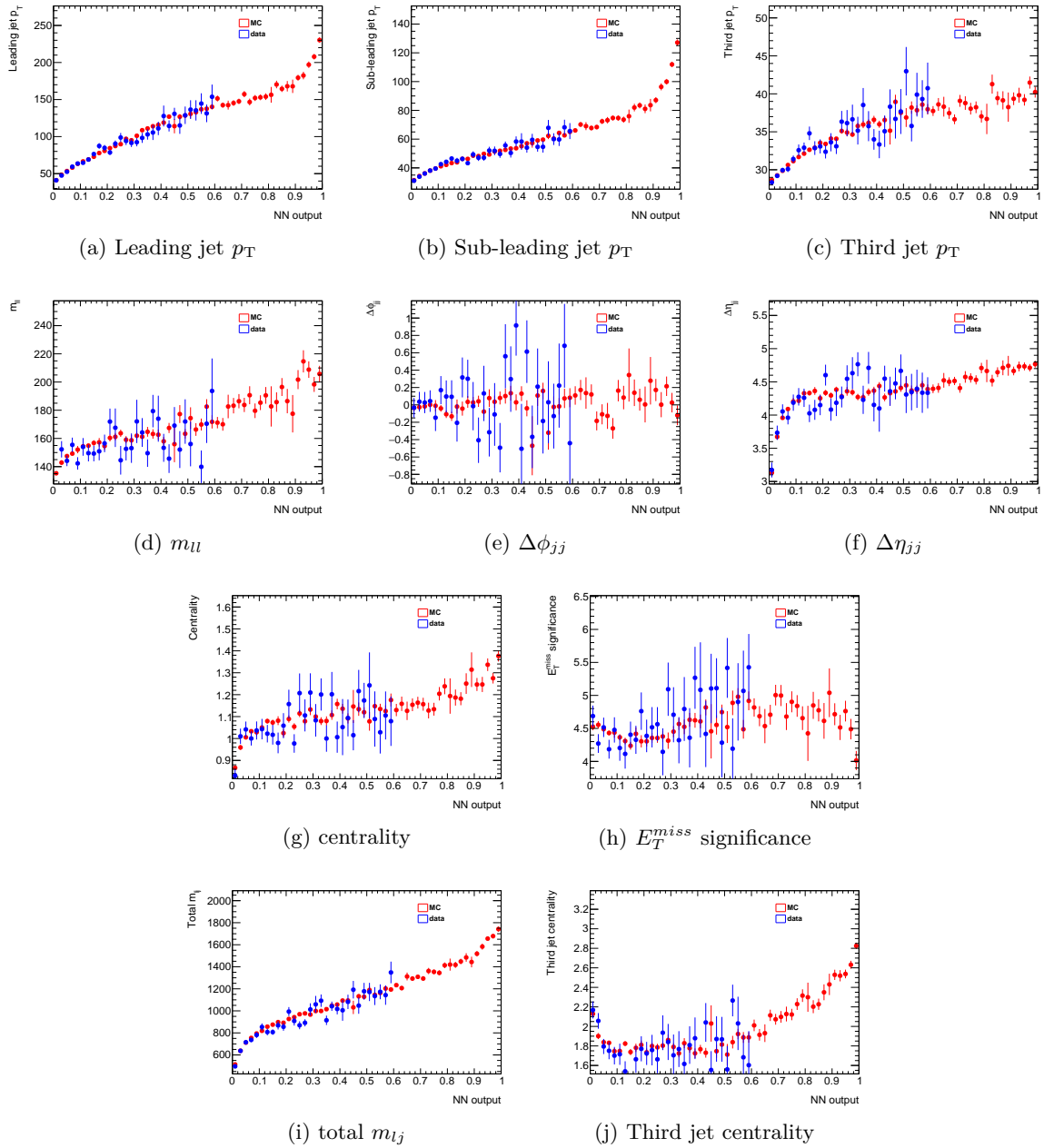


Figure 9.12: Correlation between variables used for training the NN output for $n_{\text{jets}} = 3$ in the signal region. The error bars correspond to the standard error on the mean and do not include systematic uncertainties.

It is also useful to check to make a basic data/Monte Carlo prediction comparison in the validation region. Figures 9.13 and 9.14 show this comparison for $n_{\text{jets}} = 2$ and $n_{\text{jets}} = 3$ and for all the observables used as input for the neural networks. We observe a good agreement with data.

All of these checks in the validation region give us good confidence that the input variables used

by the neural networks are well modeled by the Monte Carlo samples.

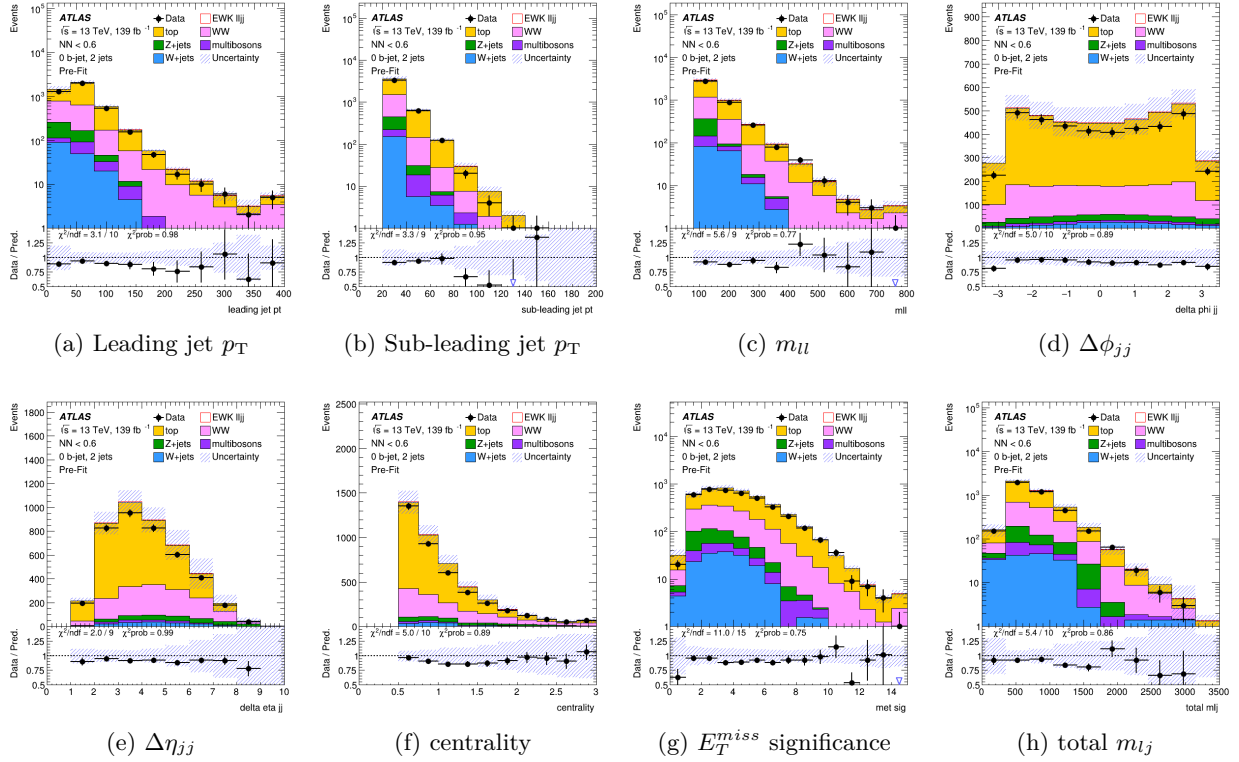


Figure 9.13: Validation region plots for $n_{\text{jets}} = 2$. We can observe a good agreement with data. All uncertainties are included.

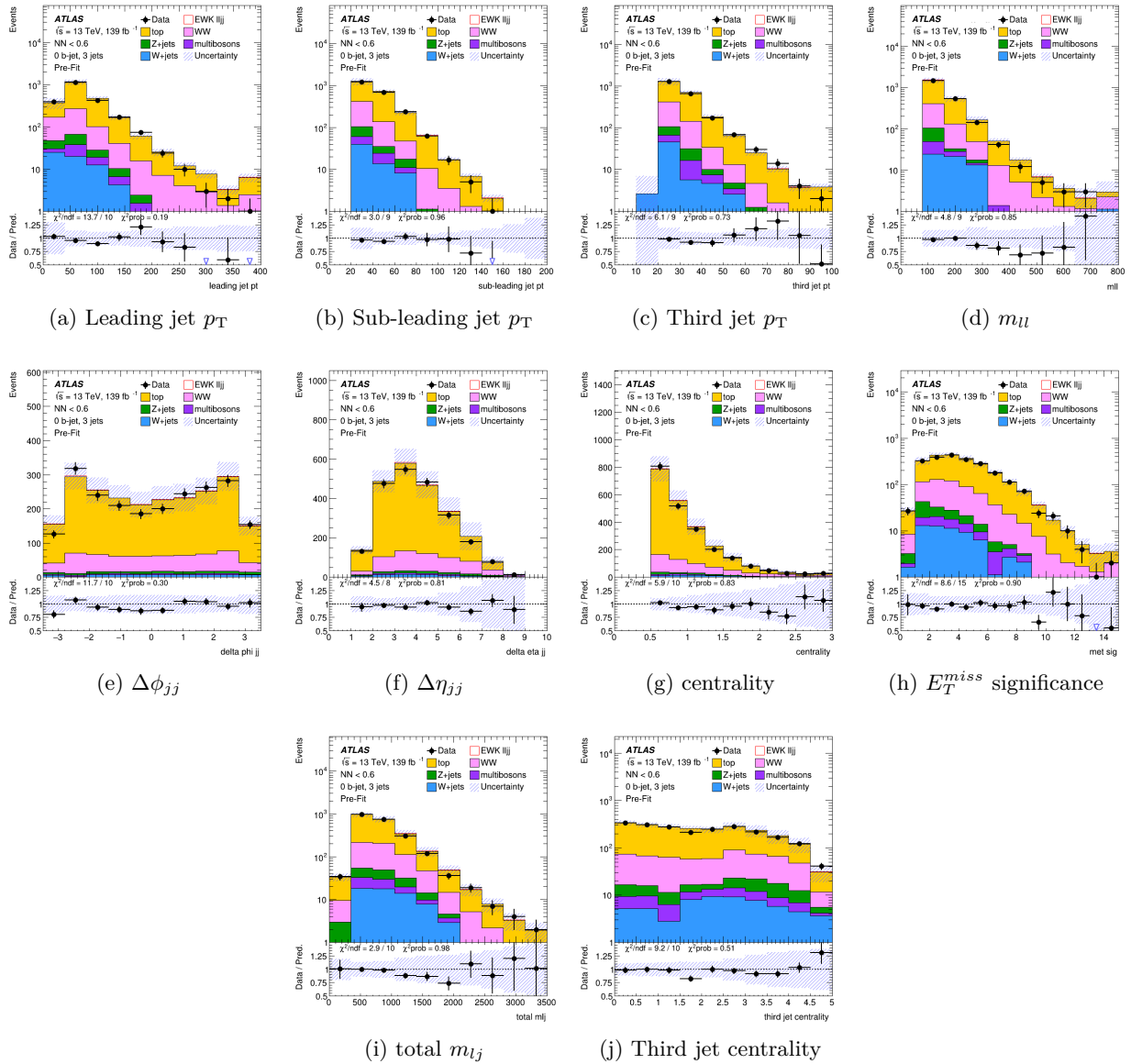


Figure 9.14: Validation region plots for $n_{\text{jets}} = 3$. We can observe a good agreement with data. All uncertainties are included.

9.3.8 Control region

As explained in chapter 8, a control region is defined where exactly one of the two leading jets is required to be a b-jet. It is also good to check the agreement in this region. Figures 9.15 and 9.16 show the Monte Carlo/data comparisons in the control region for $n_{\text{jets}} = 2$ and $n_{\text{jets}} = 3$ respectively.

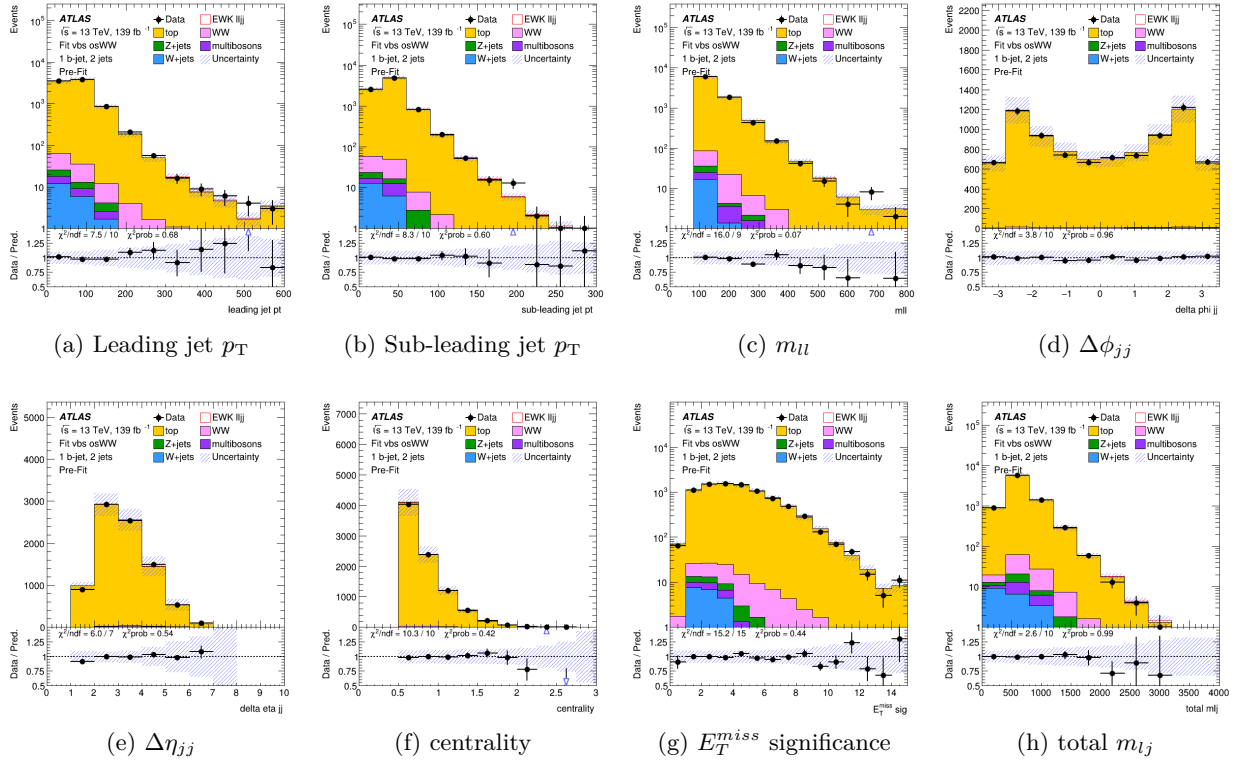


Figure 9.15: Control region plots for $n_{\text{jets}} = 2$. All uncertainties are included. A good agreement with data can be observed.

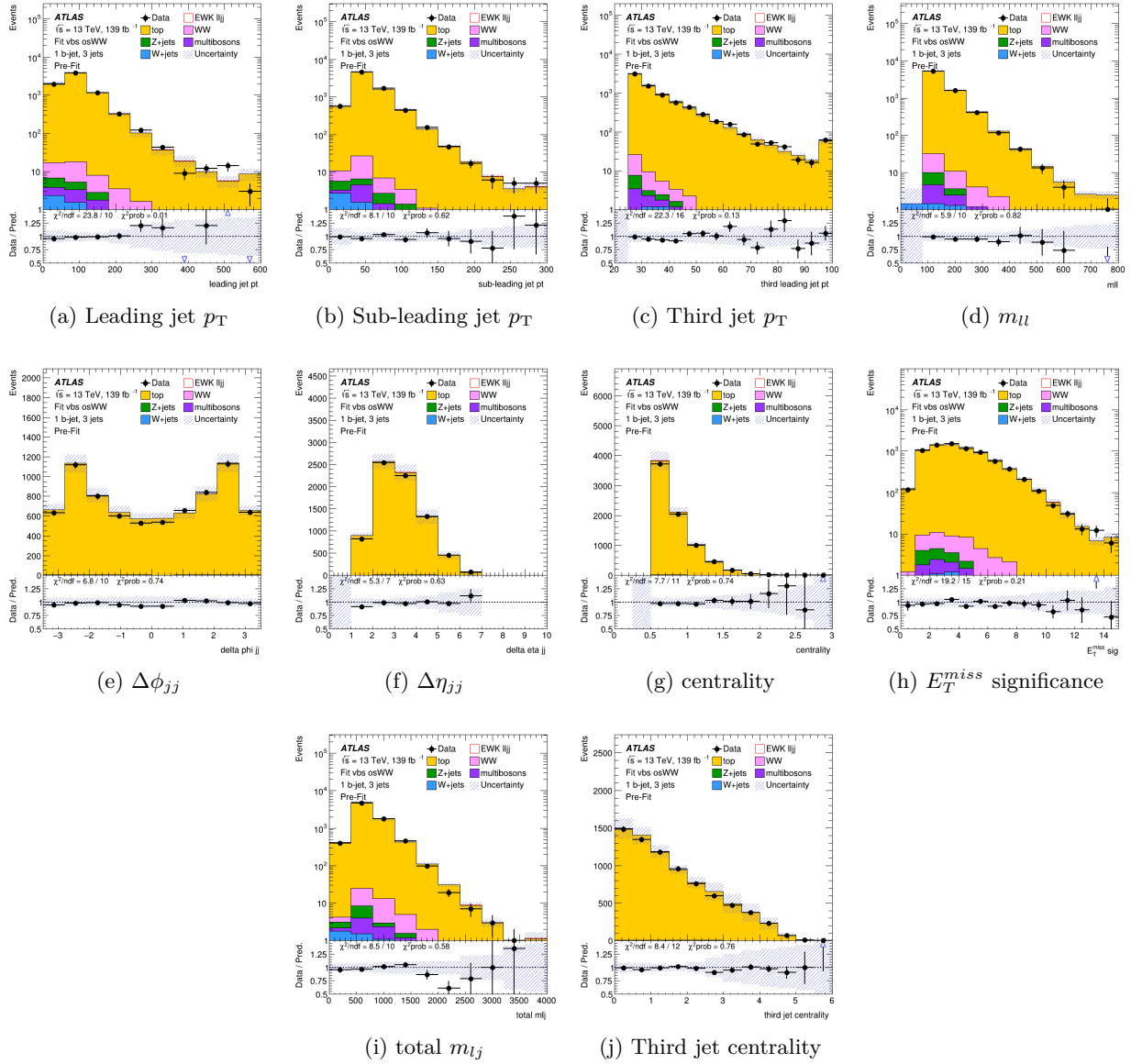


Figure 9.16: Control region plots for $n_{\text{jets}} = 3$. All uncertainties are included. A good agreement with data can be observed.

9.3.9 High NN region

It is also useful to look at the ‘high neural network output’ region, defined by adding a cut on the signal region (NN output > 0.9). The signal region is defined with fairly loose cuts to allow the neural network to do most of the work itself. This avoids the possibility of cutting too harshly and potentially lose statistics. Figure 9.17 shows the neural network output distribution for the input variables that are used by both networks and the invariant mass of the two leading jets m_{jj} .

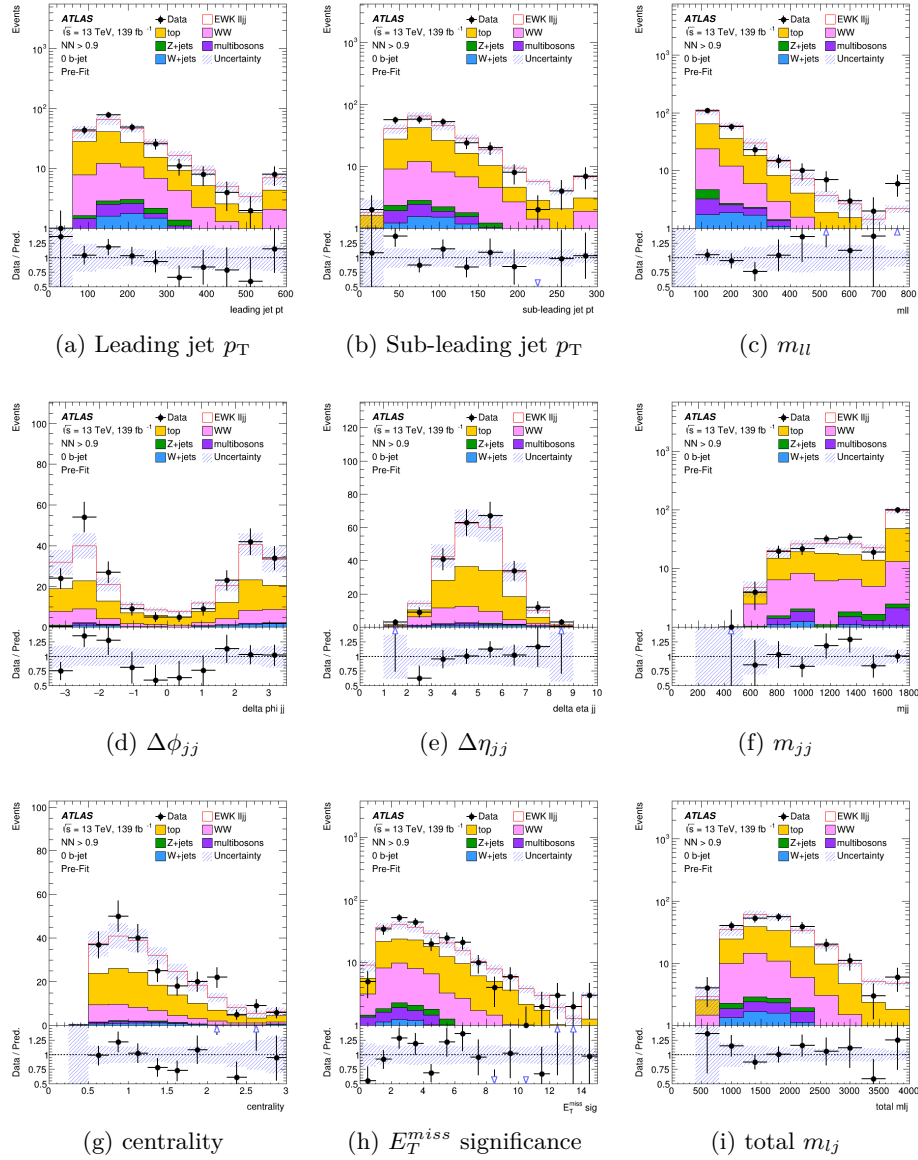


Figure 9.17: Observable in the signal region with a neural network > 0.9 cut. All uncertainties are included.

A few observations can be made:

- Events with leading jet $p_T < 100$ GeV and sub-leading jet $p_T < 50$ GeV are highly suppressed.
- No events with $m_{jj} < 500$ GeV pass the selection and events with $500 \text{ GeV} < m_{jj} < 750$ GeV are strongly suppressed.
- There are also no events with $\Delta\eta_{jj} < 2$ and events with small $\Delta\phi_{jj}$ are very suppressed.

This shows that cuts often made in vector boson scattering analysis (e.g. on $\Delta\eta_{jj}$ and m_{jj}) are done implicitly by the network itself. It is also noteworthy that the network is able to cut low

m_{jj} events off despite m_{jj} not being part of the inputs. This is because all the variables needed to compute m_{jj} are known by the network.

9.3.10 Input variables in the signal region

It is also useful to check the whole signal region without any cut on the neural network output. Figures 9.18 and 9.19 shows the input variables distributions in the signal region with a comparison with data. We can observe that, as expected, signal events are found mostly at the tail of p_T and invariant mass distributions.

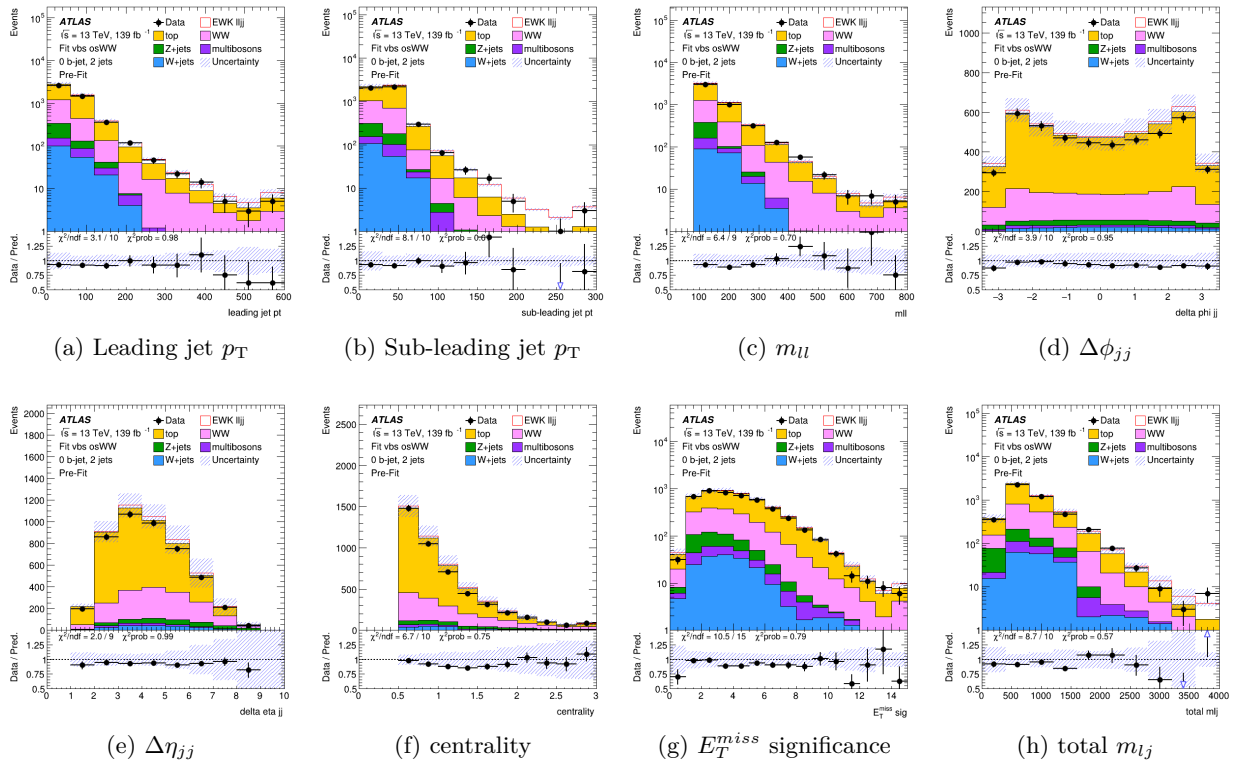


Figure 9.18: Observables used as input for the neural network in the signal region. All uncertainties are included.

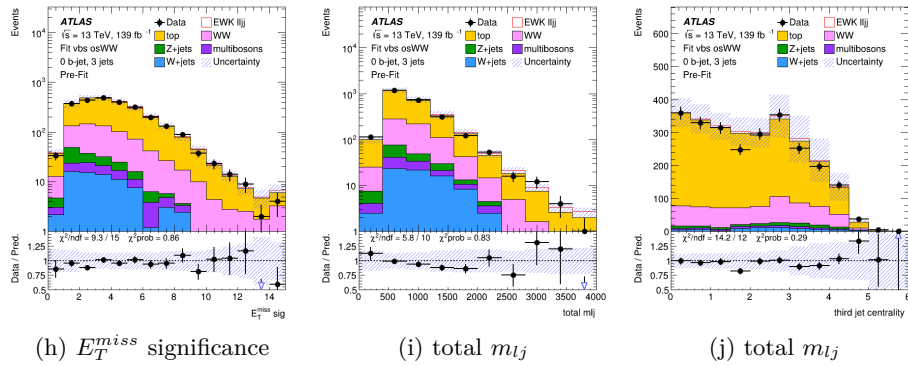
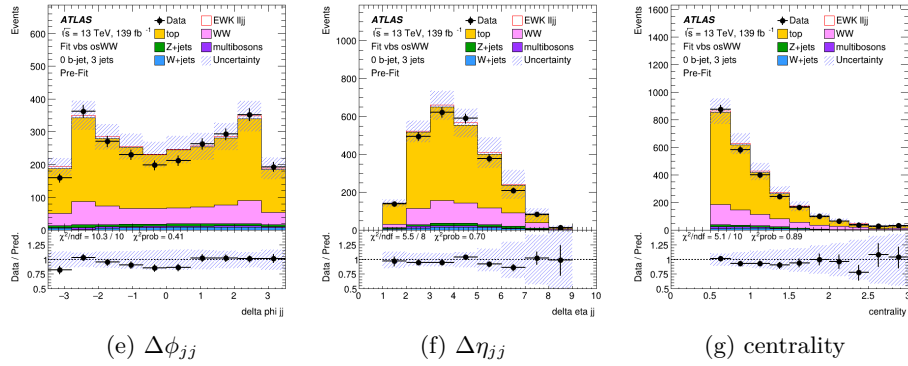
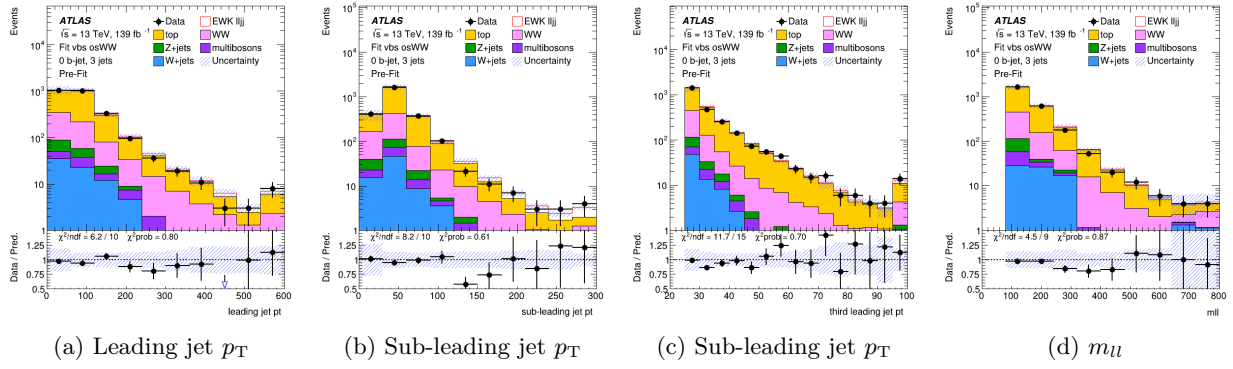


Figure 9.19: Observables used as input for the neural network in the signal region. All uncertainties are included.

Chapter 10

Uncertainties

10.1 Introduction

Various sources of uncertainties need to be taken into account in order to correctly estimate the total uncertainty on the signal strength (or on the cross-section). Besides the statistical uncertainty (discussed in chapter 11), there are also systematical uncertainties consisting of experimental uncertainties and theoretical uncertainties. Experimental uncertainties correspond to various detector calibration uncertainties and efficiency uncertainties, and theoretical uncertainties correspond to uncertainties when producing simulated events.

This chapter defines and discusses the main experimental and theoretical uncertainties for the EWK W^+W^-jj analysis affecting the total uncertainty on the signal strength.

10.2 Experimental uncertainties

The dominant experimental uncertainties come from the jet energy scale uncertainties, the b-tagging efficiency and the jet flavor composition uncertainty. Uncertainties related to electrons and muons (resolution, calibration of energy or momentum scale and isolation uncertainties), as well as uncertainties related to missing energy, pile-up and luminosity, are also taken into account but are overall smaller than the jet-related uncertainties.

The experimental uncertainties are estimated by varying a nuisance parameter (for example varying the jet calibration scale) by one sigma around the nominal value (up and down). Those uncertainties are propagated through the analysis and the final impact on the signal strength measurement is evaluated when the likelihood fit is performed (see chapter 11).

A summary and short descriptions of the experimental uncertainties can be found in tables 10.1 (for jets), 10.2 (for muons), 10.3 (for electrons) and 10.4 (for various other uncertainties). The complete and detailed list of experimental uncertainties can be found in appendix A.

Plots for signal (figure 10.1), top (figure 10.2) and QCD-WW backgrounds (figure 10.3) are shown for some of the main experimental uncertainties in the signal region. For each plot, the top

Experimental systematic uncertainties
Jet energy scale uncertainty for b jets
Jet energy scale uncertainty (for all jets)
Jet flavor composition uncertainty
Jet energy resolution uncertainty
JVT efficiency uncertainty
b-tagging efficiency uncertainties

Table 10.1: Jet experimental uncertainties.

Experimental systematic uncertainties
Muon momentum resolution uncertainty
Muon momentum scale uncertainty
Muon isolation working point scale factor uncertainty
Muon identification working point scale factor uncertainty

Table 10.2: Muon experimental uncertainties.

Experimental systematic uncertainties
Electron energy scale uncertainty
Electron identification efficiency uncertainty
Electron trigger efficiency uncertainty

Table 10.3: Electron experimental uncertainties.

part shows the nominal distribution as well as distributions for some of the main systematics while the bottom part shows the ratios with the nominal. The systematics shown on these plots are chosen to be among the (experimental) uncertainties that contribute the most to the final uncertainty (see next chapter about the likelihood fit). They are all jet-related uncertainties.

For the signal (figure 10.1), some fluctuations are seen, in particular for the jet energy resolution uncertainty, because of a few large weight events in the signal sample, due to the use of a born suppression factor. For events with a neural network output greater than about 0.8 (i.e. where the signal purity is relatively large), this is not an issue as there is enough statistic in the signal sample to smooth those variations out so the impact on the fit is negligible.

Systematic	Description
MET_SoftTrk_ResoPara	Track-based soft term related to longitudinal resolution uncertainty
MET_SoftTrk_ResoPerp	Track-based soft term related to transversal resolution uncertainty
MET_SoftTrk_Scale	Track-based soft term scale uncertainty
PRW_DATASF	Data scale factor uncertainty for pile-up reweighting
Luminosity	

Table 10.4: Various experimental uncertainties.

For the top background (figure 10.2), we can see that the uncertainty related to the b-tagging efficiency is smaller for larger values of the neural network output while the opposite is true for the Jet pile-up uncertainties. For the QCD-WW backgrounds (figure 10.3) uncertainties are smaller at larger neural network output. Overall, we can also observe that the overall experimental uncertainty for the signal is fairly small compared to the top and QCD W^+W^- backgrounds.

The same set of plots for the other backgrounds can be found in appendix B.

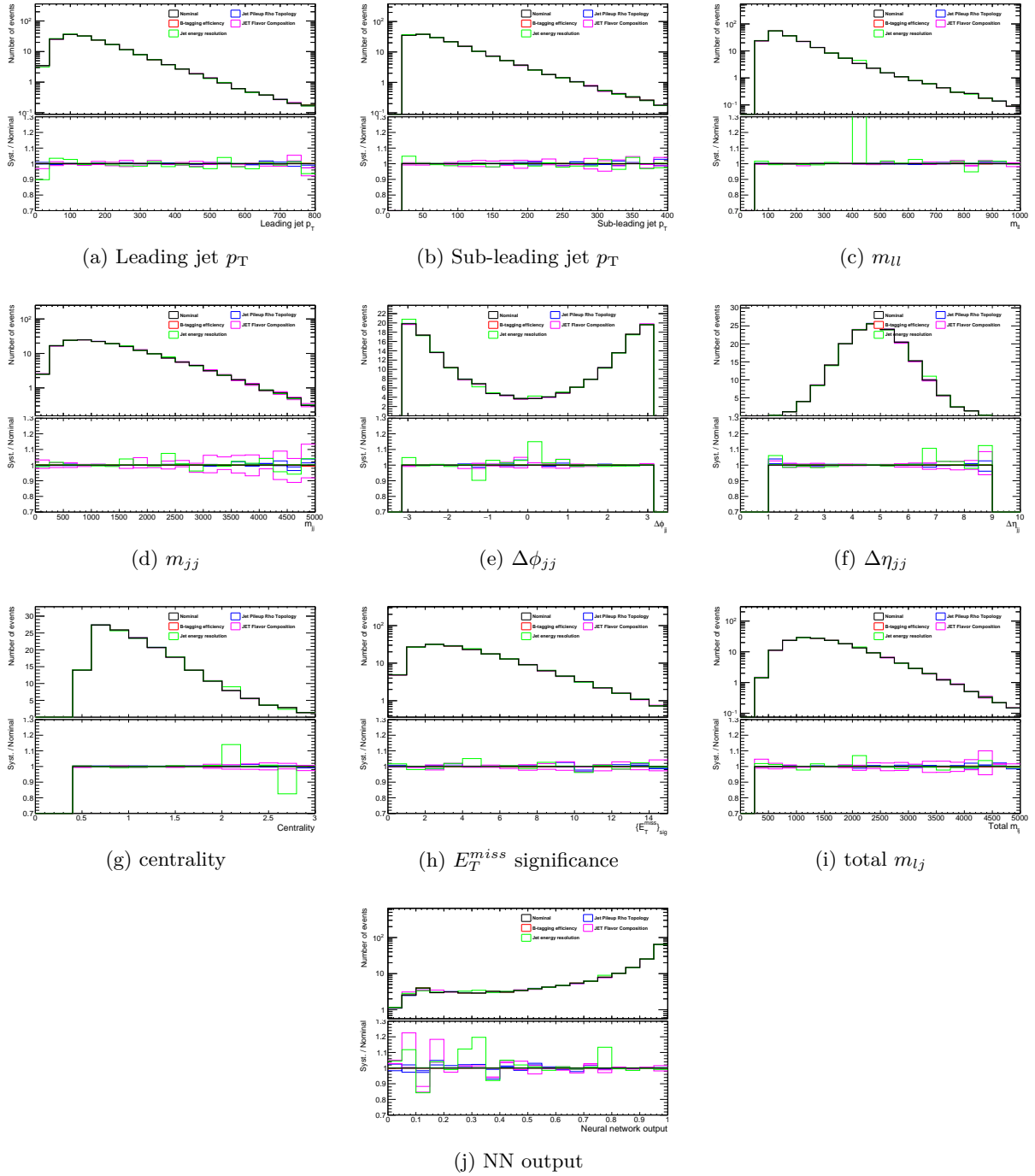


Figure 10.1: Distributions for the signal in the signal region for the main experimental uncertainties with the systematics/nominal ratio.

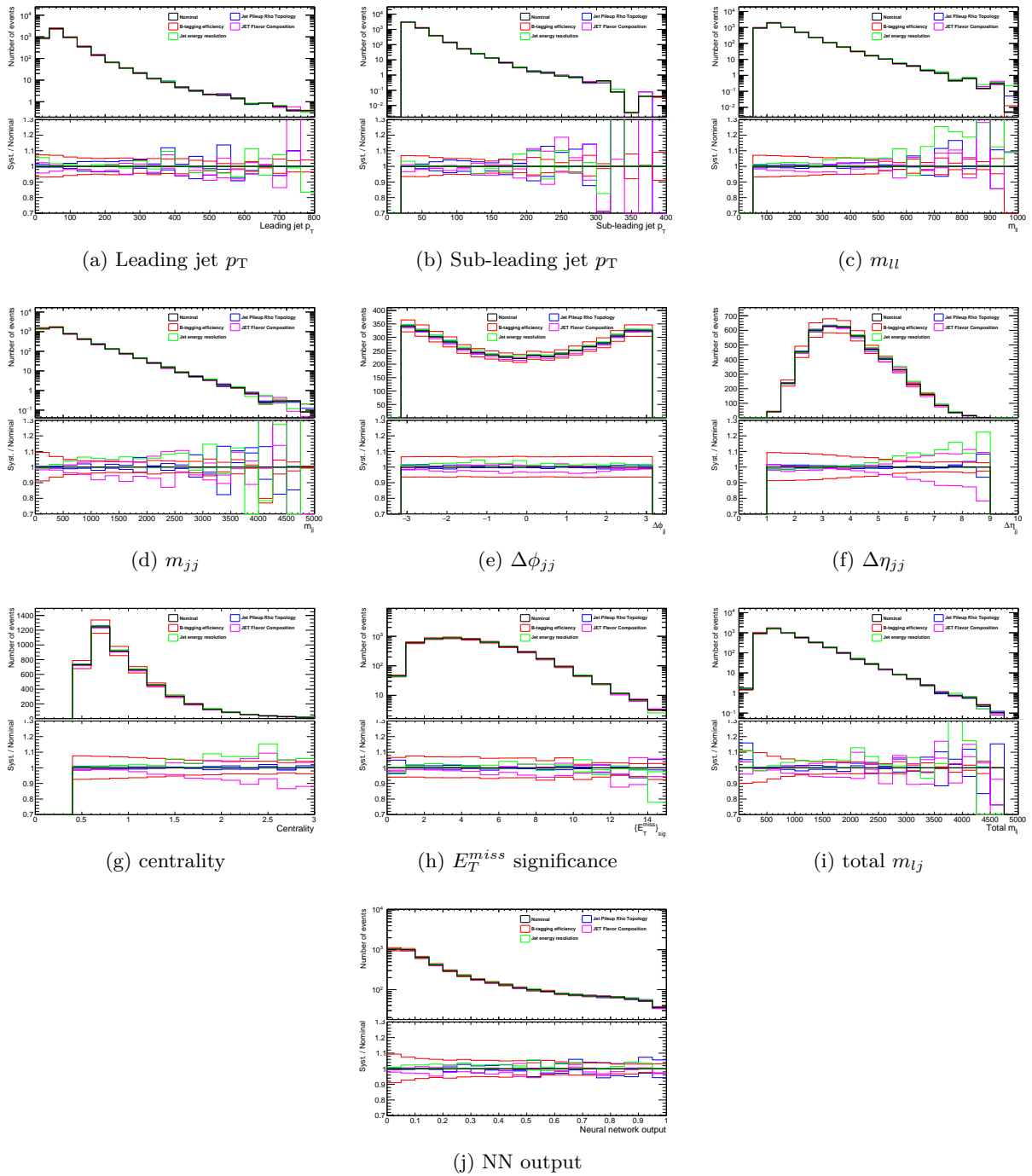


Figure 10.2: Distributions for the top background in the signal region for the main experimental uncertainties with the systematics/nominal ratio.

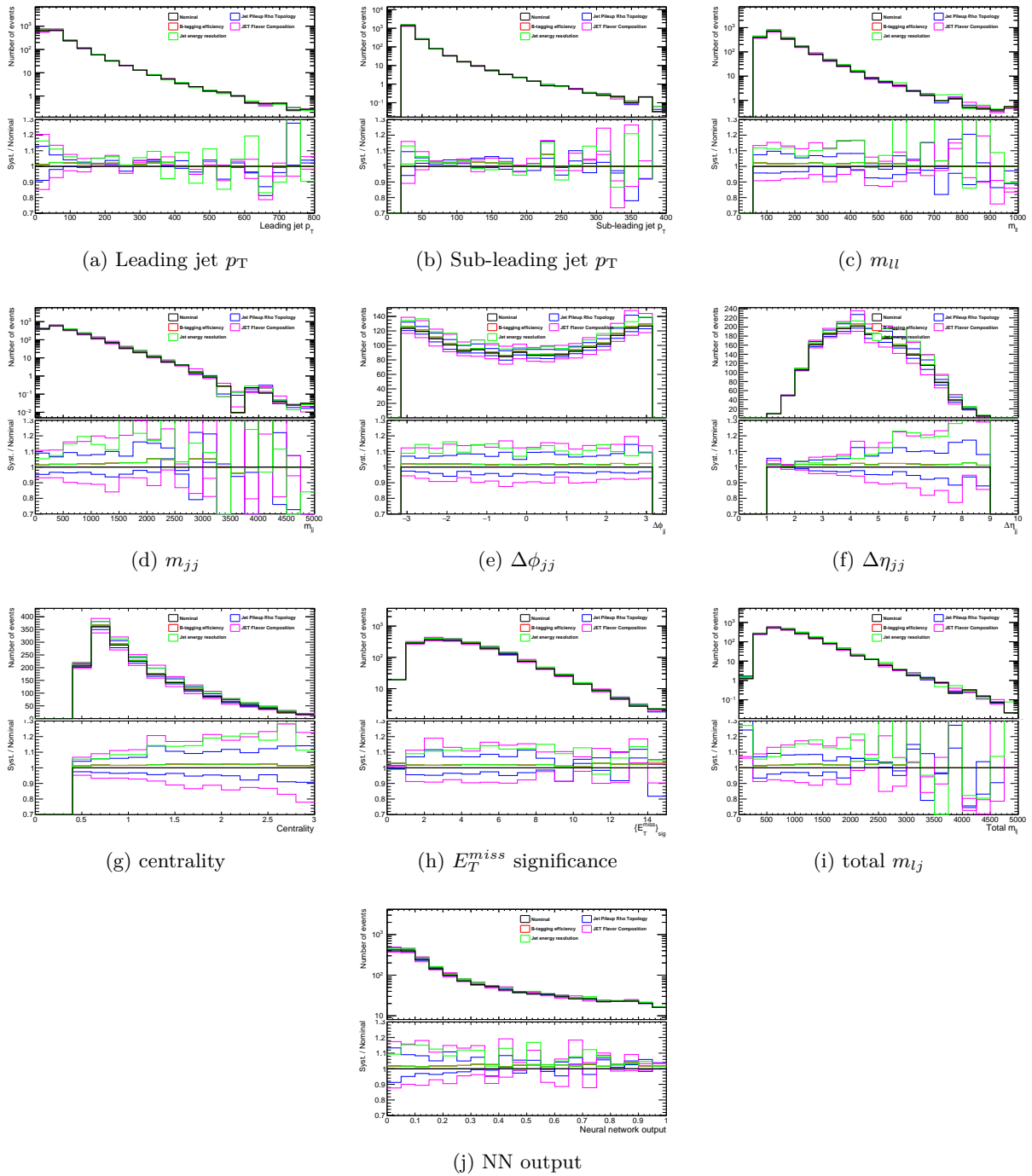


Figure 10.3: Distributions for the WW background in the signal region for the main experimental uncertainties with the systematics/nominal ratio.

10.3 Data-driven fake estimation

Another source of uncertainty is related to the fact that jets can be misidentified as leptons. Jets misidentified as leptons are called *fake leptons*. Fake leptons are why Z+jets and W+jets events can find their way into the signal region, even though an electron and a muon would not be expected in the final state.

The V+jets (Z+jets and W+jets) background is predicted to be very small using Monte Carlo but the Monte Carlo prediction is often unreliable for fake leptons estimation. A fake-enriched region was therefore defined by selecting leptons with loose identification and by requiring that at least one of the two leptons fails the isolation and the identification requirements, it is otherwise the same as the signal region. The idea is to have a region where the number of fake leptons is much larger than in the signal region, and compare simulation and data in this region to correct the V+jets prediction in the signal region. A *transfer factor* was defined in bins of the neural network output by:

$$\frac{N_{CR}^{\text{data}} - N_{CR}^{\text{not V+jets}}}{N_{CR}^{\text{V+jets}}} \quad (10.1)$$

where N_{CR}^{data} is the number of data events in the fake-enriched region, $N_{CR}^{\text{not V+jets}}$ the predicted (by Monte Carlo) number of events in the fake-enriched region that are not V+jets events and $N_{CR}^{\text{V+jets}}$ the (MC) predicted number of V+jets events (also in the fake-enriched region). The V+jets MC samples in the signal region are multiplied by the transfer factor. Figure 10.4 shows the neural network output in the fake-enriched region and 10.5 shows the binned transfer factor.

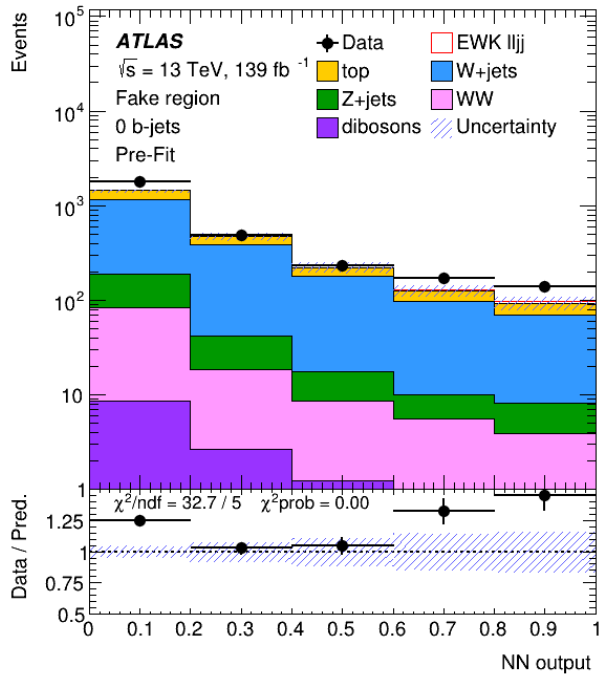


Figure 10.4: NN output in the fake-enriched region.

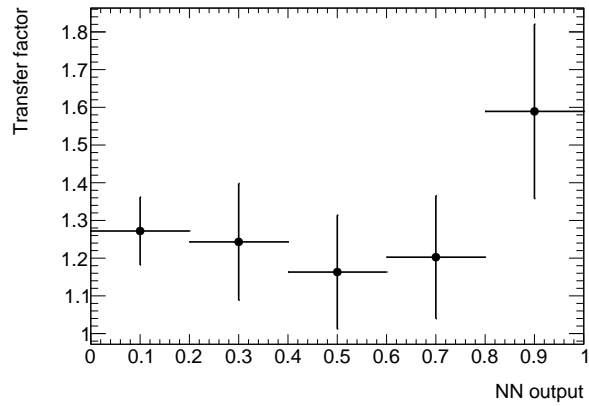


Figure 10.5: Transfer factor values for different NN output bins.

10.4 Theoretical uncertainties

10.4.1 Top theoretical uncertainties

For both the $t\bar{t}$ and the single top (Wt), we compare the nominal sample (POWHEG +PYTHIA) with alternatives samples/settings:

- Comparison with the alternative MADGRAPH5_aMC@NLO +PYTHIA sample: matrix element uncertainty.
- Comparison with the alternative POWHEG +Herwig 7 sample: parton shower uncertainty.
- The initial state radiation uncertainty is estimated by varying h_{damp} (parameter controlling the matrix element/parton shower in POWHEG), varying the strong coupling constant α_S^{ISR} used for initial state radiation and varying the renormalization and factorization scales.
- The final state radiation uncertainty is estimated by varying the strong coupling constant α_S^{FSR} used for final state radiation.

Another source of uncertainty comes from the Wt interference with $t\bar{t}$. To avoid overlap, single top samples are simulated with the $t\bar{t}$ contribution removed. There are two schemes to do that: the diagram removal (DR) and the diagram subtraction (DS) scheme, the former being the removal of $t\bar{t}$ at the amplitude level and the other at the cross-section level. The difference between the two schemes can be interpreted as the interference between $t\bar{t}$ and Wt . [128]. This uncertainty is estimated by comparing Powheg+Pythia with DR and Powheg+Pythia with DS schemes.

To take into account the cross-section uncertainty of the top background, the overall normalization of the top background is allowed to vary and is fitted when running the likelihood fit (see chapter 11).

The PDF uncertainty is estimated using matrix element level reweights for 100 PDF variations of the NNPDF30_nlo [98] set. This gives rise to a set of predictions (each one corresponding to a different PDF) and the standard deviation of this distribution is computed to set the uncertainty for each bin.

The α_S uncertainty is evaluated by setting $\alpha_S^{\text{down}} = 0.117$ and $\alpha_S^{\text{up}} = 0.119$ (the nominal value used being 0.118). Two PDFs exist that have been determined using these values, there are called NNPDF30_nlo_as_0119 and NNPDF30_nlo_as_0117. They are used to compute the α_S uncertainty, which is then given by: $\delta\alpha_S = \frac{\sigma(\alpha_S^{\text{down}}) - \sigma(\alpha_S^{\text{up}})}{2}$.

A comparison is also made with a PDF that was not produced by the NNPDF collaboration (the ‘inter-PDF’ uncertainty). The comparison is made with a PDF called: MMHT2014nlo68clas118 PDF taken from the MMHT 2014 set of PDFs [129].

Figure 10.6 shows the effects of these systematic uncertainties for various observables.

10.4.2 Signal theoretical uncertainties

For the signal sample, the uncertainties are the renormalization and factorization scale uncertainties, the PDF uncertainty, the initial and final state radiation uncertainties and the parton shower uncertainty.

To estimate the renormalization and factorization scale uncertainty, the renormalization and factorization scale are varied by either multiplying or dividing them by a factor of two. We then take the envelope of $\{\mu_r, \mu_f\} \times \{2, 1\}, \{1, 2\}, \{1, 0.5\}, \{0.5, 1\}, \{0.5, 0.5\}$ to compute the total uncertainty.

The PDF uncertainty is estimated using matrix element level reweights for 100 PDF variations of the NNPDF30_nlo [98] set and taking the standard deviation as an uncertainty bin-by-bin. The α_S uncertainty is evaluated by using NNPDF30_nlo_as_0119 and NNPDF30_nlo_as_0117 with $\delta\alpha_S = \frac{\sigma(\alpha_S^{\text{down}}) - \sigma(\alpha_S^{\text{up}})}{2}$. An inter-PDF uncertainty is added by comparing with the MMHT2014nlo68clas118 [129] and the CT14nlo [130] PDFs.

The initial and final state radiation uncertainty is estimated similarly to what is done for the top initial and final state radiation uncertainty.

Finally, the parton shower uncertainty is estimated by comparing the nominal setting (POWHEG-BOX + PYTHIA) with an alternative sample produced with POWHEGBOX + Herwig.

10.4.3 QCD-WW theoretical uncertainties

The PDF uncertainty for the QCD $qq \rightarrow WW$ background is estimated using matrix element level reweights for 60 PDF variations of the CT10nlo set and taking the standard deviation as an uncertainty bin-by-bin. The renormalization and factorization scale uncertainties are also estimated, using the same procedure as the one signal for the signal uncertainty.

The other backgrounds are much smaller and the theoretical uncertainties are not computed for those.

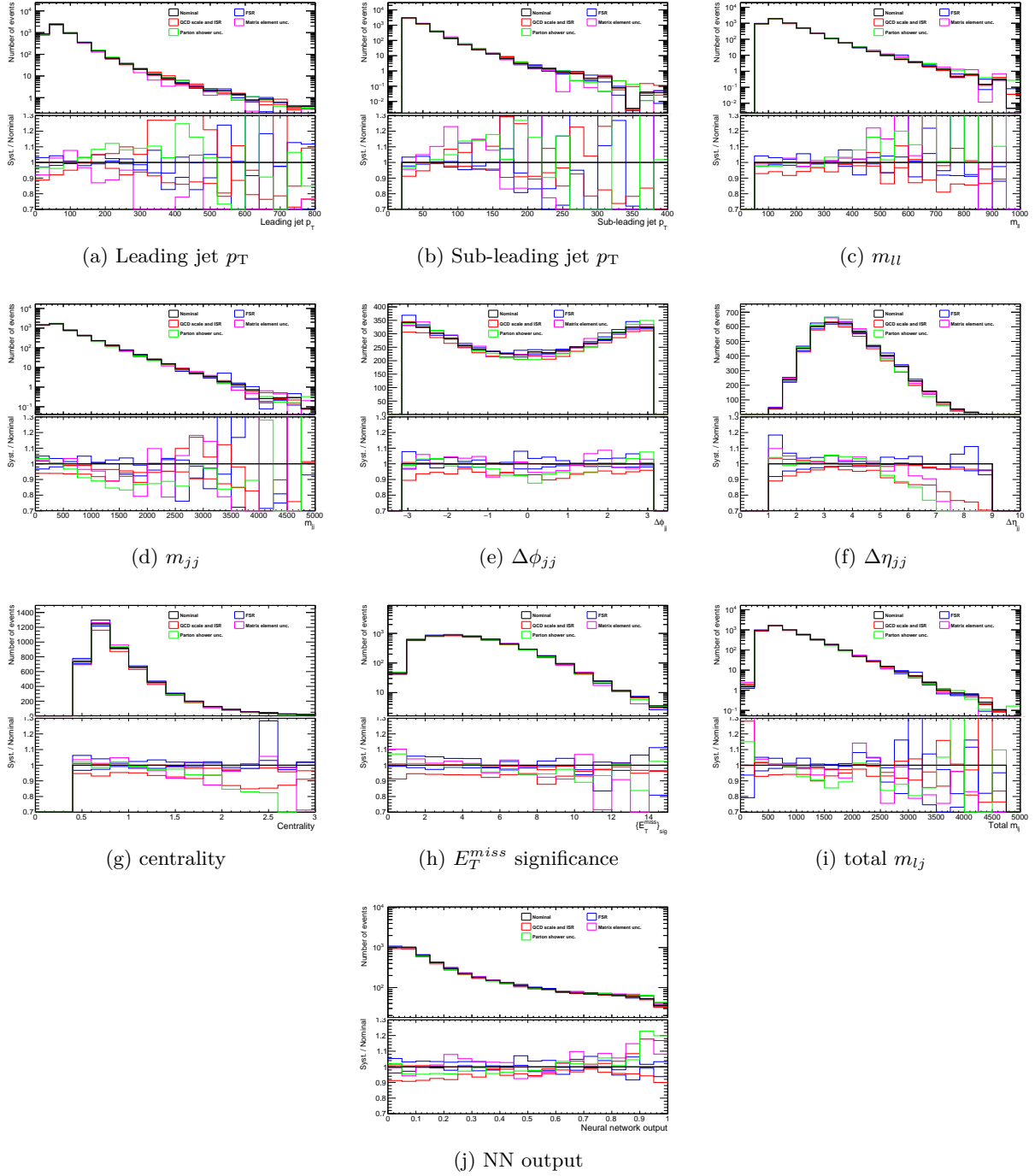


Figure 10.6: Distributions for the top background in the signal region for the main theoretical uncertainties with the systematics/nominal ratio.

10.4.4 QCD-EWK WW production interference

Another source of uncertainty is the interference between QCD and EWK W^+W^- production. A study was made, at truth level, by producing events (using MADGRAPH5_aMC@NLO) for the EWK W^+W^- production, and (in a separate sample) ‘events’ corresponding to the interference term only. This can be done with MADGRAPH5_aMC@NLO by requiring matrix elements squares to be of order α_S^2 . The following ratio is then computed bin-by-bins for several observables:

$$\frac{\text{Interference} + \text{Signal}}{\text{Signal}} \quad (10.2)$$

Where ‘Interference’ and ‘Signal’ are the predicted event yields for the interference term and the EWK W^+W^- production respectively. An example is given in figure 10.7 for the invariant mass of the two leading jets m_{jj} .

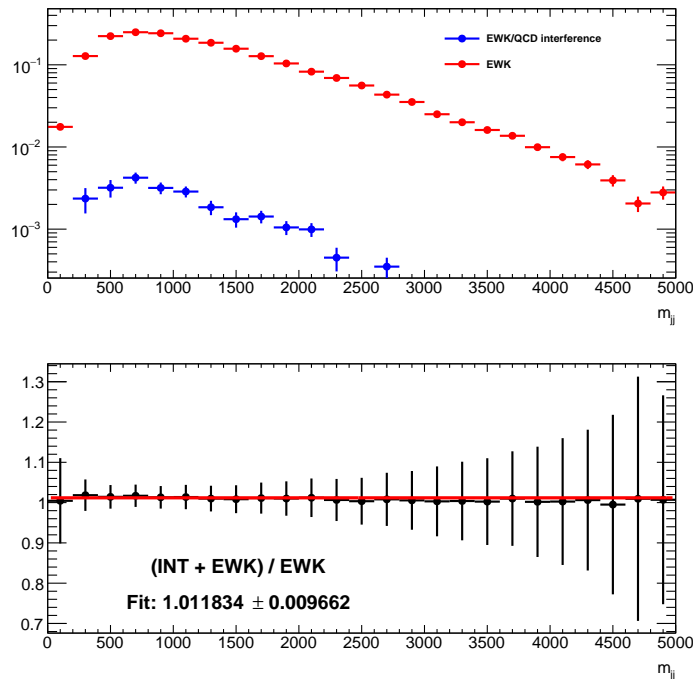


Figure 10.7: Ratio plot comparing the events from the signal and the interference. Plot produced by Rabia Omar.

Chapter 11

Likelihood Fit

11.1 Introduction

Statistical analysis is a central part of any particle physics analysis. It is through careful statistical analysis that we can objectively test various theories and models or make precise measurements with a rigorous estimation of the corresponding uncertainties.

This chapter starts with a short introduction to probabilities and to the notion of *profile likelihood fit*. The second part of the chapter describes how the likelihood fit is performed for this present analysis as well as the results obtained.

11.1.1 Probabilites

Consider two hypotheses H_0 and H_1 , for example, H_0 could be the Standard Model and H_1 a BSM model. Let n be the number of observed events in a certain phase space region. We would like to know the probability that the hypothesis H_1 is true given n . The answer is given by the *Bayes theorem* [131, 132]:

$$P(H_1|n) = \frac{P(n|H_1)P(H_1)}{P(n|H_0)P(H_0) + P(n|H_1)P(H_1)} \quad (11.1)$$

where $P(H_0)$ and $P(H_1)$ are *prior probabilities*, they are the probabilities that we assign, *before making any measurement*, that hypotheses H_0 and H_1 is true. The probability that we obtain, $P(H_1|n)$, is called a *posterior probability*.

Bayes theorem can also be used with probability densities. For example, if we want to measure the cross-section of a particular process, the number of events will be:

$$n = \mu s + b \quad (11.2)$$

where s is the number of signal events, b is the number of background events, and μ is a parameter called the *signal strength*. $\mu = 1$ corresponds to a particular model (e.g. the Standard Model) and

$\mu = 0$ corresponds to the background-only hypothesis. Bayes theorem would be written:

$$p(\mu|n) = \frac{p(n|\mu)p(\mu)}{\int p(n|\nu)p(\nu)d\nu} \quad (11.3)$$

where $p(\mu)$ is the prior probability density for the signal strength μ and $p(\mu|n)$ the corresponding posterior probability density.

Probabilities and probability densities, in Bayesian epistemology, represent our degrees of belief [133]. For example, if $P(H_0) = 0.95$ it means that we are very confident that the hypothesis H_0 is true. If our probability density has a sharp peak around $\mu = 1$, it means that we are confident that the true value of μ is very close to one.

This raises a very important question: how do we choose the prior probabilities? This is the so-called *problem of the priors* and is still heavily debated among Bayesian epistemologists and there is no consensus on how to solve this problem [133]. This does not mean that the Bayesian approach should not be used in physics (arguably it should be used more often than it is [134]) but it does mean that writing down a posterior probability as a summary of an experimental analysis would not be a good idea.

Instead of thinking about the (bayesian) probability that a certain hypothesis is true given the data, we will take a *frequentist* approach. It is another interpretation of the notion of probability. Instead of defining probabilities as degree of beliefs, frequentists interpret probabilities as a relative frequency of occurrence [135]. For example, the probability of obtaining ‘one’ when throwing a fair dice is $\frac{1}{6}$ because the ratio of the number of ‘one’ obtained over the total number of throws converges toward $\frac{1}{6}$ as the number of throws increases to infinity.

In a frequentist approach, it does not make sense to think about the probability that a certain hypothesis is true given some data. For a frequentist, a theory or a model is either true or it is not, there is no probability assigned to hypotheses. But what does make sense, either from a frequentist or a bayesian point of view, is to compute the probability of getting some data under a certain hypothesis. The upside is that we don’t have to worry about having the ‘correct’ priors for various models and/or signal strengths in order to perform such computation. The price to pay is that we do not estimate the probability of any model.

11.1.2 Profile likelihood

Let us come back to our example where we looked at the signal strength μ and the corresponding probability density function $p(n|\mu)$. For a fixed value of n , the *likelihood function* is defined by [136]:

$$L(\mu) = p(n|\mu) \tag{11.4}$$

In practice, we have to worry about many other parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$ that are called *nuisance parameters*, and correspond to uncertainties in the prediction. In contradistinction with nuisance parameters, μ is called the *parameter of interest* (for our particular example). We are also looking at more than one phase space region, and in each region, we have several bins. The data is therefore not just one number of events n , but $n_1, n_2, \dots, n_{n_{\text{bins}}}$. The likelihood function is then:

$$L(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{n_{\text{bins}}} p(n_i|\mu, \boldsymbol{\theta}) \tag{11.5}$$

The probability distribution to have n events in a particular bin, for a given signal strength μ , is given by the Poisson distribution [137, 138]:

$$p(n|\mu) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \tag{11.6}$$

Using the Poisson distribution, the likelihood function can be written as:

$$L(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{n_{\text{bins}}} \frac{[(\mu S(\boldsymbol{\theta}) + B(\boldsymbol{\theta}))^n]}{n!} e^{-[\mu S(\boldsymbol{\theta}) + B(\boldsymbol{\theta})]} \prod_{j=1}^m C(\theta_j) \tag{11.7}$$

where $C(\theta_j)$ constrains θ_j and is usually Gaussian [139].

Sometimes, *normalization factors* are also used, which are multiplicative factors scaling the background (or part of it) and are not associated with any systematic uncertainty. They also don't have any associated constraints $C(\theta_j)$.

Intuitively, we want to know what set of parameters makes our data the most likely. The idea is then to find which set of parameters $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ maximize the likelihood function, and this will give us our best estimate for the true value of these parameters. This approach is called the *maximum likelihood method*. For convenience, it is often useful to use the negative log-likelihood function

$-\ln(L)$, and we find the values of the parameters that minimize $-\ln(L)$ (and therefore maximize L) by solving the equations [140]:

$$-\frac{\partial \ln(L)}{\partial \mu} \Big|_{\mu=\hat{\mu}} = 0 \quad \text{and} \quad -\frac{\partial \ln(L)}{\partial \theta_j} \Big|_{\theta_j=\hat{\theta}_j} = 0 \quad \text{for } j = 0, 1, \dots, m \quad (11.8)$$

Maximum likelihood estimators have several favorable properties [141]:

- Consistency: they converge to the true value of the parameters.
- Small bias: the bias, defined as the difference between the expected value and the true value, is proportional to $1/N$, while the estimated uncertainty is proportional to $1/\sqrt{N}$.
- Efficiency: the variance of a parameter α is at most $\langle \alpha^2 \rangle - \langle \alpha \rangle^2$.
- Invariance: for any function of a parameter α , we have $f(\hat{\alpha}) = \widehat{f(\alpha)}$.

11.1.3 Profile likelihood ratio

The profile likelihood ratio is defined by:

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} \quad (11.9)$$

where $\hat{\theta}$ and $\hat{\mu}$ are obtained by minimizing $-\ln(L)$, i.e. they are the solutions to the equations 11.8. $\hat{\theta}(\mu)$ is obtained by minimizing $-\ln(L)$ for a fixed value of μ .

The Neyman-Person lemma [142] says that the likelihood ratio is the optimal discriminator between two competing hypotheses. For example, if the two hypotheses are the background-only hypothesis ($\mu = 0$) and the presence of signal ($\mu \geq 0$), we define the *test statistic* q_0 :

$$q_0 = \begin{cases} -2 \ln(\lambda(0)) & \text{if } \hat{\mu} \geq 0 \\ 0 & \text{if } \hat{\mu} < 0 \end{cases} \quad (11.10)$$

Note that we reject the possibility of $\mu < 0$ a priori. It can be shown that the test statistics $-2 \ln(\lambda)$ follow (asymptotically) a χ^2 distribution under the null hypothesis (in our case this corresponds to $\mu = 0$), this result is the so-called Wilks' theorem [143]. This means that we can obtain the probability density function for q_0 , that we will write $f(q_0|0)$. By integrating over this probability

density, we get the p -value:

$$p = \int_{q_{0,\text{obs}}}^{+\infty} f(q_0|0)dq_0 \quad (11.11)$$

which is the conditional probability, under the assumption that the null hypothesis is correct, of obtaining data at least as ‘extreme’ as the data collected. ‘Extreme’ means data with a larger q_0 than observed, i.e. data even less compatible with the null hypothesis than the data that has been collected; with the caveat that $\hat{\mu} < 0$ does not count against the null hypothesis.

It is customary in particle physics to compute the *significance* from the p -value. The significance is defined by:

$$Z = \Phi^{-1}(1 - p) \quad (11.12)$$

where $\Phi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$ is the cumulative distribution of the standard normal distribution. By convention, a significance of at least 3σ is considered to be evidence (for a process, a new particle, ...), and a significance of at least 5σ is considered to be discovery. 5σ corresponds to a p -value of 3×10^{-7} .

11.2 Neural network fit

11.2.1 Overview of the likelihood fit

The neural network output is used to perform a profile likelihood fit, using the TRExFitter framework [144]. TRExFitter is a very powerful tool for binned template profile likelihood fits. TRExFitter is based on several statistical tools like HistFactory [145], RooFit [146] and RooStats [147].

The parameter of interest in the fit is the signal strength of EWK $W^+W^- + \text{jets}$ production. The fit also provides the expected significance, with the null hypothesis corresponding to $\mu = 0$.

The signal events were ‘blinded’ to avoid being biased by the actual data when developing and tuning the analysis, with a blinding threshold set a 5%. As the low-NN region is dominated by background, it did not need to be blinded. The signal region was therefore split into two parts, one part with NN output < 0.6 and another one with NN output > 0.6 , for both the two-jets signal region and the three-jets signal region. NN output ≈ 0.6 is approximately the point where the signal over background ratio starts to grow larger than 5%. The NN output > 0.6 was kept completely blinded until the last moment.

To have the best estimation of the expected significance before unblinding, we perform a ‘mixed data and Monte Carlo’ fit [148] with TRExFitter. It is made by following three steps:

1. A background-only fit to the control region (and the signal region where NN output < 0.6) is performed first.
2. A ‘modified pseudo-data’ is constructed in the signal region with NN output > 0.6 using the fitted nuisance-parameter values from the first step and the Monte Carlo simulation prediction.
3. Finally, a fit is done using a mix of real data (CR and SR with NN output < 0.6) and modified pseudo-data (SR with NN output > 0.6).

For the likelihood fit with real data, no such splitting is necessary. Figures 11.1 and 11.2 show the neural network output distribution before and after the fit in the signal region, for two-jets and three-jets respectively. For the control region (where one of the leading jets is required to be a b-jet) there is no need to split between a two-jets and a three-jets region. The control region (before and after the fit) is shown in figure 11.3.

We can observe an overall good agreement between Monte Carlo predictions and data, both in the signal and control region, although with some small fluctuations in the top control region.

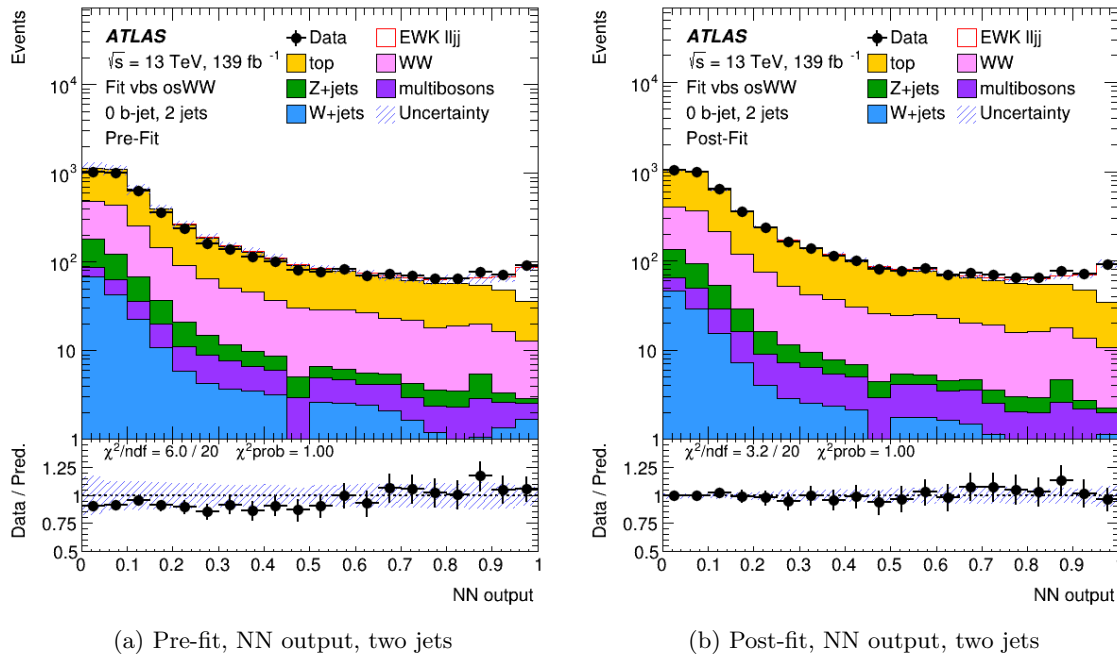


Figure 11.1: NN output in the signal region before (left) and after (right) the likelihood fit ($n_{\text{jets}} = 2$).

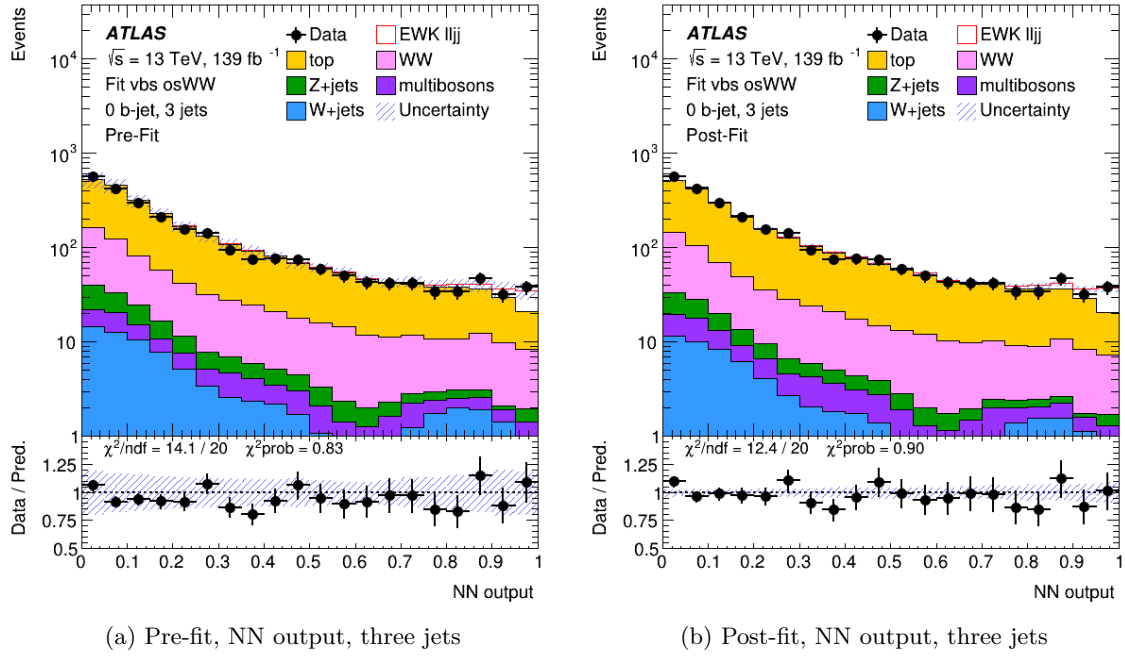


Figure 11.2: NN output in the signal region before (left) and after (right) the likelihood fit ($n_{\text{jets}} = 3$).

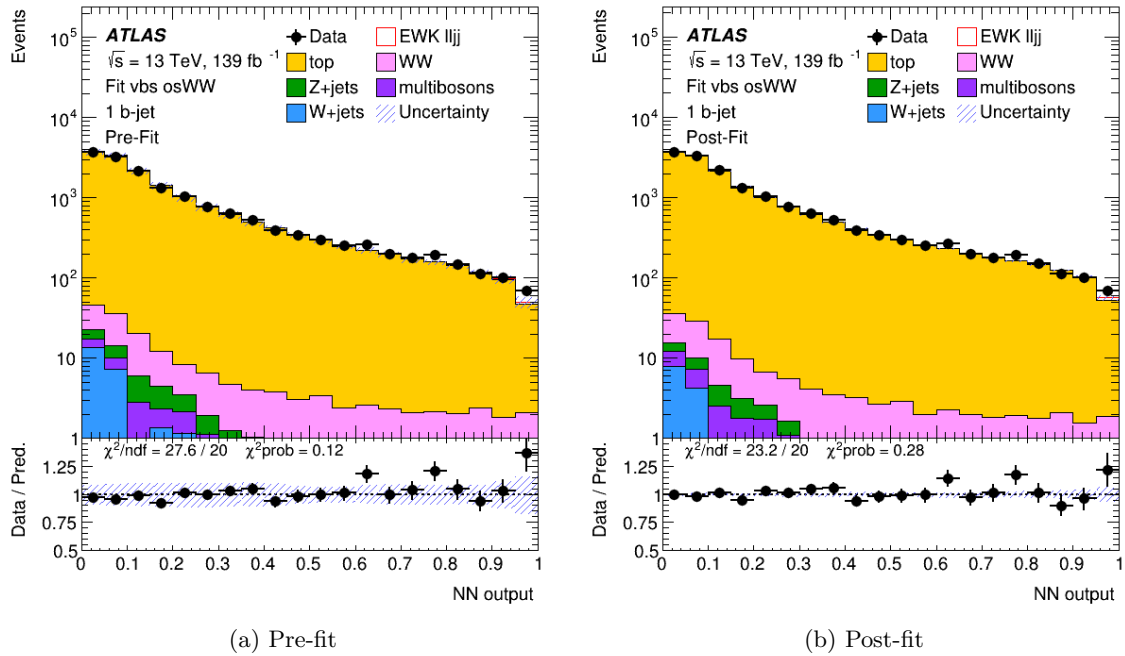


Figure 11.3: NN output in the control region before (left) and after (right) the likelihood fit.

Normalization factors are used for the top and QCD-WW backgrounds in the likelihood fit. The top normalization is well constrained by the fit, unlike the QCD-WW normalization as there is no dedicated control region for QCD-WW production. The overall impact on the total signal strength uncertainty is however limited as the QCD-WW background is only about 10-15% for large ($\gtrsim 0.7$) neural network output values.

Pruning is applied to avoid having too many nuisance parameters which would slow down the fit significantly. The pruning threshold has been set to 0.5%. A comparison was made with a test fit without pruning and the impact on the significance and total signal strength uncertainty was found to be negligible.

11.3 Smoothing of neural network distribution

A smoothing algorithm is applied to the shape of the neural network distribution for each systematic uncertainty (see figure 11.4 for two examples in the signal and control region). This is used in order to avoid ‘spiky’ distributions due to statistical fluctuations that can potentially lead to problems with the fit. The default TRExFitter smoothing algorithm (MaxVariation) is used: the distribution is rebinned until the relative Monte Carlo statistical uncertainty is below a defined tolerance and if the number of variations in the ratio is below a certain maximum value the tolerance is halved and re-binning is performed again. After all of this is done, the TH1::Smooth function is run, which is using the ‘353QH twice’ algorithm [149].

11.3.1 Fake lepton uncertainty

As explained in chapter 10, the V+jets Monte Carlo prediction is corrected using a transfer factor from a fake-enriched region. A conservative uncertainty of 100% ($^{+100\%}_{-99\%}$) is also used. With this ‘fake uncertainty’ added, the overall impact on the signal strength is about one percent. This shows that even a large uncertainty on V+jets leads to a small impact on the total uncertainty.

11.3.2 Nuisance parameters ranking

We can rank the nuisance parameters in term of how much each parameter impact the total uncertainty. Four fits are performed for each nuisance parameter with the parameter in question

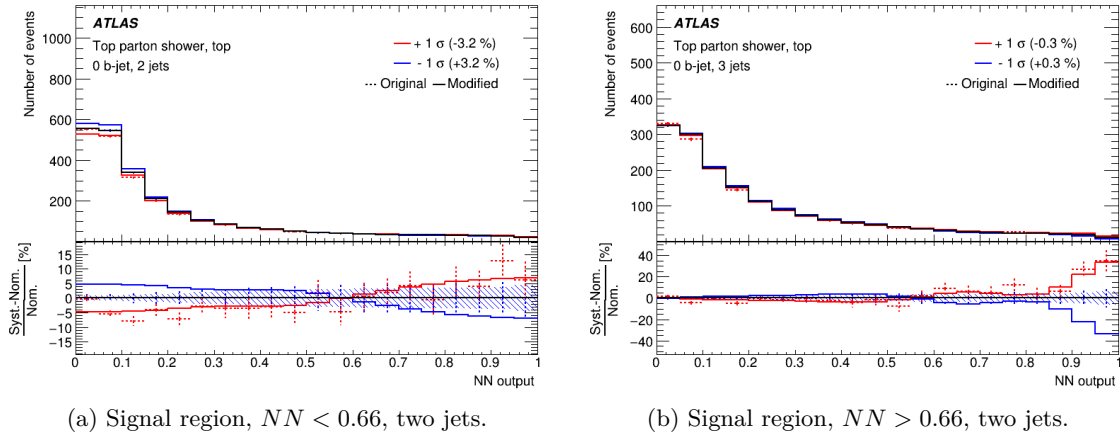


Figure 11.4: Smoothing of the systematics. The dashed (resp. solid) line shows the distribution before (resp. after) smoothing. These plots show the smoothing of the neural network output for the top parton shower uncertainty, in the signal region for two jets (left) and three jets (right)

fixed to the following four possible values:

- pre-fit value + pre-fit uncertainty
- pre-fit value - pre-fit uncertainty
- post-fit value + post-fit uncertainty
- post-fit value - post-fit uncertainty

Figure 11.6 shows the nuisance parameters ranking for different categories. The impact (for each nuisance parameter) is the difference between the signal strength computed from these fits and the one obtained from the nominal fit.

Figure 11.5 shows the correlation coefficients between the nuisance parameters (including also the signal strength).

Table 11.1 shows an overview of the uncertainties for different categories. For the systematic uncertainties, the fit is done again with the nuisance parameters in a certain group held constant at their best fit values. This fit is associated with a new uncertainty on the signal strength $\Delta\mu'$ (the uncertainty for the nominal fit being $\Delta\mu$) and the impact of this group of uncertainties is defined as:

$$\sqrt{(\Delta\mu)^2 - (\Delta\mu')^2} \quad (11.13)$$

The Monte Carlo statistical uncertainty (γ in figure 11.6) corresponds to the systematic uncertainty in the prediction due to the statistics of the Monte Carlo sample. On the other hand, what is

ATLAS Internal

$\mu(EWK WW_{jj})$	100.0	-10.5	-5.0	4.7	-5.9	-8.6	16.0	6.6	-24.6	-3.6	-19.6	-30.0	-26.5
WWXsec	-10.5	100.0	-5.3	-74.9	10.5	25.3	-34.1	28.9	-35.8	7.8	-24.4	5.5	-30.9
Single Top Interference	-5.0	-5.3	100.0	8.5	3.9	-1.1	-5.1	-3.4	1.2	28.6	-15.8	-4.4	19.4
Fake uncertainty	4.7	-74.9	8.5	100.0	13.0	-22.9	23.2	-24.7	25.9	-7.6	5.0	-3.9	11.6
Final state radiation	-5.9	10.5	3.9	13.0	100.0	-0.3	2.9	-3.5	10.2	-23.4	-20.5	-3.3	-25.5
T_JER_DataVsMC_MC16_smear	-8.6	25.3	-1.1	-22.9	-0.3	100.0	0.2	2.0	-9.7	7.4	1.3	-0.2	-3.7
JET_JER_EffectiveNP_1_smear	16.0	-34.1	-5.1	23.2	2.9	0.2	100.0	-10.6	-1.0	0.8	-15.4	-5.3	-9.8
JET_Pileup_RhoTopology	6.6	28.9	-3.4	-24.7	-3.5	2.0	-10.6	100.0	-5.1	27.1	8.7	-0.5	20.9
Top parton shower	-24.6	-35.8	1.2	25.9	10.2	-9.7	-1.0	-5.1	100.0	-44.8	43.2	-12.6	62.9
QCD scale and ISR	-3.6	7.8	28.6	-7.6	-23.4	7.4	0.8	27.1	-44.8	100.0	-19.2	-2.5	27.8
Top matrix element	-19.6	-24.4	-15.8	5.0	-20.5	1.3	-15.4	8.7	43.2	-19.2	100.0	-10.5	54.4
Signal parton shower	-30.0	5.5	-4.4	-3.9	-3.3	-0.2	-5.3	-0.5	-12.6	-2.5	-10.5	100.0	-14.0
topXsec	-26.5	-30.9	19.4	11.6	-25.5	-3.7	-9.8	20.9	62.9	27.8	54.4	-14.0	100.0
	$\mu(EWK WW_{jj})$	WWXsec	Single Top Interference	Fake uncertainty	Final state radiation	T_JER_DataVsMC_MC16_smear	JET_JER_EffectiveNP_1_smear	JET_Pileup_RhoTopology	Top parton shower	QCD scale and ISR	Top matrix element	Signal parton shower	topXsec

Figure 11.5: Nuisance parameters correlation matrix.

called ‘statistical uncertainty’ in table 11.1 is simply the expected statistical uncertainty in data. It is estimated by performing the fit while ignoring all the other sources of uncertainties (setting `StatOnlyFit = True` and `FixNPforStatOnly = True` in `TRExFitter`).

Systematic	$\frac{\sqrt{(\Delta\mu)^2 - (\Delta\mu')^2}}{\mu}$
Monte Carlo statistical uncertainty	7.70%
Top theoretical uncertainties	6.06%
Signal theoretical uncertainties	5.85%
Jet experimental uncertainties	4.87%
Luminosity	1.71%
Fake uncertainty	0.56%
B-tagging	0.39%
Lepton experimental uncertainties	0.13%
Others	0.25%
Statistical uncertainty	12.66%
Top normalization uncertainty	6.06%
QCD-WW normalization uncertainty	1.32%
Total uncertainty	18.58%

Table 11.1: Systematic uncertainties and the corresponding post-fit uncertainty on the signal strength. The statistical and total uncertainties are also shown.

The signal strength estimation obtained from the fit is shown in figure 11.7. Its observed value is:

$$\mu = 1.20 \pm 0.22 \tag{11.14}$$

The fitted top and QCD WW background normalization are also shown in figure 11.7. The observed (expected) significance given by `TRExFitter` is 7.06σ (6.21σ).

A plot of the difference between the negative log-likelihood $-\log(L)$ evaluated at some value μ with the best-fitted value $\hat{\mu}$ is shown in figure 11.8. The best-fit value $\hat{\mu}$ is at the minimum of the curve (as expected).

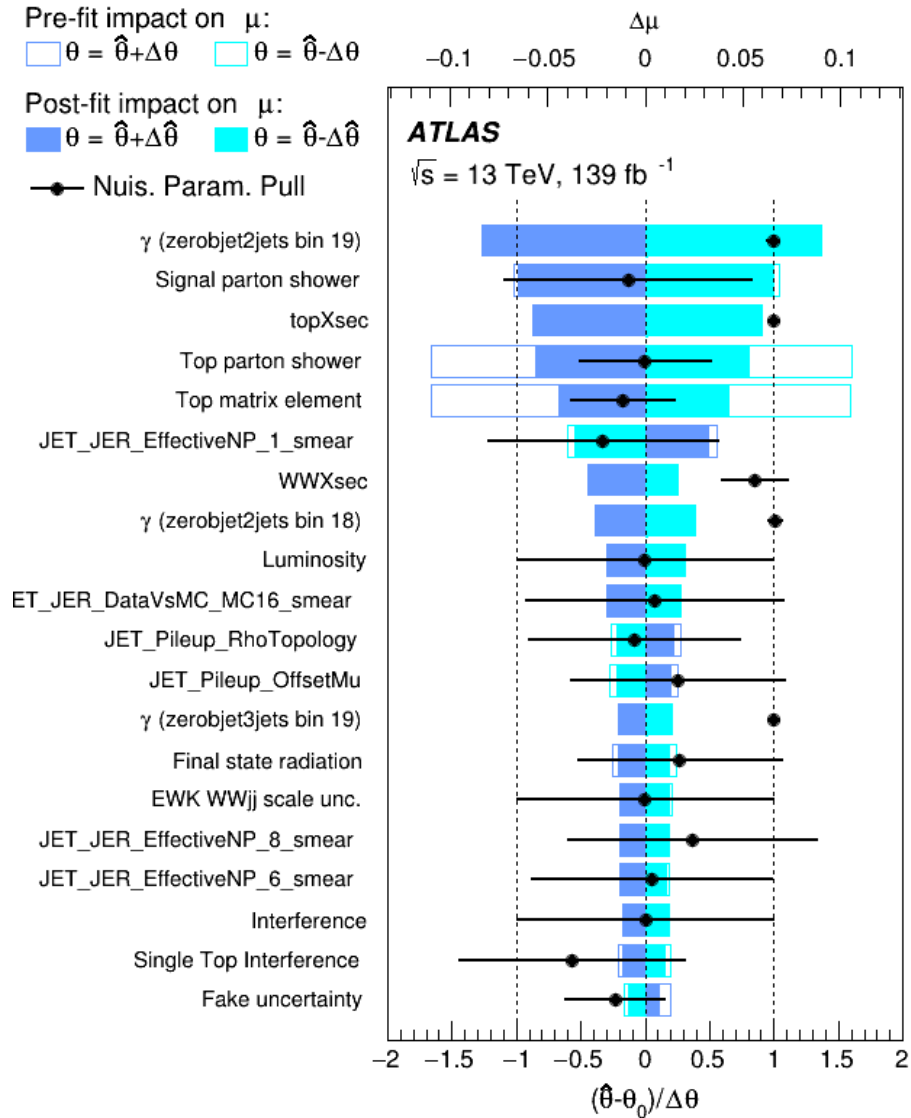


Figure 11.6: Nuisance parameter ranked by impact in order of post-fit impact. γ corresponds to the Monte Carlo statistical uncertainty in a particular bin. The upper x-axis corresponds to the uncertainty of the signal strength $\Delta\mu$, the pre-fit uncertainties are shown with the bars filled with white and the post-fit ones by fully colored bars. The lower y-axis shows the nuisance parameter pull $(\hat{\theta} - \theta_0) / \Delta\theta$ where $\hat{\theta}$ is the best-fit value of the nuisance parameter θ , θ_0 is the nominal (pre-fit) value of θ and $\Delta\theta$ is the uncertainty. The pull for the MC statistical uncertainties or any other scale factor is expected to be around one because it is just rescaling the MC prediction.

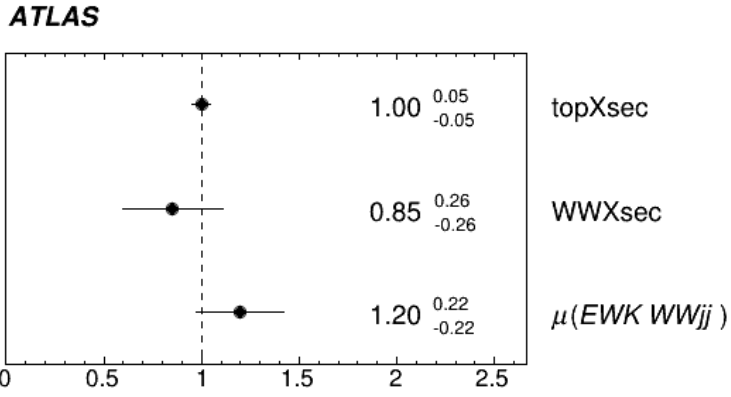


Figure 11.7: Signal strength estimation from the likelihood fit (bottom) as well as top and WW-QCD normalization.

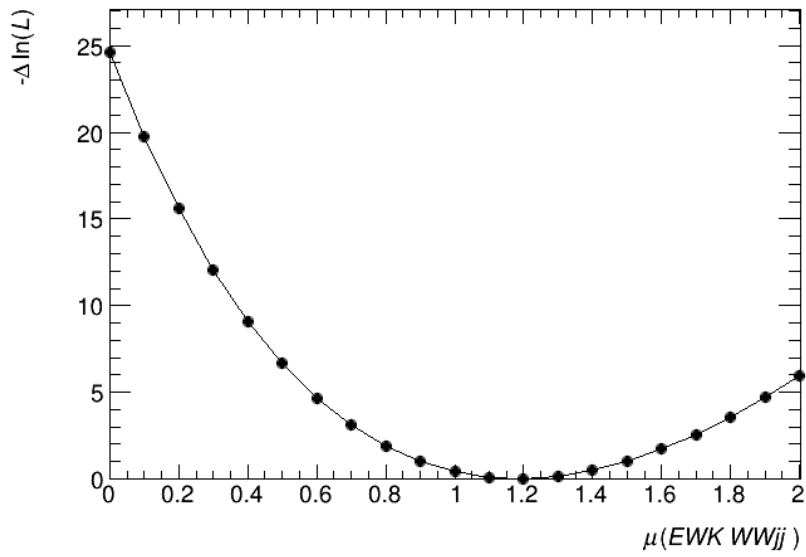


Figure 11.8: Likelihood scan for the signal strength μ .

11.4 Cross-section

The cross-section is measured in a fiducial region that is designed to be similar to the part of the signal region which is most sensitive to the signal cross-section, instead of the whole signal region. A cut on m_{jj} , not present in the reconstructed signal region, is added in the fiducial region definition. This is due to two reasons:

1. It is ambiguous whether tribosons processes corresponding to Feynman diagrams with electroweak order equal to six ought to be defined as part of our signal or not. We are using POWHEGBOX and this tribosons contribution is not included, but the same process simulated using MADGRAPH5_aMC@NLO would include it by default. Adding a cut on m_{jj} strongly suppresses this contribution and makes comparison with other generators easier.
2. Our signal region being fairly inclusive, the signal strength obtained from the fit is not very sensitive to signal events with low neural network output values. So, the point of cutting on m_{jj} in the fiducial region definition is to make sure that we are avoiding a potential over-extrapolation. Without this cut, we would be applying the same signal strength to the low m_{jj} region of the phase space even though low m_{jj} signal events make a negligible impact on the likelihood fit result.

In order to determine the most appropriate value for fiducial region cut on m_{jj} , we look at three criteria:

- The generator-level cut on m_{jj} must reduce the number of reconstructed events in low NN bins, i.e. where signal/background is very small ($< 5\%$ which corresponds to $NN < 0.6$) but not significantly reduce the number of events in high NN bins ($NN > 0.6$).
- The m_{jj} generator-level cut efficiencies must be similar to the $NN > 0.6$ reconstruction cut efficiencies because we want the fiducial region to be similar to the $NN > 0.6$ reconstructed region.
- The m_{jj} cut must suppress tribosons contributions.

Choosing the m_{jj} cut at 500 GeV satisfy all three criteria. Figure 11.9 shows a comparison of the neural network distribution in the signal region with and without this cut. Figures 11.10 and 11.11 show the cut efficiencies for several variables, comparing the m_{jj} cut and the $NN > 0.6$ cut.

The efficiencies are defined as the ratio:

$$\frac{\text{Number of events after cut}}{\text{Number of events in the signal region}} \quad (11.15)$$

for each bin. As can be seen in 11.10 and 11.11, the efficiencies are roughly similar for cuts, which provide evidence that the fiducial region with $m_{jj} > 500$ GeV is similar to the reconstructed signal region with $NN > 0.6$.

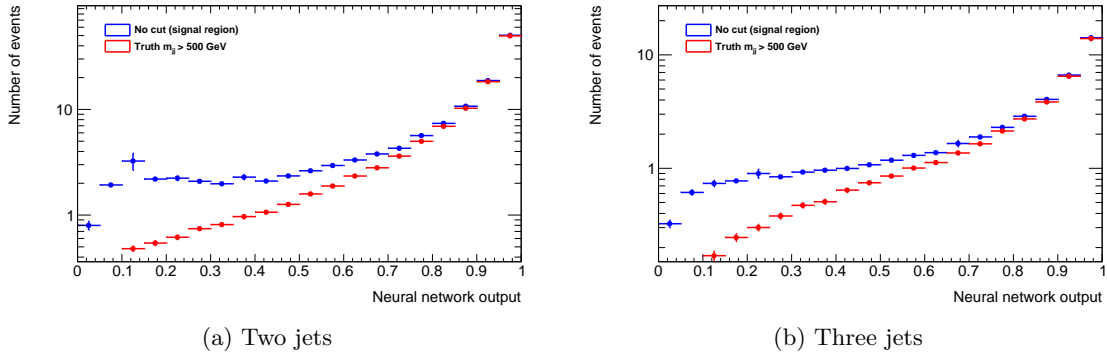


Figure 11.9: Neural network distribution for two jets (left) and three jets (right). The distribution without an additional generator-level m_{jj} cut is shown in blue and the distribution with an additional $m_{jj} > 500$ GeV generator-level cut in red. We can observe that elements with $NN < 0.6$ are suppressed.

Finally, from the tribosons m_{jj} distribution, shown in figure 11.12, it can be inferred that cutting m_{jj} at 500 GeV strongly suppresses tribosons.

The fiducial region is then defined with a selection that is similar to the reconstructed signal region, with an additional cut on m_{jj} . The fiducial region definition is shown in table 11.2.

From the fiducial region, the theoretical cross-section, as given by POWHEGBOX can be computed. It is given by:

$$\sigma_{\text{POWHEG}} = 2.20 \pm 0.14 \text{ fb} \quad (11.16)$$

where the uncertainty includes all the theoretical uncertainties. A blinded likelihood fit is performed where the signal theoretical uncertainties are split into a normalization uncertainty and a shape uncertainty. The theoretical ‘normalization uncertainties’ are then ignored, to avoid double-counting, as they are already included in equation 11.16. The cross-section given by equation 11.16 is then multiplied by this signal strength and the uncertainties are combined. The result is the expected

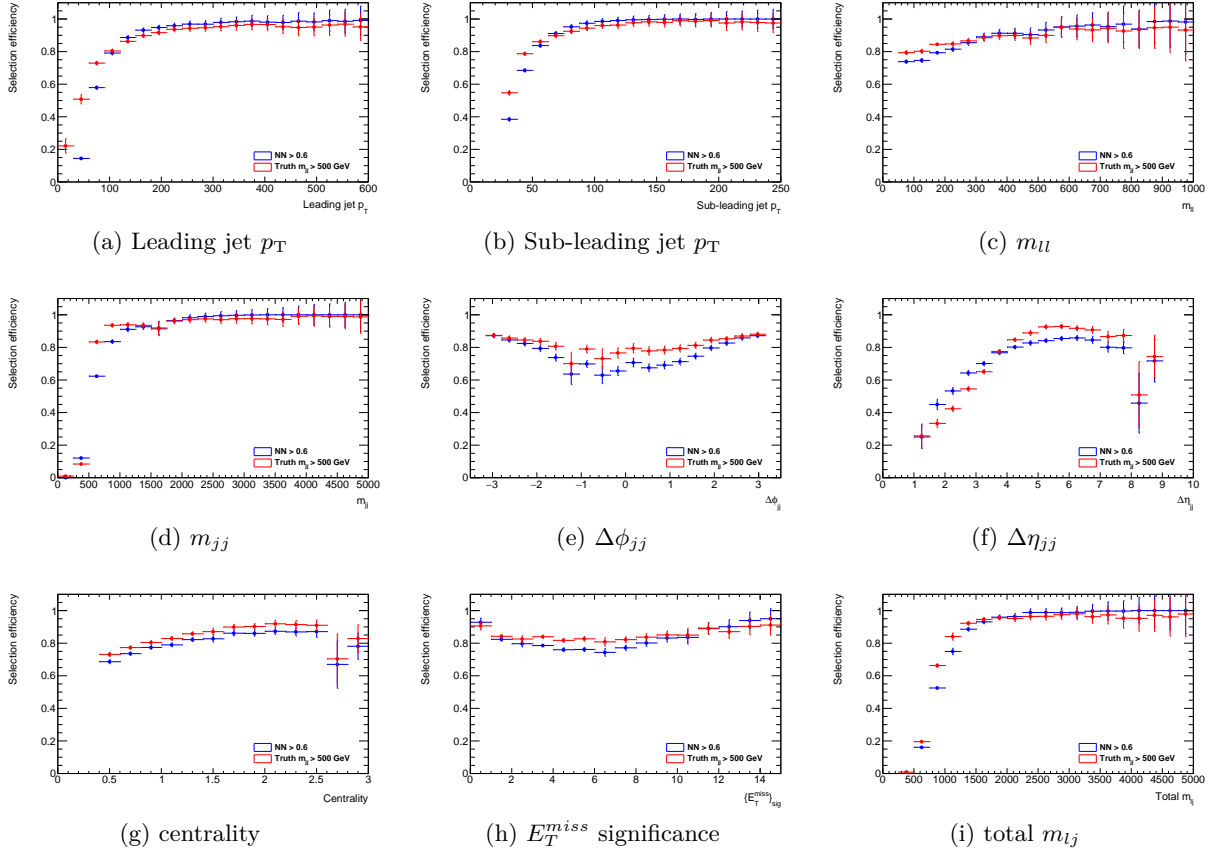


Figure 11.10: Cut efficiencies for $NN > 0.6$ and generator-level m_{jj} cut for the two jets signal region.

Requirements

One electron and one muon with opposite charges

No additional lepton

$$p_T^{\text{dressed } \ell} = p_T^\ell + \sum_i p_T^{\gamma_i} \text{ if } \Delta R(\ell, \gamma_i) < 0.1$$

Leptons p_T : $p_T > 27 \text{ GeV}$

Lepton η : $|\eta| < 2.5$

Two or three jets with $p_T > 25 \text{ GeV}$ and $|\eta| < 4.5$

$$p_T^{\text{miss}} > 15 \text{ GeV}$$

$$m_{jj} > 500 \text{ GeV}$$

centrality > 0.5

$$m_{e\mu} > 80 \text{ GeV}$$

no b-jet veto with $\eta > 2.5$

Table 11.2: Fiducial region definition.

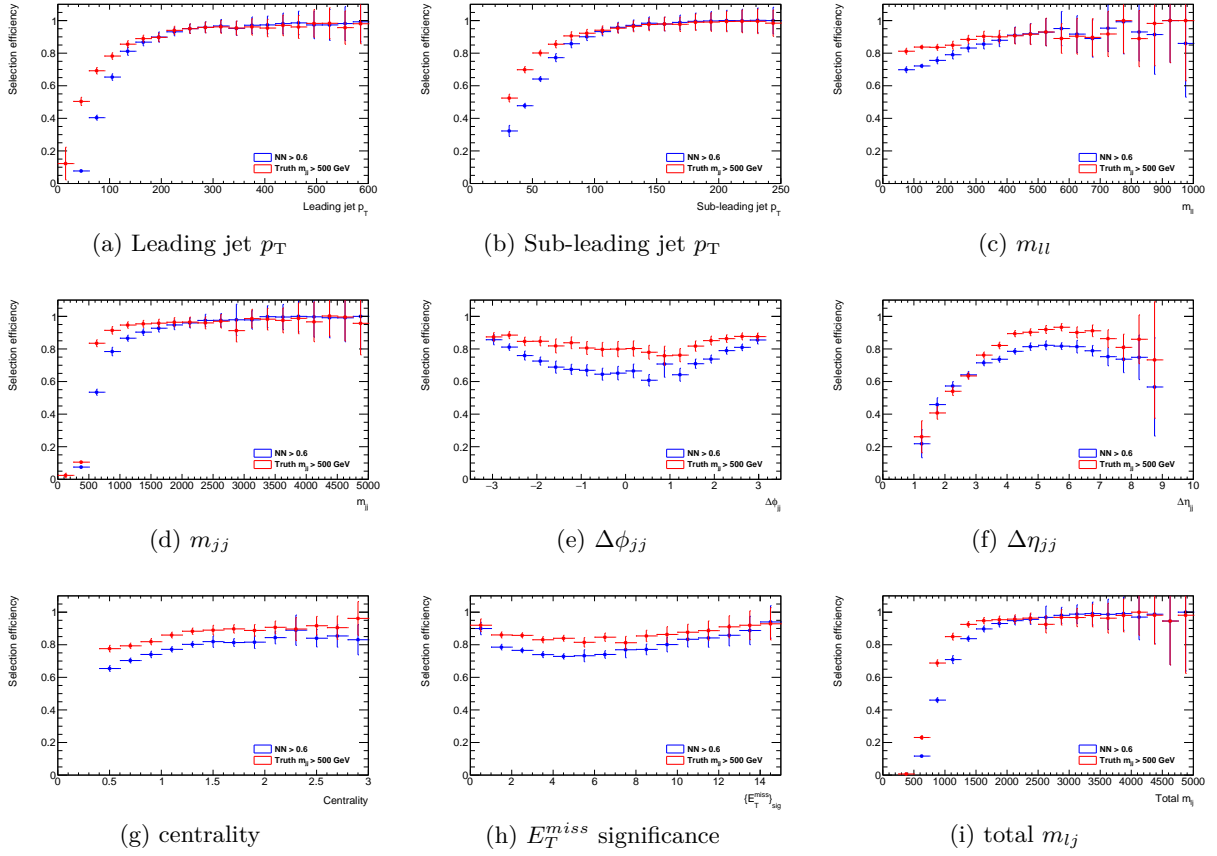


Figure 11.11: Cut efficiencies for $NN > 0.6$ and generator-level m_{jj} cut for the three jets signal region.

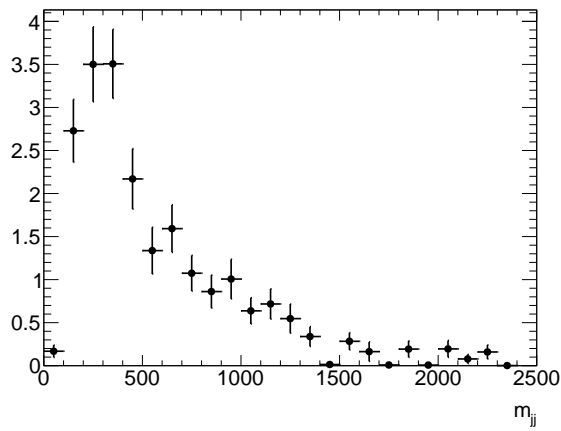


Figure 11.12: m_{jj} tribosons distribution in the signal region.

cross-section:

$$\sigma_{\text{expected}} = 2.20 \pm 0.47 \text{ fb} \quad (11.17)$$

The same procedure is then followed to obtain the observed cross-section. The result is:

$$\sigma_{\text{observed}} = 2.62 \pm 0.50 \text{ fb} \quad (11.18)$$

The measured cross-section is in agreement with the Standard Model prediction, given by the NLO QCD simulation with POWHEG.

Chapter 12

Conclusion

The Standard Model represents one of the greatest accomplishments in the history of physics. It has been tested for many decades using ever more powerful particle colliders. Despite an impressive amount of experimental and theoretical effort, we still do not know what lies beyond the Standard Model. This is why it is important to test the Standard Model as precisely and thoroughly as possible and the ATLAS experiment at the LHC is a powerful tool to accomplish this goal.

Vector boson scattering processes are particularly interesting and useful to study at the LHC. According to the Standard Model, perturbative unitarity for WW scattering is a consequence of the Higgs mechanism. However, several BSM models, such as the 2HDM model, predict that the Higgs (i.e. the lightest one for models with several Higgs) only partially unitarizes this process. If such a model is correct, the cross-section of WW vector boson scattering would be larger than predicted. Measurement of vector boson scattering cross-section is a potential path to the discovery of new physics.

This thesis presented a measurement of the electroweak production of two opposite-sign W bosons with two or three jets, using data collected with the ATLAS detector, during the LHC run-II corresponding to a total integrated luminosity of 139 fb^{-1} . This is the first measurement and first observation of this process in ATLAS.

The main difficulty with this analysis is the existence of two large backgrounds that are challenging to reduce: the top and the QCD W^+W^- backgrounds. Machine learning techniques were therefore used to optimize signal/background separation. Two neural networks (corresponding to two and three jets) were trained and used to separate the signal from the aforementioned backgrounds. Validation studies were performed, checking the modeling of the neural network input variables in a background-enriched region defined by cutting on the neural network output ($\text{NN} > 0.6$). A top control region, where one of the two leading jets is required to be a b-jet, was defined to better constrain the top background.

A likelihood fit is then performed to extract the signal strength, using the maximum likelihood estimation method, taking into account all the experimental and theoretical uncertainties affecting

the result. The resulting signal was found to be:

$$\mu = 1.21 \pm 0.22 \tag{12.1}$$

The observed (expected) significance was found to be 7.06σ (6.21σ). A fiducial region we also defined, and was designed to be similar to the most sensitive $NN > 0.6$ part of the signal region.

The observed (expected) cross-section is:

$$\sigma = 2.62 \pm 0.51 \text{ fb} \text{ (} 2.20 \pm 0.47 \text{ fb)} \tag{12.2}$$

The observed signal strength and cross-section are in agreement with Standard Model predictions.

Bibliography

- [1] M. Gonzalez-Garcia and M. Maltoni, “Phenomenology with massive neutrinos”, en, *Physics Reports* **460**, 1–129 (2008).
- [2] Wikipedia contributors MissMJ and Cush, *Standard model of elementary particles*, (2020) https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [3] M. D. Schwartz, *Quantum field theory and the standard model* (Cambridge University Press, New York, 2014).
- [4] M. Talagrand, *What Is a Quantum Field Theory?* (Cambridge University Press, Cambridge, 2022).
- [5] I. Brivio and M. Trott, “The standard model as an effective field theory”, en, *Physics Reports* **793**, 1–98 (2019).
- [6] P. A. Zyla et al., “Review of Particle Physics”, en, *Progress of Theoretical and Experimental Physics* **2020**, 083C01 (2020).
- [7] T. Maudlin, “On the Unification of Physics:” de, *Journal of Philosophy* **93**, edited by J. Smylie, 129–144 (1996).
- [8] S. Weinberg, “V-A was the key”, *J. Phys.: Conf. Ser.* **196**, 012002 (2009).
- [9] J. Goldstone, “Field theories with « Superconductor » solutions”, en, *Nuovo Cim* **19**, 154–164 (1961).
- [10] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, “Experimental Test of Parity Conservation in Beta Decay”, en, *Phys. Rev.* **105**, 1413–1415 (1957).
- [11] D. D. Tong, “Lectures on quantum field theory”, en, Part III Cambridge University Mathematics Tripos, Michaelmas, 155 (2006).
- [12] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, en, *Phys. Rev. Lett.* **10**, 531–533 (1963).
- [13] M. Kobayashi and T. Maskawa, “ $C P$ -Violation in the Renormalizable Theory of Weak Interaction”, en, *Prog. Theor. Phys.* **49**, 652–657 (1973).

- [14] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, “Evidence for the 2π Decay of the K_2^0 Meson”, en, *Phys. Rev. Lett.* **13**, 138–140 (1964).
- [15] S. D. Rindani, “Strong Gauge Boson Scattering at the LHC”, en, in *Physics at the Large Hadron Collider*, edited by A. Datta, B. Mukhopadhyaya, A. Raychaudhuri, A. K. Gupta, C. L. Khetrapal, T. Padmanabhan, A. Raychaudhuri, and M. Vijayan (Springer India, New Delhi, 2009), pp. 145–155.
- [16] B. W. Lee, C. Quigg, and H. B. Thacker, “Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass”, en, *Phys. Rev. Lett.* **38**, 883–885 (1977).
- [17] C. T. Hill and E. H. Simmons, “Strong dynamics and electroweak symmetry breaking”, en, *Physics Reports* **381**, 235–402 (2003).
- [18] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* **716**, 1 (2012).
- [19] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716**, 30 (2012).
- [20] D. Espriu and B. Yenko, “Longitudinal W W scattering in light of the “Higgs boson” discovery”, en, *Phys. Rev. D* **87**, 055017 (2013).
- [21] J. Chang, K. Cheung, C.-T. Lu, and T.-C. Yuan, “W W scattering in the era of post-Higgs-boson discovery”, en, *Phys. Rev. D* **87**, 093005 (2013).
- [22] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva, “Theory and phenomenology of two-Higgs-doublet models”, en, *Physics Reports* **516**, 1–102 (2012).
- [23] D. Buarque Franzosi et al., “Vector boson scattering processes: Status and prospects”, en, *Reviews in Physics* **8**, 100071 (2022).
- [24] ATLAS Collaboration, “Standard Model Summary Plots February 2022”, (2022).
- [25] CMS Collaboration, *Summaries of CMS cross section measurements*, (2022) <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined>.
- [26] R. Covarelli, M. Pellen, and M. Zaro, “Vector-Boson scattering at the LHC: Unraveling the electroweak sector”, en, *Int. J. Mod. Phys. A* **36**, 2130009 (2021).

- [27] J. M. Campbell, J. W. Huston, and W. J. Stirling, “Hard interactions of quarks and gluons: a primer for LHC physics”, en, *Rep. Prog. Phys.* **70**, 89–193 (2007).
- [28] A. Buckley et al., “General-purpose event generators for LHC physics”, en, *Physics Reports* **504**, 145–233 (2011).
- [29] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, “Parton distributions in the LHC era: MMHT 2014 PDFs”, en, *Eur. Phys. J. C* **75**, 204 (2015).
- [30] B. Webber, “Parton shower Monte Carlo event generators”, en, *Scholarpedia* **6**, 10662 (2011).
- [31] ATLAS Collaboration, “Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J. C* **73**, 2304 (2013).
- [32] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton fragmentation and string dynamics”, en, *Physics Reports* **97**, 31–145 (1983).
- [33] B. R. Webber, “A QCD model for jet fragmentation including soft gluon interference”, en, *Nuclear Physics B* **238**, 492–528 (1984).
- [34] D. Amati and G. Veneziano, “Preconfinement as a property of perturbative QCD”, en, *Physics Letters B* **83**, 87–92 (1979).
- [35] M. Cardoso, N. Cardoso, and P. Bicudo, “Lattice QCD computation of the colour fields for the static hybrid quark-gluon-antiquark system, and microscopic study of the Casimir scaling”, en, *Phys. Rev. D* **81**, arXiv:0912.3181 [hep-lat], 034504 (2010).
- [36] N. Metropolis and S. Ulam, “The monte carlo method”, *Journal of the American statistical association* **44**, 335–341 (1949).
- [37] S. Weinzierl, *Introduction to Monte Carlo methods*, en, arXiv:hep-ph/0006269, June 2000.
- [38] T. Kloek and H. K. van Dijk, “Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo”, *Econometrica* **46**, 1 (1978).
- [39] G. Peter Lepage, “A new algorithm for adaptive multidimensional integration”, en, *Journal of Computational Physics* **27**, 192–203 (1978).
- [40] F. Siegert, “Monte-Carlo event generation for the LHC”, en, PhD thesis (Durham U., 2010).

- [41] 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”, *J. High Energy Phys.* **2014**, arXiv:1405.0301 [hep-ph], 79 (2014).
- [42] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, “Event generation with SHERPA 1.1”, *J. High Energy Phys.* **2009**, 007–007 (2009).
- [43] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *J. High Energy Phys.* **2010**, 43 (2010).
- [44] P. Nason, “A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms”, *J. High Energy Phys.* **2004**, 040–040 (2004).
- [45] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *J. High Energy Phys.* **2007**, 070–070 (2007).
- [46] P. F. Monni, P. Nason, E. Re, M. Wiesemann, and G. Zanderighi, “Erratum to: MiNNLOPS: a new method to match NNLO QCD to parton showers”, *J. High Energy Phys.* **2022**, 31 (2022).
- [47] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, en, arXiv:2203.11601 [hep-ex, physics:hep-ph], Mar. 2022.
- [48] M. Bähr et al., “Herwig++ physics and manual”, en, *Eur. Phys. J. C* **58**, 639–707 (2008).
- [49] J. Bellm et al., “Herwig 7.0/Herwig++ 3.0 release note”, en, *Eur. Phys. J. C* **76**, 196 (2016).
- [50] L. Evans and P. Bryant, “LHC Machine”, *J. Inst.* **3**, S08001–S08001 (2008).
- [51] *Where did it all begin?*, <https://home.web.cern.ch/about/who-we-are/our-history>.
- [52] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*, CERN Yellow Reports: Monographs (CERN, Geneva, 2004).
- [53] P. Mouche, “Overall view of the LHC. Vue d’ensemble du LHC”, General Photo (2014).
- [54] G. Aad et al. (ATLAS), “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3**, S08003 (2008).

- [55] S. Chatrchyan et al. (CMS), “The CMS Experiment at the CERN LHC”, JINST **3**, S08004 (2008).
- [56] A. A. Alves Jr. et al. (LHCb), “The LHCb Detector at the LHC”, JINST **3**, S08005 (2008).
- [57] K. Aamodt et al. (ALICE), “The ALICE experiment at the CERN LHC”, JINST **3**, S08002 (2008).
- [58] A. Tricomi et al., “The LHCf experiment at the LHC: Physics goals and status”, Nucl. Phys. B Proc. Suppl. **196**, edited by J.-N. Capdevielle, R. Engel, and B. Pattison, 30–35 (2009).
- [59] G. Anelli et al. (TOTEM), “The TOTEM experiment at the CERN Large Hadron Collider”, JINST **3**, S08007 (2008).
- [60] J. Pinfold, “The MoEDAL experiment at the LHC”, EPJ Web Conf. **145**, edited by B. Pattison, 12002 (2017).
- [61] H. Abreu et al. (FASER), “The FASER Detector”, (2022).
- [62] R. Bailey and P. Collier, *Standard Filling Schemes for Various LHC Operation Modes*, tech. rep. (CERN, Geneva, 2003).
- [63] E. Mobs, “The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019”, General Photo (2019).
- [64] *Public atlas luminosity results for run-2 of the lhc*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [65] R. Linck, “Electroweak Production of Two Opposite-Sign W Bosons using the ATLAS Detector”, Presented 11 Aug 2021 (Oct. 2021).
- [66] H. ten Kate, “The ATLAS superconducting magnet system at the Large Hadron Collider”, en, Physica C: Superconductivity **468**, 2137–2142 (2008).
- [67] J. Pequeno, “Computer generated image of the ATLAS inner detector”, 2008.
- [68] A. Abdesselam et al., “The barrel modules of the ATLAS semiconductor tracker”, en, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **568**, 642–671 (2006).

- [69] B. Mandelli, “The Pixel Detector of the ATLAS Experiment for the Run 2 at the Large Hadron Collider”, en, Nuclear and Particle Physics Proceedings **273-275**, 1166–1172 (2016).
- [70] F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*, Vol. 275, Springer Tracts in Modern Physics (Springer International Publishing, Cham, 2017).
- [71] Y. Takubo (ATLAS), *ATLAS IBL operational experience*, tech. rep. (CERN, Geneva, 2017).
- [72] G. Aad et al. (ATLAS), “Operation and performance of the ATLAS semiconductor tracker in LHC Run 2”, JINST **17**, P01013 (2022).
- [73] V. A. Mitsou and On behalf of the ATLAS TRT Collaboration, “THE ATLAS TRANSITION RADIATION TRACKER”, en, in Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications (July 2004), pp. 497–501.
- [74] S. Morgenstern (ATLAS), *ATLAS LAr Calorimeter Performance in LHC Run-2*, tech. rep. (CERN, Geneva, 2019).
- [75] P. Francavilla and On behalf of the ATLAS Collaboration), “The ATLAS Tile Hadronic Calorimeter performance at the LHC”, J. Phys.: Conf. Ser. **404**, 012007 (2012).
- [76] B. Aubert et al., “Performance of a liquid argon electromagnetic calorimeter with an “accordion” geometry”, en, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **309**, 438–449 (1991).
- [77] J. Pequeno, “Computer generated image of the ATLAS Muons subsystem”, 2008.
- [78] G. Aad et al. (ATLAS), “Operation of the ATLAS trigger system in Run 2”, JINST **15**, P10004 (2020).
- [79] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse, and A. Salzburger, “The new ATLAS track reconstruction (NEWT)”, J. Phys.: Conf. Ser. **119**, 032014 (2008).
- [80] M. Aaboud et al. (ATLAS), “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2”, Eur. Phys. J. C **77**, 673 (2017).
- [81] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting”, en, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262**, 444–450 (1987).

- [82] T. G. Cornelissen et al., “Updates of the ATLAS Tracking Event Data Model (Release 13)”, (2007).
- [83] F. Meloni (fortheATLAS), *Primary vertex reconstruction with the ATLAS detector*, tech. rep. 12 (CERN, Geneva, 2016).
- [84] G. Aad et al. (ATLAS), “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”, *Eur. Phys. J. C* **77**, 490 (2017).
- [85] G. Aad et al. (ATLAS), “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data”, *JINST* **14**, P12006 (2019).
- [86] G. Aad et al. (ATLAS), “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **76**, 292 (2016).
- [87] G. Aad et al. (ATLAS), “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **81**, 578 (2021).
- [88] M. Aaboud et al. (ATLAS), “Jet reconstruction and performance using particle flow with the ATLAS Detector”, *Eur. Phys. J. C* **77**, 466 (2017).
- [89] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *J. High Energy Phys.* **2008**, 063–063 (2008).
- [90] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *en*, *Phys. Rev. D* **48**, 3160–3166 (1993).
- [91] Y. Dokshitzer, G. Leder, S. Moretti, and B. Webber, “Better jet clustering algorithms”, *J. High Energy Phys.* **1997**, 001–001 (1997).
- [92] G. P. Salam, “Towards jetography”, *en*, *Eur. Phys. J. C* **67**, 637–686 (2010).
- [93] *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, tech. rep., All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-013> (CERN, Geneva, 2017).
- [94] M. Aaboud et al. (ATLAS), “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **78**, 903 (2018).

- [95] G. Aad et al. (ATLAS), “Performance of electron and photon triggers in ATLAS during LHC Run 2”, *Eur. Phys. J. C* **80**, 47 (2020).
- [96] G. Aad et al. (ATLAS), “Performance of the ATLAS muon triggers in Run 2”, *JINST* **15**, P09015 (2020).
- [97] R. L. Hayes (ATLAS), *ATLAS Muon Trigger Performance*, tech. rep. (CERN, Geneva, 2019).
- [98] R. D. Ball et al. (NNPDF), “Parton distributions for the LHC run II”, *JHEP* **04**, 040 (2015).
- [99] J. Pumplin et al., “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis”, *JHEP* **07**, 012 (2002).
- [100] ATLAS Collaboration, “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *JHEP* **09**, 145 (2014).
- [101] B. Jager and G. Zanderighi, “Electroweak $W+W$ -jj production at NLO in QCD matched with parton shower in the POWHEG-BOX”, *JHEP* **04**, 024 (2013).
- [102] A. Denner, R. Franken, T. Schmidt, and C. Schwan, “NLO QCD and EW corrections to vector-boson scattering into W^+W^- at the LHC”, (2022).
- [103] ATLAS Collaboration (ATLAS Collaboration), *ATLAS Pythia 8 tunes to 7 TeV data*, tech. rep. ATL-PHYS-PUB-2014-021 (CERN, Geneva, Nov. 2017).
- [104] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC”, *Eur. Phys. J. C* **63**, 189 (2009).
- [105] D. J. Lange, “The EvtGen particle decay simulation package”, *Nucl. Instrum. Meth. A* **462**, 152 (2001).
- [106] T. Gleisberg and S. Höche, “Comix, a new matrix element generator”, *JHEP* **12**, 039 (2008).
- [107] S. Schumann and F. Krauss, “A parton shower algorithm based on Catani–Seymour dipole factorisation”, *JHEP* **03**, 038 (2008).
- [108] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, “A critical appraisal of NLO+PS matching methods”, *JHEP* **09**, 049 (2012).
- [109] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, “QCD Matrix Elements + Parton Showers”, *JHEP* **11**, 063 (2001).

- [110] S. Höche, F. Krauss, S. Schumann, and F. Siegert, “QCD matrix elements and truncated showers”, *JHEP* **05**, 053 (2009).
- [111] G. Aad et al. (ATLAS), “Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”, *Eur. Phys. J. C* **76**, 581 (2016).
- [112] C. M. Bishop, *Pattern recognition and machine learning*, Information science and statistics (Springer, New York, 2006).
- [113] D. Beniaguev, I. Segev, and M. London, “Single cortical neurons as deep artificial neural networks”, en, *Neuron* **109**, 2727–2739.e3 (2021).
- [114] A. Gidon et al., “Dendritic action potentials and computation in human layer 2/3 cortical neurons”, en, *Science* **367**, 83–87 (2020).
- [115] P. Kidger and T. Lyons, “Universal Approximation with Deep Narrow Networks”, en, 22.
- [116] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, *On Calibration of Modern Neural Networks*, Number: arXiv:1706.04599 arXiv:1706.04599 [cs], Aug. 2017.
- [117] A. Komatsuzaki, “One Epoch Is All You Need”, Publisher: arXiv Version Number: 1, 10.48550/ARXIV.1906.06669 (2019).
- [118] Hecht-Nielsen, “Theory of the backpropagation neural network”, in *International 1989 Joint Conference on Neural Networks (1989)*, 593–605 vol.1.
- [119] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult”, *IEEE Trans. Neural Netw.* **5**, 157–166 (1994).
- [120] K. Fukushima, “Cognitron: A self-organizing multilayered neural network”, en, *Biol. Cybernetics* **20**, 121–136 (1975).
- [121] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines”, in *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML’10 (June 2010)*, pp. 807–814.
- [122] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks”, in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, Vol. 15, edited by G. Gordon, D. Dunson, and M. Dudík, *Proceedings of Machine Learning Research* (Nov. 2011), pp. 315–323.

- [123] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, <http://www.deeplearningbook.org> (MIT Press, 2016).
- [124] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting”, *The Journal of Machine Learning Research* **15**, 1929–1958 (2014).
- [125] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, en, arXiv:physics/0703039, July 2009.
- [126] F. Chollet et al., *Keras*, <https://keras.io>, 2015.
- [127] R. Brun and F. Rademakers, “ROOT — An object oriented data analysis framework”, en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389**, 81–86 (1997).
- [128] C. D. White, S. Frixione, E. Laenen, and F. Maltoni, “Isolating Wt production at the LHC”, *J. High Energy Phys.* **2009**, 074–074 (2009).
- [129] L. Harland-Lang, A. Martin, P. Motylinski, and R. Thorne, “Parton distributions in the LHC era: MMHT 2014 PDFs”, *Eur. Phys. J. C* **75**, 204 (2015).
- [130] S. Dulat et al., “New parton distribution functions from a global analysis of quantum chromodynamics”, en, *Phys. Rev. D* **93**, 033006 (2016).
- [131] T. Bayes Rev., “An essay toward solving a problem in the doctrine of chances”, *Phil. Trans. Roy. Soc. Lond.* **53**, 370–418 (1764).
- [132] J. Joyce, “Bayes’ Theorem”, in *The Stanford encyclopedia of philosophy*, edited by E. N. Zalta, Fall 2021 (Metaphysics Research Lab, Stanford University, 2021).
- [133] H. Lin, “Bayesian Epistemology”, in *The Stanford encyclopedia of philosophy*, edited by E. N. Zalta and U. Nodelman, Fall 2022 (Metaphysics Research Lab, Stanford University, 2022).
- [134] R. D. Cousins, “Why isn’t every physicist a Bayesian?”, en, *American Journal of Physics* **63**, 398–410 (1995).
- [135] A. Hájek, “Interpretations of Probability”, in *The Stanford encyclopedia of philosophy*, edited by E. N. Zalta, Fall 2019 (Metaphysics Research Lab, Stanford University, 2019).

- [136] K. Cranmer, *Practical Statistics for the LHC*, arXiv:1503.07622 [hep-ex, physics:physics], Mar. 2015.
- [137] S. D. Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile: précédées des règles générales du calcul des probabilités* (Bachelier, 1837).
- [138] A. D. Moivre, “De mensura sortis, seu, de probabilitate eventuum in ludis a casu fortuito pendentibus”, en, Phil. Trans. R. Soc. **27**, 213–264 (1710).
- [139] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke (ROOT), “HistFactory: A tool for creating statistical models for use with RooFit and RooStats”, (2012).
- [140] G. Cowan, *Statistical data analysis*, Oxford science publications (Clarendon Press ; Oxford University Press, Oxford : New York, 1998).
- [141] A. Collaboration, *Statistical methods for the lhc*, <https://atlas-stats-doc-dev.web.cern.ch/atlas-stats-doc-dev/statisticaltests>.
- [142] J. Neyman and E. S. Pearson, “IX. On the problem of the most efficient tests of statistical hypotheses”, en, Phil. Trans. R. Soc. Lond. A **231**, 289–337 (1933).
- [143] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, en, Ann. Math. Statist. **9**, 60–62 (1938).
- [144] *Trexfitter framework*, <https://trexfitter-docs.web.cern.ch/trexfitter-docs/>.
- [145] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke (ROOT Collaboration), *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep. (New York U., New York, 2012).
- [146] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, arXiv:physics/0306116, June 2003.
- [147] L. Moneta et al., *The RooStats Project*, arXiv:1009.1003 [physics], Feb. 2011.
- [148] *Mixed data and MC fits for realistic predictions*, https://trexfitter-docs.web.cern.ch/trexfitter-docs/model_studies/mixed/.
- [149] J. H. Friedman, “Data Analysis Techniques for High-Energy Particle Physics”, in 3rd CERN School of Computing (Oct. 1974), p. 271.

Appendix A

List of experimental uncertainties

Systematic	Description
JET_BJES_Response	Jet energy scale uncertainty for b jets
JET_EffectiveNP_Detector	Jet energy scale uncertainty (from the detector)
JET_EffectiveNP_Mixed	Jet energy scale uncertainty (various sources)
JET_EffectiveNP_Modelling	Jet energy scale uncertainty (modeling components)
JET_EffectiveNP_Statistical	Jet energy scale uncertainty (stat. components)
JET_Flavor_Composition	Jet flavor composition uncertainty
JET_JER_EffectiveNP	Jet energy resolution uncertainty
JET_JER_DataVsMC_MC16	Jet energy resolution uncertainty (from data/MC difference)
JET_Pileup_OffsetMu	Jet energy scale uncertainty on pile-up (dependent on μ)
JET_Pileup_OffsetNPV	Jet energy scale uncertainty on pile-up (dependent on the number of primary vertices)
JET_Pileup_PtTerm	Jet energy scale uncertainty on pile-up (dependent on p_T)
JET_Pileup_RhoTopology	Jet energy scale uncertainty on pile-up (dependent on ρ density)
JET_EtaIntercalibration_Modelling	Jet energy scale uncertainty on η intercalibration modelling
JET_JvtEfficiency	JVT efficiency uncertainty
FT_EFF_Eigen_B	b-tagging efficiency uncertainties (b-jet scale factor uncertainty)
FT_EFF_Eigen_C	b-tagging efficiency uncertainties (c-jet scale factor uncertainty)
FT_EFF_Eigen_L	b-tagging efficiency uncertainties (light jet scale factor uncertainty)
FT_EFF_extrapolation	b-tagging efficiency uncertainty on the extrapolation to high- p_T jets
FT_EFF_extrapolation_from_charm	b-tagging efficiency uncertainty on τ jets

Table A.1: Jet experimental uncertainties.

Systematic	Description
MUON_ID	Muon momentum resolution uncertainty from inner detector
MUON_SCALE	Muon momentum scale uncertainty
MUON_SAGITTA_RESBIAS	
MUON_EFF_ISO_SYS	Muon isolation working point scale factor uncertainty
MUON_EFF_RECO_SYS	

Table A.2: Muon experimental uncertainties.

Systematic	Description
EG_SCALE_ALL	Electron energy scale uncertainty
EL_EFF_ID_CorrUncertaintyNP	Electron identification efficiency uncertainty
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP	Electron identification efficiency uncertainty
EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	Electron trigger efficiency uncertainty

Table A.3: Electron experimental uncertainties.

Systematic	Description
MET_SoftTrk_ResoPara	Track-based soft term related to longitudinal resolution uncertainty
MET_SoftTrk_ResoPerp	Track-based soft term related to transversal resolution uncertainty
MET_SoftTrk_Scale	Track-based soft term scale uncertainty
PRW_DATASF	Data scale factor uncertainty for pile-up reweighting
Luminosity	

Table A.4: Various experimental uncertainties.

Appendix B

Additional experimental uncertainty plots

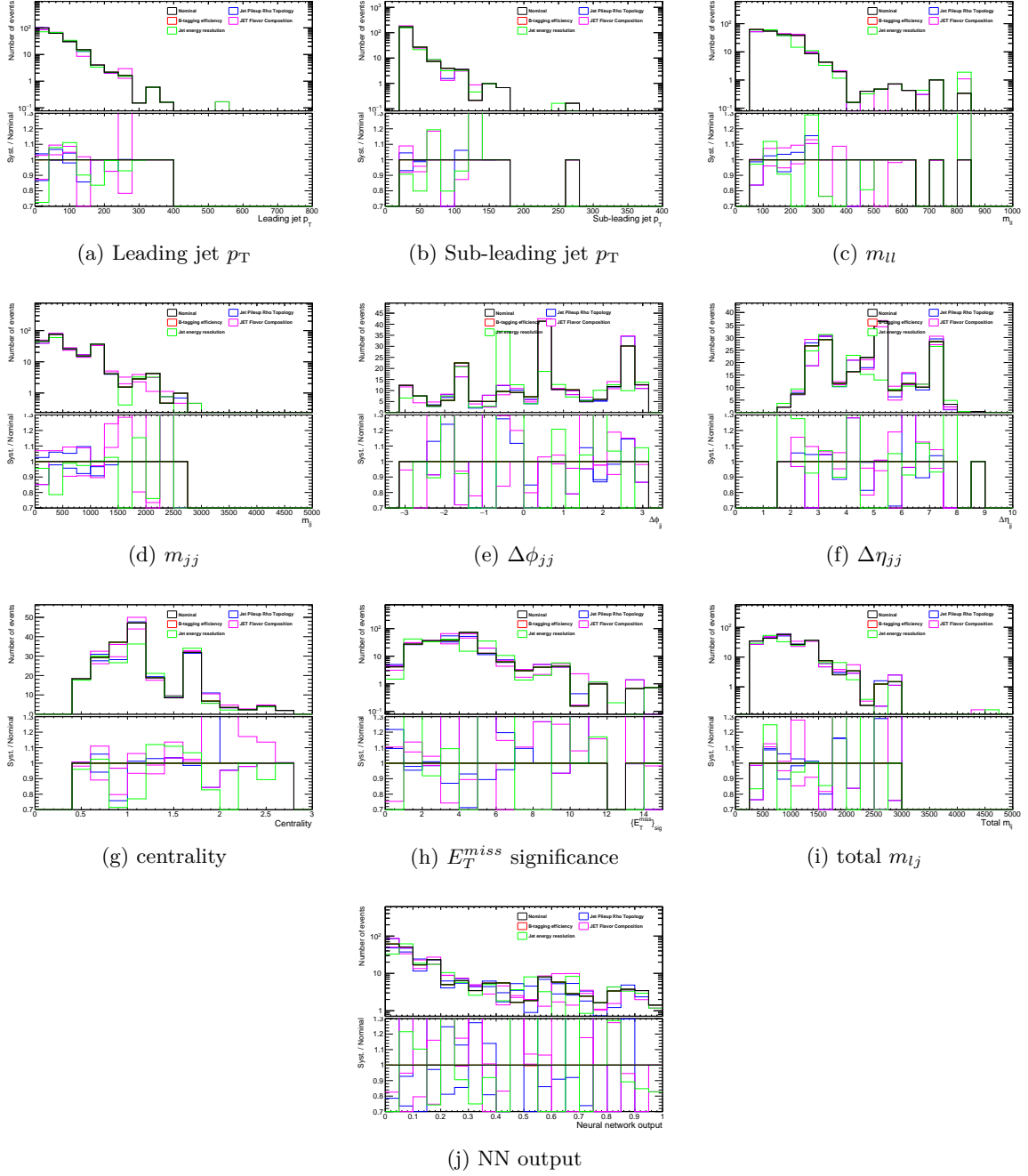


Figure B.1: Distributions for the W+jets background in the signal region for the main experimental uncertainties with the systematics/nominal ratio.

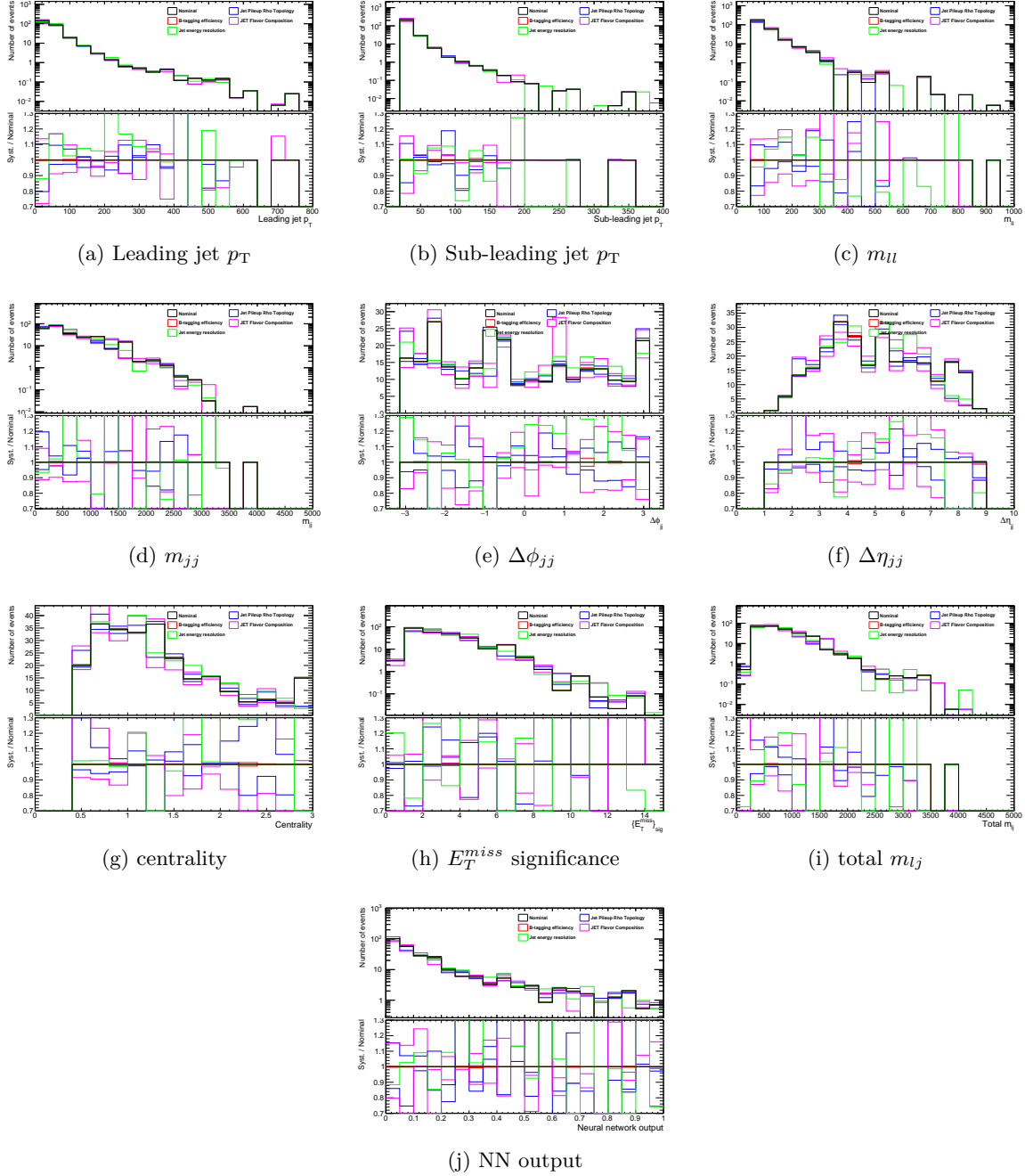


Figure B.2: Distributions for the Z+jets background in the signal region for the main experimental uncertainties with the systematics/nominal ratio.

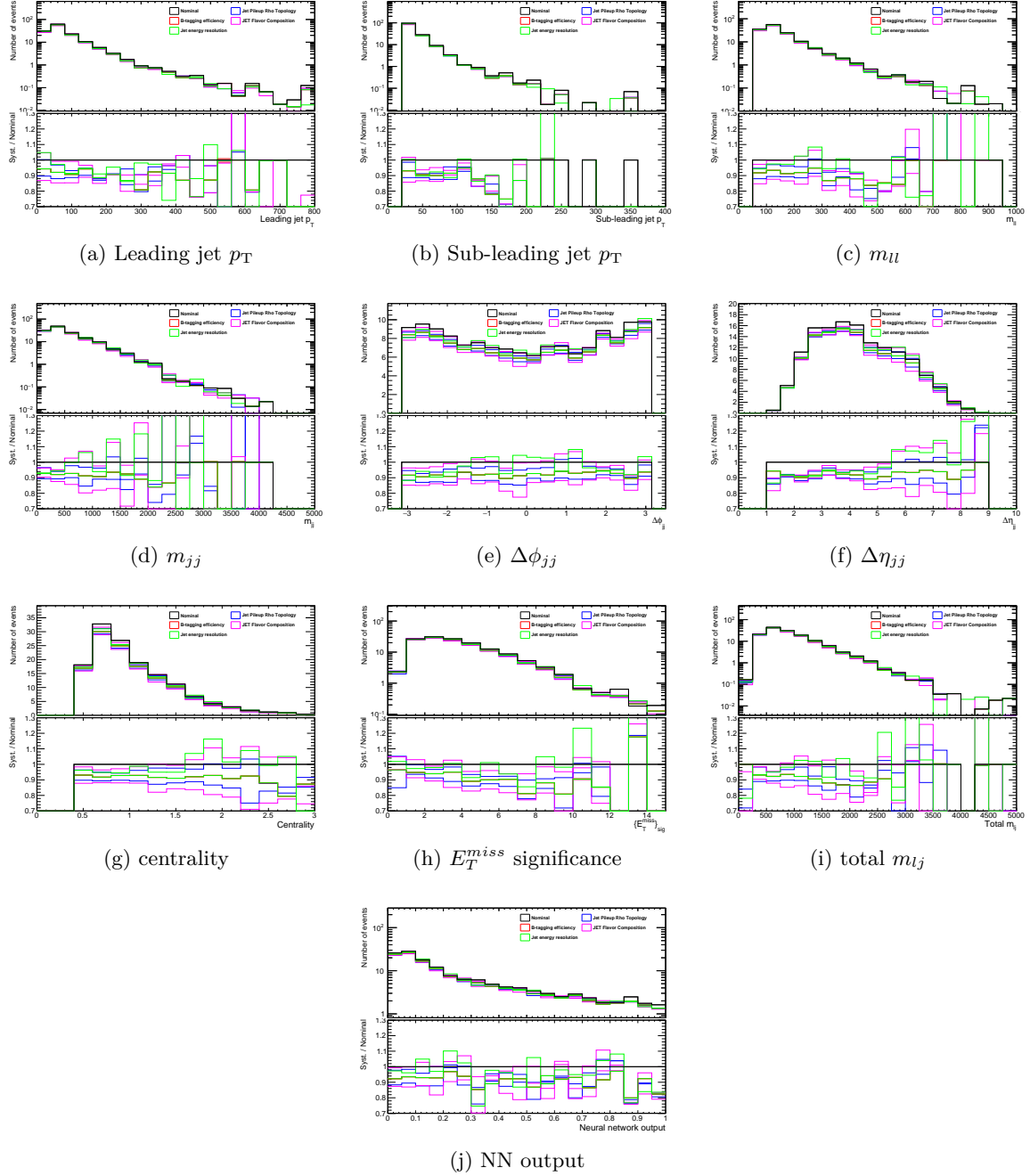


Figure B.3: Distributions for the multibosons background in the signal region for the main experimental uncertainties with the systematics/nominal ratio.

Christophe Roland

Education

- **Indiana University Bloomington, USA**

- PhD in Physics, March 2023

Thesis: *Study of W^+W^- Vector Boson Scattering with the ATLAS detector.*

Thesis advisor: Sabine Lammers

- MS in Physics, 2015

- **University of Mons, Belgium**

- MS in Physics, 2013 - magna cum laude

Thesis: *Systèmes à N corps par la méthode des champs auxiliaires.*

Thesis advisor: Claude Semay

- B.S. in physics, 2011

Research experience

- **University of Mons, Belgium**

- Graduate research assistant (2012)

Advisor: Yves J. Brihaye

Collaborated on finding how to achieve the localization of right and left chiral fermions on a brane world in the Koley-Kar brane-world model.

- Graduate research assistant (2013)

Advisor: Claude Semay

Generalized the auxiliary field method (a procedure to obtain approximate solutions, eigenvalues and eigenstates in quantum mechanics), to solve the N-body problem for identical particles in D dimensions and with arbitrary kinetics.

- **The ATLAS experiment, CERN**

- Graduate research assistant, Indiana University (2015-2023)

Advisor: Sabine Lammers

- * Analysis contact and sole analyzer for the measurement of the electroweak production of opposite sign W bosons with two jets. Tuning and training neural networks for

the analysis, estimation of the uncertainties, likelihood fit, cross-section computation and signal Monte Carlo production.

- * Estimation of EWK-QCD W+jets interference and QCD W+jets Monte Carlo production for the measurement of the electroweak production of W+jets.
- * Estimation of the fake efficiency for the W+jets background in the opposite sign WW + 0jets analysis.
- * Tuning of the low threshold and signal shaping for argon gas-mixture in the TRT, work on calibration and production of data and Monte Carlo samples for the TRT software group.

Physics schools

- The Coordinated Theoretical-Experimental Project on QCD (CTEQ) School on QCD and Electroweak Phenomenology. University of Pittsburgh, USA. July 18–28, 2017.
<http://www.physics.smu.edu/scalise/cteq/schools/summer17>
- Electroweak Symmetry Breaking (EWSB) Spring School. Maratea, Italy. Apr. 15–21, 2018.
<https://indico.cern.ch/event/673580>

Publication

- *ATLAS Collaboration et al. “Measurement of electroweak production of W^+W^- in association with jets with the ATLAS Detector”* (publication expected in 2023)
- *ATLAS Collaboration et al. “Measurement of fiducial and differential W^+W^- production cross-sections at $\sqrt{s} = 13\text{TeV}$ with the ATLAS detector”*. Eur. Phys. J. C 79.10 (2019), p. 884
<https://doi.org/10.1140/epjc/s10052-019-7371-6> arXiv: 1905.04242
- *C. Semay, C. Roland “Approximate solutions for N-body Hamiltonians with identical particles in D dimensions”*, Results in Physics 3 (2013) 231–234
<https://doi.org/10.1016/j.rinp.2013.10.001>

Talks

- *Triple and quartic gauge boson couplings* - 41st International Symposium on Physics in Collision
ATL-COM-PHYS-2022-774

<https://cds.cern.ch/record/2825030>

<https://indico.cern.ch/event/1158815/contributions/4947859/>

- *Measurements of multiboson production using the ATLAS detector* - The 2019 Phenomenology Symposium

ATL-COM-PHYS-2019-437

<https://cds.cern.ch/record/2672839>

<https://indico.cern.ch/event/777988/contributions/3410601/>

Teaching experience

- Private tutoring – Helping students for mathematics and physics courses (2011-2013)
- Indiana University, Bloomington: Associate Instructor in the Department of Physics - Grading homework and exams, tutoring students (2014-2016)

Outreach

- Website: <http://freesciences.be/>
- Collaborated with the ‘L’art de la science’ youtube channel: <https://www.youtube.com/@lartdelascience9673>