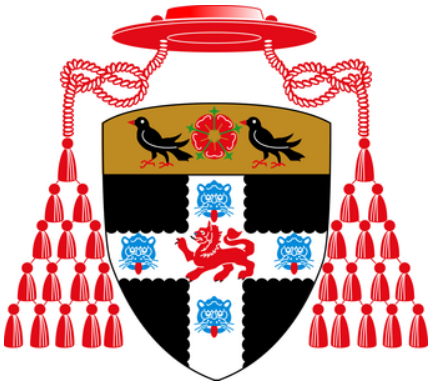
PRECISION MEASUREMENTS OF THE STANDARD MODEL; PDFS, PILE-UP, AND PRODUCTION OF W+JETS AT 13 TEV WITH THE ATLAS DETECTOR

EIMEAR CONROY

Christ Church, University of Oxford



A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Long Vacation 2023

PRECISION MEASUREMENTS OF THE STANDARD MODEL; PDFS, PILE-UP, AND PRODUCTION OF W+JETS AT 13 TEV WITH THE ATLAS DETECTOR

EIMEAR CONROY

Christ Church, University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Long Vacation 2023

Abstract

The question of the nature of the universe at its most fundamental level is one which has been with humankind since ancient times. The development of the theoretical basis of the Standard Model of elementary particles during the 20$^{\text{th}}$ century, and its subsequent experimental validation, is a triumph of modern physics. However, the Standard Model is known to be incomplete, and, in the Large Hadron Collider era at the high-energy and high-intensity frontier, our understanding and modelling of the Standard Model has increasingly become limiting for searches for physics beyond it. In this thesis, a measurement performed using 36.2 fb^{-1} of integrated luminosity recorded by the ATLAS experiment during Large Hadron Collider Run 2 of a crucial benchmark process for Standard Model modelling is presented; W boson production in association with jets. This is the first such measurement in the $W \to \mu\nu$ decay channel at $\sqrt{s}$ =13 TeV. Differential cross-sections, in observables sensitive to the modelling of quantum chromodynamics, are presented for both charge-independent and charge-separated W production, and compared to two cutting-edge calculations evolved to next-to-next-to-leading order in quantum chromodynamics. In addition, this process strongly constrains fits of the Parton Distribution Functions of the proton; an analysis of the structure of the proton using several previous ATLAS measurements, including W boson production in association with jets at $\sqrt{s}$=8 TeV, will also be presented.

Dedicated to my family,

to the memory of my grandparents, Betty, Donie, and Jim,

and to my Granny Mary;

“What’s for you won’t pass you”.

Acknowledgements

No man is an island, and no Ph.D journey would ever be possible alone. As such, I would like to take this opportunity to thank the army of incredible individuals and organisations, without whom this document, and what it represents, would never have been achieved.

Thank you to my professors and mentors at University College Dublin, for igniting my passion for experimental physics, and who believed in the crazy pipe-dream of the Irish girl who wanted to be a particle physicist before even she did; and to my wonderful Dublin friends, who taught me what it feels like to find your tribe.

To all the rowers and coxes at Christ Church Boat Club, thank you for the belonging you've made me feel since the very beginning, and for introducing me to a new lifelong passion, my Oxford journey would not have been the same without this incredible family. Thank you to my LTA family, for making my time in Switzerland the adventure of a lifetime; to Holly McGrath, for managing to be the best Meyrin flatmate I could have ever hoped for, and somehow, an even better friend.

Thank you to all the students and postdocs of Oxford particle physics (both bona fide and honorary), for being a constant source of joy, friendship and support throughout

the past four years; to Bill Balunas, Gabriel Gallardo, Claudia Merlassino, Aaron O'Neill, Holly Pacey, Santiago Paredes, and Siyuan Yan, for bringing the light in between the dark shadows of a pandemic; and to Chris Brown, Tim Bruckler, Maggie Chen, Ynyr Harris, Mei Holmberg, Simon Koch, Alessandro Ruggiero, Beth Spear and Iza Veliscek for all the laughter this year, terrible pub quiz results, ski antics and your enduring friendship.

Thank you to my partners in crime from the Christ Church GCR: Alex Ilsley, Ben Joseph, Victoria Walls and Arthur Wotton, for welcoming me back to Oxford with open arms, and for making this town feel like a home; and to Peter Watson, for being by my side from day one.

Thank you the the Oxford Department of Particle Physics for providing the highest standard of academic training I could have ever hoped for; to the past and present members of the Standard Model and Beyond subgroup: Alan Barr, Claire Gwenlan, Gabriel Gallardo, Radek Grabarczyk, Ynyr Harris, Claudia Merlassino, Koichi Nagai, Aaron O'Neill, Gavin Pownall, Alessandro Ruggiero, Jonas Wuerzinger and Tom Yildirim, for both your wonderful camaraderie and incredible insights into my physics problem of the week; to Mandy Cooper-Sarkar, for so patiently sharing your wealth of knowledge and experience; and to Todd Huffman, for being an excellent partner in cat-herding the Oxford ATLAS group.

Thank you to the ATLAS Collaboration, for welcoming me with open arms, learning how to become a real scientist as part of this community has been one of the most rewarding experiences of my life. Thank you to the colleagues who took me under their

wing: Bill Balunas, Hugo Beauchemin, Francesco Giuli, Max Swiatlowski, Holly Pacey and others, I have learned so much from you all. Thank you to the UK Science and Technology Facilities Council and the Oxford Scatcherd European Scholarship, for funding this incredible PhD journey, without their support this would not have been possible.

To my supervisor, Claire Gwenlan - one paragraph could never do the past four years justice. Thank you for seeing potential in me, and for your unwavering belief. Thank you for your peerless academic guidance, for your in patience answering every question, no matter how trivial, and for correcting every mistake, no matter how basic. Thank you for your steadfast encouragement, support and friendship; I truly could not have wished for a better supervisor, and I cannot wait to continue working with you as a postdoc.

Finally, thank you to my family, I love you all more than words can say. To my grandparents, Betty, Donie and Jim, I hope, somewhere, I've made you proud. To Granny Mary, without your encouragement I never would have taken this plunge, thank you for being a rock of love and support my entire life. Finally, to Mam, Dad, Ultan and Maeve, I wouldn't be half of who I am today without each and every one of you, from the bottom of my heart, thank you; *go raibh míle maith agaibh, beidh sibh i gcónaí i mo chroí.*

Preface

The work presented in this thesis has been performed as a member of the ATLAS experiment at CERN, which consists of over 3000 active members; membership of such a large collaboration therefore necessitates the sharing of work. As such, the theoretical and experimental background of a Standard Model measurement performed using the ATLAS experiment, as presented in Chapters 1-3, is not the author's own work. The results presented in Chapters 4-10 are predominantly the author's own work, except where explicitly indicated in the text or with a citation in a figure caption; this is a non-exhaustive description of the total body of research completed throughout the author's Ph.D study. The author's contributions to these chapters are summarised below; figures presenting work performed by the author are easily identifiable by an 'ATLAS Work in progress' plot title:

Balance JVT pile-up tagger, lead tool developer

- **Chapter 4, 'Pile-up tagging in the ATLAS central region'**: The creation of a new pile-up jet tagging tool for the ATLAS central region, developing on existing code for the ATLAS forward region; the validation of the pile-up jet content of the topology used for performance testing the new tool; initial performance studies of the tool, including its impact on missing transverse energy reconstruction.
  - This is an interim result, with the outcome of a full performance study and calibration of the latest version of the tool expected to be published in 2024.

W+jets production at 13 TeV, $W \to \mu\nu$ channel primary analyser

- **Chapter 5, 'W+jets analysis introduction'**: Optimising the binning scheme for the observables to be unfolded, in concert with the electron channel analysts, evaluation of the expected detector smearing associated with these binnings; implementation of the analysis selections in the muon channel;

- **Chapter 6, Background estimation:** Evaluation of MC simulated backgrounds in the W+jets muon channel; determination of the selections used to define the Signal, Control and Validation regions in the muon channel; evaluation of the signal + electroweak background contribution to the Control Region and Validation Region; determination and implementation of the prescaled trigger strategy for the muon channel multijet background calculation; determination of the variables used to parameterise multijet background efficiencies in the muon channel and optimisation of the efficiency binning schemes; the evaluation of the multijet background in the Signal Region and comparison to a simulated sample of dijet events; comparison of the measured data in the muon channel Validation Region to Standard Model predictions, which include an estimate on the multijet background;

- **Chapter 7, 'Unfolding':** Evaluation of the migration matrices for the W+jets muon channel, including the implementation of a jet matching scheme; performance of the technical unfolding closure test in the muon channel;

- **Chapter 8, 'Uncertainties':** Evaluation of the statistical and systematic uncertainties in the W+jets muon channel, including those associated with the multijet background estimate;

- **Chapter 9, 'Results':** Comparison of the measured data in the W+jets muon channel to two Standard Model predictions, including statistical and systematic uncertainties; determination of charge independent $W \to \mu\nu$+jets differential cross-sections, charge-separated differential cross-sections and ratios of the W^+/W^- differential cross-sections, for inclusive jet multiplicities up to 5 jets, including an evaluation of the statistical and systematic uncertainty;

  – This is an interim result; the full paper will include a $W \to e\nu$ channel analysed by collaborators in other institutions, a combination of channels, and a QCD analysis, which will be performed by the author. Publication is expected in 2024.

<u>Parton Distribution Function fits</u>

- **Chapter 10, 'Parton Distribution Functions with ATLAS data'**: Implementation of the two-point jet systematic uncertainty decorrelation in the ATLASpdf21 Parton Distribution Function fit, evaluation of the impact of each scheme on the shape of the fitted parton distributions and on the fit χ^2.

  – This analysis has been made public, and was published by European Physical Journal C in 2021 [1].

Contents

Introduction

"Tús maith, leath na hoibre
(A good start is half the work)"

— Seanfhocal
(Irish proverb)

The question of the fundamental nature of the universe is one which has been with humankind since ancient times. Indeed, the idea that the matter which comprises the universe around us is composed of elementary building blocks is as old as natural philosophy itself, with the ancient Greek philosophers of the doctrine of atomism proposing that the physical universe is composed of invisible, *atoms*, literally meaning 'indivisible', which determined the properties of the matter they made up.

Such a concept would prove to be incredibly prescient as the philosophical argumentation of natural philosophy gave way to the empirical observations of natural science. Atomic theory was provided experimental support in the 19th century through the development of stoichiometry by John Dalton, which indicated that each chemical element is composed of a unique type of constituent. Believing these particles to be indivisible, Dalton and his contemporaries named them 'atoms'. Also in the 19th century came Dmitri

Mendeleev's periodic table, introducing the idea of classifying the fundamental components of nature on the basis of their characteristics.

However, the assumption that atoms were the indivisible fundamentals of nature would be challenged as the turn of the 20^{th} century gave rise to the birth of experimental particle physics. From the discovery of the electron by Thompson, to the scattering experiments of Rutherford, Geiger and Marsden, leading to the discovery of the proton, and the uncovering of the menagerie of elementary fermions and vector bosons by cosmic ray and collider experiments as the century progressed, the "periodic table" of truly fundamental particles, to the extent our current understanding, began to take shape. In concert with this century of discovery arose the theoretical framework of the Standard Model, which achieved a description of the elemental universe with an accuracy and robustness of which the ancient forebearers of the discipline could only have dreamt.

In the modern era, high-energy collider scattering experiments, exploiting Einstein's seminal relationship between energy at matter, continue to advance the energy and intensity frontiers of particle physics. However, while many may consider the 2012 discovery of the Higgs boson using such a machine, the Large Hadron Collider, as crowning achievement of the Standard Model, the evidence for physics beyond it is extensive and compelling. Increasingly, the limitations of exploring such new horizons originate from our understanding of the Standard Model itself; from the intricacies of the internal structure of the proton to the stubbornly challenging task of predicting the behaviour of quantum chromodynamics, there is much still to uncover within this nominally "complete" model.

This thesis is focused on such a endeavour. Chapter 1 presents an overview of the relevant theoretical background. Chapter 2 describes the experimental machinery employed: the Large Hadron Collider and the ATLAS experiment. A description of particle reconstruction in the ATLAS detector is presented in Chapter 3, with a focus on the identification and rejection of pile-up jets in Chapter 4. Chapter 5 introduces the primary analysis of this thesis, which is a crucial benchmark measurement for the modelling of the Standard Model: The production of W bosons in association with jets at $\sqrt{s}$=13 TeV, for the first time in the $W \to \mu\nu$ decay channel at this centre-of-mass energy. The estimation of the backgrounds for this process is described in Chapter 6, while correction for detector effects is discussed in Chapter 7. The uncertainties associated with this measurement are detailed in Chapter 8 and, ultimately, the measured results are presented in Chapter 9. Finally, a description of measurement of the structure of the proton using a variety of ATLAS measurements is presented in Chapter 10.

Chapter 1

Theoretical framework

"Three quarks for Muster Mark!"

— James Joyce, 'Finnegan's Wake'

Our current best understanding of the nature of matter at its most essential is described by the *Standard Model* (SM). This foundational model of particle physics describes the 12 fundamental[1] particles that comprise the constituent matter of the universe, and the fundamental forces through which the particles interact, which themselves are mediated by force-carrying particles.

In the SM paradigm, illustrated in Figure 1.1, particles are initially categorised on the basis of their spin. Spin-$\frac{1}{2}$ particles obey Fermi-Dirac spin statistics, whereby indistinguishable particles are prevented from occupying the same quantum state in accordance

[1] In this context, the term 'fundamental' refers to an entity which cannot be broken down further into simpler constituent components.

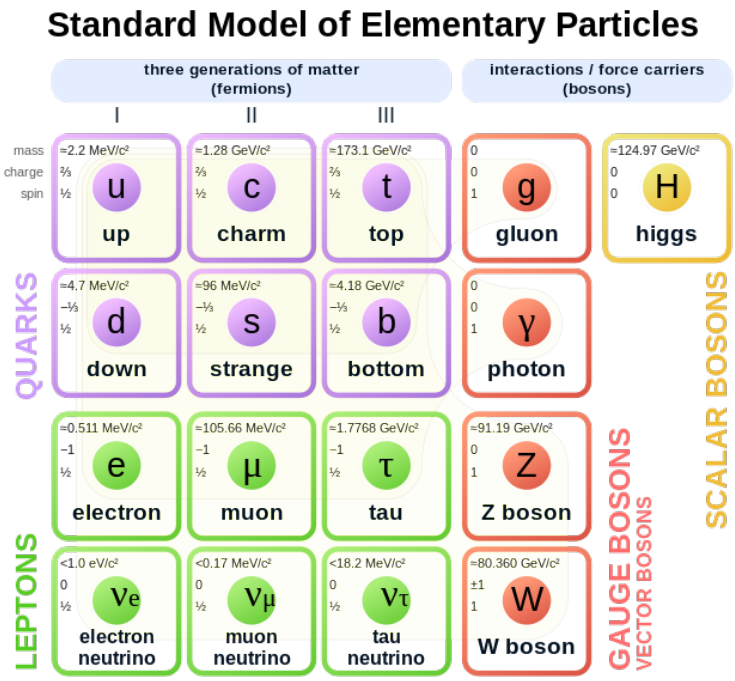


Figure 1.1: An overview of the particle content of Standard Model. Image from [2] using data from [3].

with the Pauli Exclusion Principle. Therefore, they are known as *fermions*. Conversely, two indistinguishable integer spin particles, known as *bosons*, are permitted to simultaneously hold the same quantum numbers and therefore obey Bose-Einstein statistics. According to the Standard Model, fermions are the particles which form matter, while bosons mediate the *fundamental forces*; electromagnetism, strong, weak. At present, no 'graviton' mediator of the gravitational force has been experimentally observed, nor has the force been incorporated into the mathematics describing the SM. However, the gravitational force has been robustly documented in experimental physics and described by the theory of general relativity; the union of the relativistic description of gravity with the Standard Model is an active area of research. However, on the scale of high-energy particle colliders, gravitational forces have a negligible effect on observations and are generally

therefore neglected.

The fermions are further subdivided into two sub-categories: *quarks*, which experience the *strong force* mediated by massless *gluons* that couple to *colour* charge, and *leptons*, which do not. Additionally, all quarks and charged leptons experience the *electromagnetic* (EM) interaction, which is mediated by the massless *photon*. All 'up-type' quarks on the first row on Figure 1.1 have an EM charge of $+\frac{2}{3}$, while 'down-type' quarks in the second row have a charge of $-\frac{1}{3}$. In the lepton sector, *electrons, muons* and τ*-leptons* have a charge of -1, while the significantly lighter neutrinos are electrically neutral. For each fermion, there also exists a conjugate *anti-particle* with an opposite-sign EM charge.

The leptons and quarks are further divided into three *generations*, with fermion mass increasing for each successive generation[2]. The *weak force* mediates transformations between different fermions of the same generation, with a small amount of generational mixing permitted (see Section 1.4.1 for a description of quark-sector mixing). The weak force is mediated by the massive $W^{\pm}, Z$ bosons, and can be shown to be intrinsically linked with the EM force via *electroweak unification*, discussed in Section 1.4.2.

Finally, electroweak bosons and charged fermions acquire mass through the spontaneous breaking of electroweak symmetry (see Section 1.4.3), resulting in a universal *Higgs field* and an associated massive, spin-0 *Higgs boson*. The Higgs boson was the final SM particle to be experimentally observed, which was accomplished for the first time by the ATLAS and CMS collaborations in 2012 [4][5].

[2]It has not yet been discovered whether or not the very light neutrinos obey the same mass hierarchy as the other SM fermions; however, a detailed discussion of neutrino mass is beyond the scope of this thesis.

1.1 The Standard Model

Mathematically, the Standard Model is the quantum field theory of elementary particles, arising from perturbations of quantum fields, and their interactions, which are expressed using the Lagrangian formalism. To begin, the dynamics of relativistic free fermions are described in a Lorentz invariant form; as spin-half particles, fermion dynamics are governed by the Dirac equation:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0, \tag{1.1}$$

where γ^μ represents the Lorentz invariant 4×4 *Dirac matrices*, ∂_μ is the four-derivative, m is the fermion mass, and $\psi = \psi(x)$ is the 4×1 fermion field spinor, which is a function of spacetime x.

The Dirac equation has four solutions. Two correspond to the spin-up and spin-down positive-energy *particles*, and two correspond to negative-energy particles which propagate backwards in time. The negative-energy solutions can also be expressed as positive-energy, physical *anti-particles*, which propagate forward in time and have a sign-inverted EM charge to their particle counterparts. The probability current j^μ, which describes fermion interactions through the exchange of a force-carrying mediating particle, can be expressed in terms of the Dirac matrices and spinors and will be proportional to:

$$j^\mu \propto \bar{\psi}\gamma^\mu\psi,$$

where the adjoint spinor $\bar{\psi} = \psi^\dagger\gamma^0$, $\psi^\dagger = (\psi^*)^T$ is the Hermitian conjugate of the spinor and γ^0 is the first Dirac matrix.

Secondly, the interactions between the fermions must be incorporated into the model. This is accomplished by the formulation of the Standard Model as a *gauge theory*; a category of field theory where the Lagrangian is invariant under local gauge transformations, and therefore also the dynamics of the system. For the purposes of this thesis, the term 'Lagrangian' is used to refer to the Lagrangian density $\mathcal{L} = \mathcal{L}(\psi(x), \partial_\mu\psi(x))$. A *gauge* is a specific additional requirement on potentials in a physical system, intended to mediate unnecessary degrees of freedom. One example from classical electrodynamics is the Coulomb gauge, which imposes the requirement that the vector potential of an electric field $\mathbf{A}$ be divergenceless; $\nabla \cdot \mathbf{A} = 0$, which allows $\mathbf{A}$ to be specified uniquely.

A *gauge transformation* is a transformation between choices of gauge according to a particular *gauge symmetry* which obeys the Lie algebra of group generators. In the quantum-mechanical context, a symmetry is the requirement that all physical predictions be invariant under a given wavefunction transformation $\psi \to \psi' = \hat{A}\psi$, where $\hat{A}$ is some generic transformation. A 'global symmetry' implies invariance under a transformation which is identical at all space-time points, whereas a more stringent 'local symmetry' can be a function of space and time. According to Noether's Theorem [6], local gauge symmetries in the Lagrangian or Hamiltonian are always associated with conserved currents and physical quantities.

To ensure the Lagrangian remains invariant under local gauge transformations, addi-

tional *gauge fields* are introduced for each group generator. When this theory is quantised, the gauge fields subsequently give rise to *gauge bosons*, which correspond to force-carrying particles. The properties of these gauge fields are what dictate the features of the gauge bosons and, subsequently, the interactions they mediate.

The Standard Model is a non-abelian [3] gauge theory which has the symmetry group $U(1)_Y \times SU(2)_L \times SU(3)$. This introduces a total of 12 gauge bosons: 1 photon, 3 electroweak vector bosons and 8 gluons. All fundamental particles and their interactions are represented as terms of a universal Standard Model Lagrangian $\mathcal{L}_{SM}$. The following Sections 1.2-1.4 will detail the contributions to $\mathcal{L}_{SM}$ arising from the different interactions of fundamental particles, and how they result in the particle content of the Standard Model.

1.2 Quantum Electrodynamics

Quantum Electrodynamics (QED) is the abelian gauge theory of electromagnetic interactions, which describes the behaviour of particles which carry an EM charge. This charge can be either positive or negative, with like charges repelling and opposite charges attracting. These interactions are mediated by the photon, which is itself electromagnetically neutral. The QED Lagrangian is constructed from two terms: one for the Dirac equation, representing fermion dynamics, and one derived from the EM field potential:

$$\mathcal{L}_{\mathrm{QED}} = \mathcal{L}_{\mathrm{Dirac}} + \mathcal{L}_{\mathrm{EM}}. \tag{1.2}$$

[3] 'Non-abelian' implies that the operators of a particular symmetry group do not commute; $\hat{a}\hat{b} \neq \hat{b}\hat{a}$.

The $\mathcal{L}_{\text{Dirac}}$ in Eq. 1.2 can be derived from Eq. 1.1 for a free-particle spinor field:

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi. \tag{1.3}$$

QED obeys a U(1) gauge symmetry; therefore, to ensure Eq. 1.3 is invariant under a U(1) local gauge transformation, the following transformations must be made to Eq. 1.3:

$$\begin{aligned} \psi &\xrightarrow{\text{U(1)}} \psi' = \exp\left[ie\alpha(x)\right]\psi, \\ \partial_{\mu} &\xrightarrow{\text{U(1)}} D_{\mu} = \partial_{\mu} + ieA_{\mu}, \end{aligned}$$

where e is the EM coupling strength and $\alpha(x) \in \mathbb{R}$ is the U(1) group generator. D_{μ} is known as the *covariant derivative*, which introduces A_{μ}, a vector gauge field for the EM interaction which must transform as:

$$A_{\mu} \xrightarrow{\text{U(1)}} A'_{\mu} = A_{\mu} - \partial_{\mu}\alpha(x), \tag{1.4}$$

under U(1) symmetry to retain a gauge invariant Lagrangian. The gauge-invariant expression of Eq. 1.1 subsequently becomes:

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - e\bar{\psi}\gamma^{\mu}A_{\mu}\psi,$$

which now contains a term for the interaction between the fermion and the photon field A_{μ}. A kinetic term for A_{μ} can be derived from the relativistic expression of Maxwell's Equations assuming a free photon field:

$$\mathcal{L}_{\mathrm{EM}} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu},$$

where $F^{\mu\nu} = (\partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu})$ is the gauge invariant EM field strength tensor. The total QED Lagrangian therefore becomes:

$$\mathcal{L}_{\mathrm{QED}} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - e\bar{\psi}\gamma^{\mu} A_{\mu}\psi - \frac{1}{4} F^{\mu\nu} F_{\mu\nu}. \tag{1.5}$$

From this Lagrangian, all of the physics of classical, relativistic and quantum electromagnetism can be recovered. As Eq.1.5 does not contain a photon mass term $\frac{1}{2} m_{\gamma}^{2} A^{\mu} A_{\mu}$, it can be inferred from the QED Lagrangian that mass of the photon m_{γ} must be zero; indeed, the addition of a photon mass term would violate local gauge invariance.

Finally, guided by the form of Eq. 1.1, the expression

$$j^{\mu}_{\mathrm{EM}} = e\bar{\psi}\gamma^{\mu}\psi.$$

for the conserved EM current can be extracted from Eq. 1.5.

1.3 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the quantum field theory associated with particles which interact via the strong force. In QCD, the conserved quantity is *colour* $\in$ {red,

green, blue$\}$, a quantum number only possessed by quarks. The presence of three colour charges is motivated by measurements of the ratio

$$R = \frac{\sigma(e^+e^- \to q\bar{q})}{\sigma(e^+e^- \to \mu^+\mu^-} = 3\sum_q e_q^2,$$

where the index q runs over all quark flavours and e_q is the EM charge of quark q, which indicate that each quark flavour is associated with a triplet of colour charges. The three QCD colour charges lead to an SU(3) local gauge symmetry, and three additional degrees of freedom of the fermion spinor field.

To maintain the invariance of the Dirac Lagrangian under an SU(3) gauge transformation, the quark spinor field and four derivative must transform as

$$\begin{aligned}\psi(x) &\xrightarrow{\text{SU(3)}} \psi' = \exp[ig_{\text{S}}\hat{\mathbf{T}}\cdot\boldsymbol{\alpha}(x)]\psi(x),\\ \partial_\mu &\xrightarrow{\text{SU(3)}} D_\mu = \partial_\mu + ig_{\text{S}}\hat{\mathbf{T}}\cdot\mathbf{G}_\mu,\end{aligned}$$

where, g_{S} is the coupling strength of the strong interaction, $\alpha^a(x)$ are eight functions of spacetime, and $\hat{\mathbf{T}} = \{T^a\}$ are the eight generators of SU(3), related to the eight 3×3 Gell-Mann matrices, $\mathbf{\Lambda}$, by $T^a = \frac{1}{2}\Lambda^a$. The requirement that the QCD Lagrangian is invariant under SU(3) transformation introduces eight new *gluon* fields, G_μ^a, which must transform as

$$G_\mu^c \xrightarrow{\text{SU(3)}} G_\mu^{c'} = G_\mu^c - \partial_\mu\alpha_a - g_{\text{S}}f_{abc}\alpha_a G_\mu^b,$$

where f_{abc} are the structure constants of the SU(3) group, defined by the commutation relation $[\Lambda_a, \Lambda_b] = 2if_{abc}\Lambda_c$. Similarly to the QED Lagrangian, these new $\mathbf{G}_\mu$ fields are associated with a kinetic term, resulting in the QCD Lagrangian

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \bar{\psi}_q(i\gamma^\mu\partial_\mu - m_q)\psi - \frac{g_{\mathrm{S}}}{2}\bar{\psi}_q\gamma^\mu\Lambda^a G^a_\mu\psi_q - \frac{1}{4}G^{\mu\nu,a}G^a_{\mu\nu}, \tag{1.6}$$

where $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_{\mathrm{S}}f_{abc}G^b_\mu G^c_\nu$ is the gluon field strength tensor. Similarly to the case for QED, there is no mass term for the gluon field in Eq. 1.6, as this would result in the violation of the local gauge invariance under SU(3) symmetry. However, a crucial difference of the QCD field strength tensor from its QED equivalent is the presence of the $g_{\mathrm{S}}f_{abc}G^b_\mu G^c_\nu$ term, introduced via the non-abelian nature of the SU(3) group, in contrast to the abelian nature of the U(1) group. The presence of both G^b_μ and G^c_ν in this extra term provides a mechanism for gluon self-interaction in QCD, which will be discussed in Section 1.3.1.

The conserved colour current can be read directly from Eq. 1.6 as

$$j^{\mu,a}_{\mathrm{QCD}} = \frac{g_{\mathrm{S}}}{2}\bar{\psi}_q\gamma^\mu\Lambda^a\psi_q.$$

Conventionally, the coupling strength of strong interactions is expressed using the *strong coupling constant*, α_{S} where $\alpha_{\mathrm{S}} = \frac{g^2_{\mathrm{S}}}{4\pi}$. Intimately connected to the definition of the strong coupling constant is the concept of *renormalisation*, which is used to resolve divergences which arise from the infinite number of possible Feynman diagrams, corresponding to virtual loops, at a given interaction vertex. Without renormalisation, an infinite

number higher-order corrections would be required to precisely calculate the vertex coupling term in cross-section calculations. However, this is prevented by instead absorbing the corrections into the definition of α_{S} , which will thereafter necessarily depend on Q^2 , the negative squared for momentum transfer, $Q^2 = -q^2$.

If α_{S} is known at some starting *renormalisation scale*, μ_{R}^2, the evolution, or *running*, of α_{S} with Q^2 can be expressed at leading-order as:

$$\alpha_{\mathrm{S}}(Q^2) = \frac{\alpha_{\mathrm{S}}(\mu_{\mathrm{R}}^2)}{1 + B\alpha_{\mathrm{S}}(\mu_{\mathrm{R}}^2)\ln\left(\frac{Q^2}{\mu_{\mathrm{R}}^2}\right)}, \tag{1.7}$$

where $B = \frac{11N_c - 2Nf}{12\pi}$, $N_c = 3$ is the number of colours and $N_f = 6$ is the number of quark flavours. Eq. 1.7 results in an α_{S} that decreases sharply with Q^2 . Renormalisation can also be applied to QED vertices; however, the lack of photon self-interaction terms in the QED Lagrangian results in an expression for $e(Q^2)$ that increases with Q^2.

1.3.1 Impacts of gluon self-interaction

The term proportional to $G_\mu^b G_\nu^c$ in the gluon field strength tensor, arising due to non-abelian nature of the SU(3) group, results in triple and quartic gluon couplings. These gluon *self-interaction* vertices imply that gluons carry colour charge, unlike in QED, where the photon is electrically neutral. These gluon self-interactions also preclude the existence of free quarks, in a hypothesis called *colour confinement*.

Given two quarks interacting strongly via gluon exchange, gluon self-interaction per-

mits those gluons to also interact with each other via virtual gluon exchange. If the two quarks were subsequently pulled apart, the gluon self-interactions would squeeze the colour field lines into a 'colour tube', in which the stored energy is proportional to the distance between the quarks. Therefore, beyond a very small separations, it becomes energetically favourable for free quarks to arrange themselves into net-colourless bound states called *hadrons* via a process called *hadronisation*. At particle colliders, colour confinement prevents the observation of free quarks or gluons produced in high-energy interactions. Instead, as illustrated in Figure 1.2, as high energy quarks separate, they hadronise to create new colourless bound states, which in turn themselves fragment and hadronise. After many iterations, this process produces collimated showers of hadrons called *jets*, which are the objects observed by detectors.

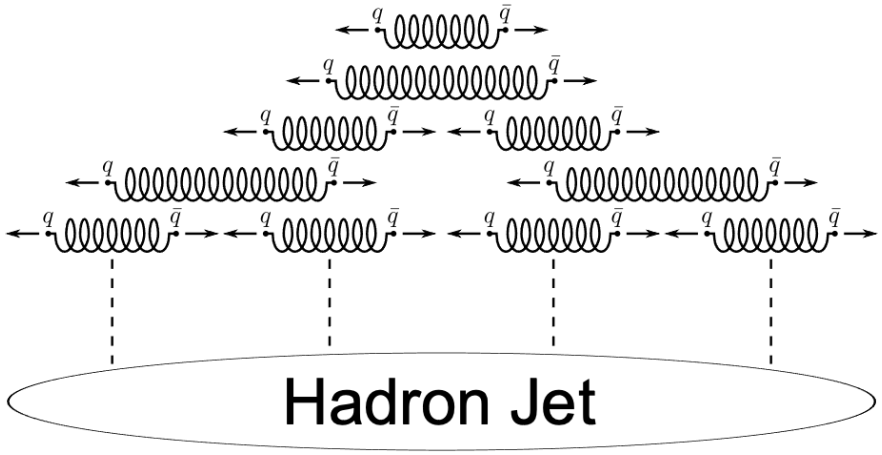


Figure 1.2: An illustration of hadronic jet production as a result of the colour confinement of quarks. Figure from [7].

The above discussion of confinement is mainly applicable to the low-energy regime, where α_{S} is too large to be calculated using perturbation theory. However, as α_{S} quickly reduces in strength with increasing Q^2 as described by Eq. 1.7, perturbation theory can be used in the high-energy regime and high-energy quarks, or strong interactions on short

distance scales, can be treated as quasi-free. This feature of QCD is known as *asymptotic freedom*. Typically, QCD cross-sections are calculated perturbatively in powers of α_{S}; however, even at high energies, α_{S} is not sufficiently small that higher-order corrections can be neglected. Therefore, fixed-order QCD calculations used in LHC experiments will generally be calculated at next-to leading-order (NLO), or next-to next-to leading-order (NNLO) in α_{S} where possible.

1.3.2 Parton Distribution Functions

Unlike the fermions discussed so far in this chapter, the protons used in hadron colliders such as the LHC are not 'fundamental' objects. Introduced by Feynman in 1969 [8], the *quark-parton model* (QPM) proposes that protons are a bound state of point-like elementary constituents. Throughout the late 20$^{\text{th}}$ century, experimental tests of the parton properties showed that the partons proposed by the QPM are in fact the quarks and gluons described by QCD.

The QPM introduces an explicit form for the *structure functions* of the proton and a dependence of the *Parton Distribution Functions* (PDFs), $f_i(x)$, which give the probability that parton i is carrying x fraction of the proton momentum, known as the *Bjorken* x variable. In the QPM, the PDFs are initially unknown, but universal for a given type of hadron.

The QPM also introduces a set of *sum rules* for the PDFs. Firstly, the *momentum sum rule* states that the summed contribution from each parton must return the total

momentum of the proton:

$$\sum_{q,\bar{q},g} \int_0^1 dx\ x f_i(x) = 1. \tag{1.8}$$

It is strong evidence for the gluon content of the proton that, without the inclusion of the gluon, the integral in Eq. 1.8 is measured to be approximately 0.5. Furthermore, the *number sum rules* state that the quark PDFs, integrated over x, should return the *valence* content of the proton:

$$\begin{aligned}
\int_0^1 dx\ (f_u(x) - f_{\bar{u}}(x)) &= 2, \\
\int_0^1 dx\ (f_d(x) - f_{\bar{d}}(x)) &= 1, \\
\int_0^1 dx\ (f_s(x) - f_{\bar{s}}(x)) &= 0,
\end{aligned}$$

i.e. three *valence quarks* which carry the quantum numbers of the proton; two up and one down resulting on a net unit positive charge. These valence quarks are bound together by gluons, which have the ability to split to generate a *sea* of virtual $q\bar{q}$ pairs primarily at low-x.

For the QPM to be valid, the quarks must be considered to be semi-free; however as discussed in Section 1.3.1, colour confinement implies that free quarks cannot be experimentally observed. This apparent contradiction is resolved by another result of QCD from Section 1.3.1; asymptotic freedom leads to an α_{S} that is sufficiently weak on the

scale of the proton radius to consider the quarks quasi-free particles, while also increasing with distance to recover quark confinement at larger distance scales.

Another basic assumption of the QPM is the *factorisation theorem*[4], which states that the cross-section for a particle-level, hard scatter process can be written as a convolution of pertubatively calculable parton-level cross-sections, and the non-perturbative PDFs:

$$\sigma_{pp\to X} = \sum_{i,j\in q\bar{q},g} \int_0^1 dx_i \int_0^1 dx_j f_i(x_i,\mu_F^2) f_j(x_j,\mu_F^2) \sigma_{ij\to X}(x_i,x_j,\mu_F,\mu_R), \tag{1.9}$$

where μ_R is the renormalisation scale discussed in Section 1.3.1 used to absorb divergences in the high-energy, *ultra-violet* (UV) regime. The dependence of the PDFs on μ_F, known as the *factorisation scale*, arises from the the application of QCD theory to the QPM in the *QCD-improved QPM*.

For a general parton-parton interaction, QCD theory allows for radiative interactions to take place before or after the 'hard-scatter' interaction, which give rise to divergences in the cross-section in the Infra-Red and Collinear (IRC) regime. In a procedure similar to renormalisation, these divergences are regularised by absorbing them into the PDFs, resulting in a μ_F dependence.

The Q^2 dependence of the PDFs can be calculated directly, via a formalism developed by Dokshittzer, Gribov, Lipatov, Altarelli and Parisi [9][10][11], collectively referred to as *DGLAP*. The DGLAP Q^2 evolution equations incorporate QCD radiative corrections in

[4]The factorisation theorem has only been strictly proven for DIS; however, in practice it is applied to a broad range of processes.

terms of *splitting functions*, denoting the probability of each type of radiative correction. These splittings are visualised in Figure 1.3 and can be described as:

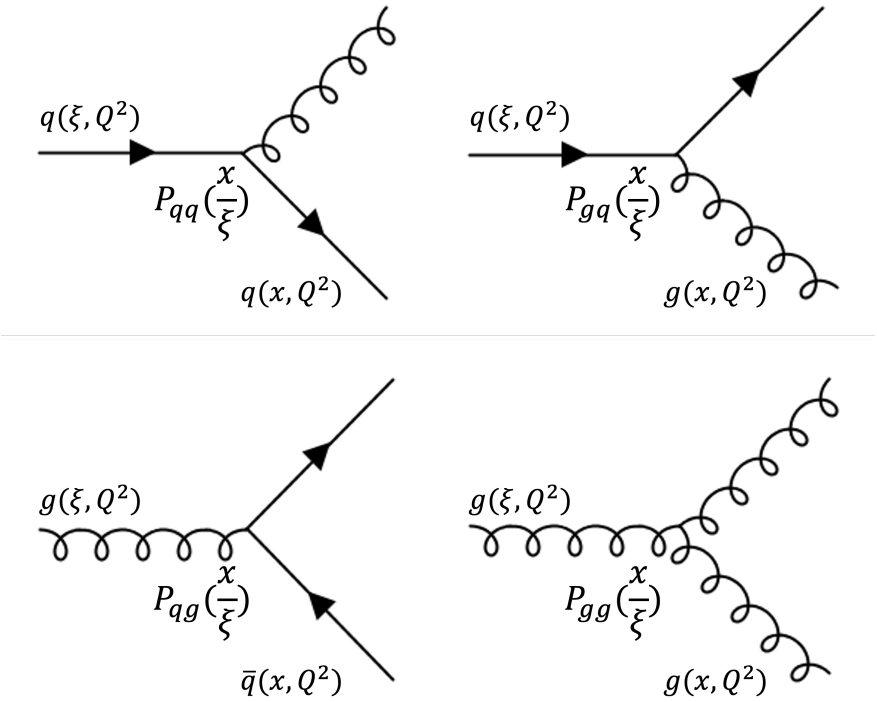


Figure 1.3: A visualisation of the four DGLAP splitting kernels.

- P_{qq}: The probability of $q \to qg$ gluon Bremsstrahlung from a quark;
- P_{gq}: The probability of $q \to gq$ gluon production from a quark;
- P_{gg}: The probability of $g \to q\bar{q}$ quark pair production from gluon splitting;
- P_{gg}: The probability of $g \to gg$ gluon pair production from gluon splitting.

At Leading Order (LO) in α_{S} , the splitting functions have the form:

$$
\begin{aligned}
P_{qq}(z) &= \frac{4}{3}\left[\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z)\right] \\
P_{qg}(z) &= \frac{1}{2}[z^2 + (1-z)^2], \\
P_{gq}(z) &= \frac{4}{3}\left[\frac{1+1-z^2}{z}\right], \\
P_{gg}(z) &= 6\left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z)\right] + \frac{33-2N_f}{6}\delta(1-z),
\end{aligned}
$$

where $z = x/\xi$ is the ratio of the momentum of the emitted parton to the original parton, and N_f is the number of quark flavours. The δ terms are included to correct for singularities as $z \to 1$ associated with soft gluon emission. The subscript $_+$ implies the following treatment when the splitting functions are integrated over z:

$$
\int_0^1 dz g(z)_+ f(x) = \int_0^1 dz g(z)(f(x) - f(1)).
$$

The Q^2 evolution of the quark and gluon PDFs, denoted q_i and g respectively can then be expressed in terms of the DGLAP splitting functions as:

$$
\frac{\partial}{\partial \ln Q^2}\begin{pmatrix} q_i(x,Q^2) \\ g(x,Q^2) \end{pmatrix} = \frac{\alpha_{\mathrm{S}}(Q^2)}{2\pi}\sum_{j\in q,g}\frac{d\xi}{\xi}\begin{pmatrix} P_{q_i,q_j}(\frac{x}{\xi},\alpha_{\mathrm{S}}(Q^2)) & P_{q_i,g}(\frac{x}{\xi},\alpha_{\mathrm{S}}(Q^2)) \\ P_{g,q_i}(\frac{x}{\xi},\alpha_{\mathrm{S}}(Q^2)) & P_{gg}(\frac{x}{\xi},\alpha_{\mathrm{S}}(Q^2)) \end{pmatrix}\begin{pmatrix} q_j(\xi,Q^2) \\ g(\xi,Q^2) \end{pmatrix}.
$$

The result of the above discussion is that, despite the PDFs not being calculable using perturbation theory, their assumed universality and known Q^2 evolution allow them to be extracted from fits to experimental data, often from several processes simultaneously to

increase the fit sensitivity to x, Q^2 and type of parton. The PDFs are parameterised at some initial scale, Q_0^2, evolved using the DGLAP equations to the scale of PDF-sensitive experimental data, and, using the factorisation theorem, convoluted with perturbatively calculated parton cross-sections to produce particle-level predictions. These predictions can then be used in a χ^2 fit to the data to extract the PDFs. Further discussion of PDF fits to experimental data can be found in Chapter 10.

1.4 The Weak Force

The weak force is responsible for interactions which transform one type of fermion into another, and is the only force experienced by all Standard Model fermions. One prominent example is the process of nuclear β-decay, where a d-quark in a neutron converts into a u-quark of a proton via the emission of an electron and an electron anti-neutrino:
$d \rightarrow u + e + \bar{\nu}_e$.

The weak force couples to, and conserves, a quantum number called *weak isospin*, I. For left-handed [5] fermions, $\mathbf{I} = \frac{1}{2}$ and has a third component $I_3 = \pm\frac{1}{2}$; 'up-type' quarks and charged leptons have $I_3 = +\frac{1}{2}$ while 'down-type' quarks and neutrinos have $I_3 = -\frac{1}{2}$. For right-handed anti-fermions, $\mathbf{I} = \frac{1}{2}$ and the sign of I_3 is reversed. Therefore, for left-handed fermions and right-handed anti-fermions, the weak force acts on *weak isospin* doublets of the fermion field spinors,

[5] Left-handed particles have their spin pointed in the same direction as their momentum, right-handed particle shave their spin pointed in the opposite direction to their momentum.

$$\chi_q = \begin{pmatrix} \psi_u \\ \psi_d \end{pmatrix}_L, \quad \chi_l = \begin{pmatrix} \psi_l \\ \psi_{\nu_l} \end{pmatrix}_L.$$

where the subscript L implies the left-handed component of the spinor field only. For right-handed fermions and left-handed anti-fermions, $\mathbf{I} = I_3 = 0$, indicating that the weak force couples only to left-handed fermions and right-handed anti-fermions.

The weak isospin doublet results in an SU(2)$_L$ gauge symmetry for the weak force. To ensure the gauge invariance of the Dirac Lagrangian, the doublets must transform under SU(2)$_L$ according to:

$$\begin{aligned} \chi &\xrightarrow{\mathrm{SU(2)_L}} \chi' = \exp[ig_{\mathrm{W}}\boldsymbol{\alpha}(x)\cdot\hat{\mathbf{T}}]\chi(x), \\ \partial_\mu &\xrightarrow{\mathrm{SU(2)_L}} D_\mu = \partial_\mu + ig_{\mathrm{W}}\hat{\mathbf{T}}\cdot\mathbf{W}_\mu \end{aligned}$$

where $\boldsymbol{\alpha}(x)$ is a set of three functions of spacetime. $\hat{\mathbf{T}} = \{\mathrm{T}^1, \mathrm{T}^2, \mathrm{T}^3\}$ is the generator of the SU(2)$_L$ group, which depends on the Pauli spin matrices $\boldsymbol{\sigma}$ according to $\hat{\mathbf{T}} = \frac{\boldsymbol{\sigma}}{2}$. and obeys the commutation rule $[\mathrm{T_i}, \mathrm{T_j}] = \mathrm{i}\epsilon_{\mathrm{ijk}}\mathrm{T_k}$ where ϵ_{ijk} is the totally antisymmetric Levi-Cevita tensor. The requirement that the weak Lagrangian be invariant under SU(2)$_L$ transformation introduces three new weak gauge fields W_μ^k, which must transform as:

$$W_\mu^k \xrightarrow{\mathrm{SU(2)_L}} W_\mu^{k'} = W_\mu^k - \partial_\mu\alpha^k - g_{\mathrm{W}}\epsilon_{ijk}\alpha^i W_\mu^j. \tag{1.10}$$

The final term in Eq. 1.10 is a result of the non-abelian nature of the SU(2)$_L$ group, which does not allow terms containing $(\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})(\mathbf{W}\cdot\boldsymbol{\sigma})$ and $(\mathbf{W}\cdot\boldsymbol{\sigma})(\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})$ to cancel. A gauge invariant kinetic term for the W_μ^k gauge fields must also be added to the weak Lagrangian;

this can be expressed in the usual $-\frac{1}{4}\mathbf{W}^{\mu\nu}\cdot\mathbf{W}_{\mu\nu}$ form, if $\mathbf{W}^{\mu\nu}$ is defined as:

$$\mathbf{W}^{\mu\nu} = \partial^{\mu}\mathbf{W}^{\nu} - \partial^{\nu}\mathbf{W}^{\mu} - g_{\mathrm{W}}\mathbf{W}^{\mu}\times\mathbf{W}^{\nu}, \tag{1.11}$$

with the final cross-product term in Eq. 1.11 arising due to the non-abelian nature of SU(2)$_L$ and indicating that the gauge fields associated with the weak force have the ability to self-interact; when Eq. 1.11 is expanded it includes terms for both triple and quartic couplings of the weak gauge bosons.

If the fermion masses are assumed to be zero[6], the weak Lagrangian therefore can be expressed as:

$$\mathcal{L}_{\mathrm{Weak}} = i\bar{\chi}\gamma^{\mu}\partial_{\mu}\chi - \frac{1}{2}g_{\mathrm{W}}\bar{\chi}\gamma^{\mu}\sigma^{k}W^{k}_{\mu}\chi - \frac{1}{4}W^{\mu\nu,k}W^{k}_{\mu\nu}. \tag{1.12}$$

The conserved weak currents $j^{k}_{\mu} = -\frac{g_{\mathrm{W}}}{2}\bar{\chi}\gamma^{\mu}\sigma^{k}\chi$ can be extracted directly from the χ, W^{k}_{μ} interaction term in Eq. 1.12. The first two of these currents can be expressed in terms of the weak isospin raising and lowering operators $\sigma^{\pm}_{\mu} = \frac{1}{2}(\sigma^{1}\pm i\sigma^{2})$ as:

$$j^{\pm}_{\mu} = -\frac{g_{\mathrm{W}}}{\sqrt{2}}\bar{\chi}\gamma^{\mu}\sigma^{\pm}\chi,$$

which couple to the physical charged weak bosons, which can be expressed as linear combinations of the first two W^{k}_{μ} fields,

$$W^{\pm}_{\mu} = \frac{1}{\sqrt{2}}(W^{1}_{\mu}\pm iW^{2}_{\mu}).$$

[6]Masses for weakly-interacting fermions will be reintroduced via the Higgs mechanism discussed in Section 1.4.3.

The weak force is the only fundamental interaction which violates P parity, the sign inversion of positional coordinates through the origin. This feature was observed in nuclear β-decay by Wu et al. in 1957 [12] and must be included in the model of the weak force.

All currents resulting from SM interactions can be classified as one of the set of Lorentz invariant $\in \{$scalar, pseudoscalar, vector, axial vector, tensor$\}$, based on the Dirac matrix structure of the interaction vertex and the gauge boson spin, and are intrinsically linked to the parity properties of the boson. The scalar products of two vector or two axial vector currents are invariant under a parity transformation; however, the scalar product of a vector current and an axial vector current is not. Therefore, to allow for parity violation the weak current must be a linear combination of vector and axial vector currents. Furthermore, it has been shown experimentally that the weak current is a $(V - A)$ interaction, and that it is maximally parity violating. This implies that the associated gauge bosons couple equally to both the vector, g_{V} and axial vector, g_{A}, components of the weak current: $g_{\mathrm{V}} = g_{\mathrm{A}} = g_{\mathrm{W}}$.

This can be included in the model of the weak interaction by making the substitution

$$\gamma^{\mu} \xrightarrow[\text{violation}]{\text{parity}} \gamma^{\mu}(1 - \gamma^{5}),$$

where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, as $\bar{\psi}\gamma^{\mu}\psi$ transforms as a vector under P parity, while $\bar{\psi}\gamma^{\mu}\gamma^{5}\psi$ transforms as an axial vector. This substitution results in new expressions for the currents associated with the physical $W^{\pm}$ bosons:

$$j_{\mu}^{\pm} = -\frac{g_{\mathrm{W}}}{\sqrt{2}}\bar{\chi}\gamma^{\mu}(1-\gamma^{5})\sigma^{\pm}\chi, \tag{1.13}$$

and for the weak neutral current

$$j_{\mu}^{3} = -\frac{g_{\mathrm{W}}}{2}\bar{\chi}\gamma^{\mu}(1-\gamma^{5})\sigma^{3}\chi. \tag{1.14}$$

It can be observed from Eq. 1.13 and Eq. 1.14 the weak charged current changes fermion flavour due to its dependence on the off-diagonal Pauli matrices σ_1, σ_2, while the weak neutral current does not, as it depends only on σ_3. It can also be noted that Eq. 1.13 and Eq. 1.14 also contain a dependence on the left-handed projection operator

$$P_{\mathrm{L}} = \frac{1}{2}(1-\gamma^{5}),$$

which, when applied to a fermion spinor field, returns only the left-handed components of fermions and right-handed components of anti-fermions. This recovers the chiral structure of the weak interaction discussed previously in this section.

1.4.1 Quark Mixing

The final element required to complete the description of the weak force in the Standard Model is a description of quark mixing via the weak current. This concept was originally introduced via the Cabibbo mechanism [13], which was proposed before the discovery of c, t and b quarks, and explains the difference observed in the probability for d and s quarks to decay into u quarks. In this model, each 'weak' quark eigenstate, the object

which couples to the weak interaction, is comprised of an unequal superposition of the physical quark 'mass eigenstates', related by a 2×2 unitary matrix, with the rotation angle known as the Cabibbo angle, θ_C:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}. \tag{1.15}$$

This idea allowed Glashow, Iliopoulos and Maini to propose the GIM mechanism [14] as an explanation for the lack of a flavour changing neutral-current (FCNC) in 1970, prior to the discovery of the charm quark. In this model, decay modes in which the FCNC is allowed to proceed via the exchange of a virtual u quark are strongly suppressed by destructive interference with other modes which occur via the exchange of a c quark.

The Cabibbo mechanism is extended to a three quark generation Standard Model via the 3×3 unitary *CKM matrix* [15], relating the weak eigenstates to the mass eigenstates via:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{1.16}$$

which introduces an additional factor of V_{ij} to weak charged current interaction vertices between quark flavour i and j (or V_{ij}^* for an interaction between i and $\bar{j}$). The CKM matrix can be described using three rotation angles and a complex phase, δ, such that:

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta'} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta'} & 0 & c_{13} \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (1.17)$$

where $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$. The complex phase in the CKM matrix is a natural mechanism for the introduction of CP violation into the Standard Model via the weak interaction.

1.4.2 Electroweak Unification

From the discussion in Section 1.4, it can be observed that the neutral weak field W^3, which does not change fermion flavour, only couples left-handed fermions or right-handed anti fermions. However, the physical Z boson has been experimentally observed to couple to both left-handed and right-handed fermions, albeit not equally. This apparent discrepancy is resolved in the Glashow, Salam and Weinberg (GSW) model of *electroweak unification* [16][17][18].

In this model, the U(1) gauge symmetry from QED is replaced with a U(1)$_Y$ symmetry which couples to a quantum number called *weak hypercharge* Y. This gives rise to a new gauge field B_μ which couples to both left and right handed particles and has the interaction vertex interaction term $g'\frac{Y}{2}\bar{\psi}\gamma^\mu B_\mu\psi$, the same form as the QED photon interaction term. The GSW model postulates that the physical photon, A_μ and Z boson are formed from a mixture of the W^3_μ and B_μ fields such that:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_{\mathrm{W}} & \sin\theta_{\mathrm{W}} \\ -\sin\theta_{\mathrm{W}} & \cos\theta_{\mathrm{W}} \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix},$$

where the mixing angle θ_{W} is known as the *Weinberg angle*. By demanding the terms in the electroweak Lagrangian containing A_μ are consistent with the original QED formulation, it can be shown that the EM and weak coupling strengths and quantum numbers can be related according to:

$$\begin{aligned} e &= g_{\mathrm{W}} \sin\theta_{\mathrm{W}} = g' \cos\theta_{\mathrm{W}}, \\ Y &= 2(Q - I_3). \end{aligned}$$

The resulting symmetry of the electroweak force is SU(2)$_L$×U(1)$_Y$, with a covariant derivative of the form

$$D_\mu = \partial_\mu + ig_{\mathrm{W}}\mathbf{T}\cdot\mathbf{W}_\mu + ig'\frac{Y}{2}B_\mu. \tag{1.18}$$

1.4.3 The Higgs Mechanism

Local gauge invariance in the Standard Model is broken by boson mass terms in the Lagrangian. This is not an issue for QED and QCD, which are both mediated by gauge bosons which are massless; however, the $W^\pm$ and Z bosons which mediate the weak force have been experimentally observed to have mass. Additionally, massive fermions would violate SU(2)$_L$ gauge invariance, as left and right-handed chiral states would transform differently. The *Higgs mechanism* resolves this apparent contradiction by describing

how particles acquire mass through interactions with a *Higgs field*.

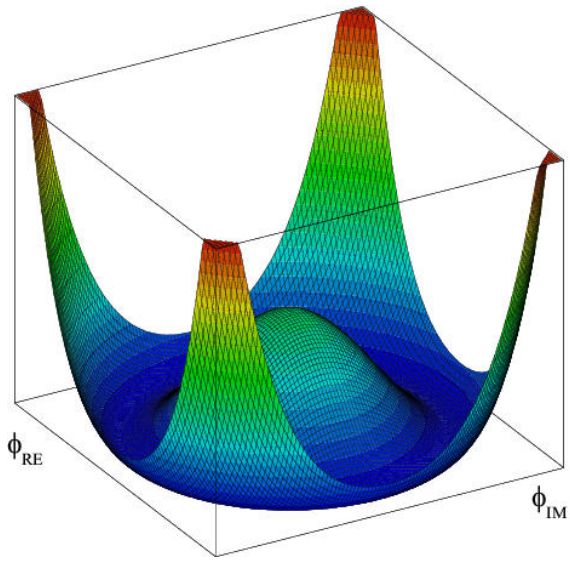


Figure 1.4: The 'Mexican hat potential'; a simple model for the shape of the Higgs field ϕ. Image from [19].

The simplest model of the Higgs field is a weak isospin doublet of complex scalar fields ϕ_i, with four degrees of freedom:

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

where ϕ^0 is neutral, and the sign of the charged field ϕ^+ can be inverted by the operation $(\phi^+)^* = \phi^-$. The Lagrangian of this doublet can be expressed as

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \tag{1.19}$$

where D_μ is the SU(2)$_L$×U(1)$_Y$ covariant derivative from Eq. 1.18. The first term in Eq. 1.19 is the kinetic term of the field, and the second term is the *Higgs potential* $V(\phi)$, taking the form

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2,$$

where μ and λ are scalar constants. The term proportional to $\phi^\dagger\phi$ is the kinetic term for a scalar particle associated with the field, while the $\mu^2\phi^\dagger\phi$ term can be considered the particle's mass term. The lowest possible energy state of ϕ corresponds to the minimum of $V(\phi)$, and is known as the *vacuum expectation value*, v. If $\mu^2 < 0$, this $V(\phi)$ has an infinite set of degenerate minima which satisfy

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{v^2}{2} = \frac{|\mu|^2}{2\lambda}, \tag{1.20}$$

which implies that the minimum of the potential is located at $\phi = v = \sqrt{\frac{|\mu|^2}{\lambda}} \neq 0$. When a choice of a particular v is made, the symmetry of ϕ is *spontaneously broken*. If a vacuum state where $\phi_1 = \phi_2 = \phi_4 = 0$, $\phi_3 = v$ is chosen, the resulting vacuum expectation of

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix},$$

can be expanded about the minimum to give

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1(x) + i\phi_2(x)\\v + \eta(x) + i\phi_4(x)\end{pmatrix}, \tag{1.21}$$

which introduces one massive scalar field $\eta(x)$ and three massive, unphysical *Goldstone fields* $\phi(x)_{1,2,4}$ to the Lagrangian. The Goldstone fields can be removed from the

Lagrangian through the choice of a particular gauge, the *unitary gauge*, which causes the Goldstone fields to be absorbed into the longitudinal polarisation components of the W and Z fields present in D_μ, while also providing them mass terms. Eq. 1.21 can be rewritten in the unitary gauge:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{1.22}$$

where $h(x)$ is the same scalar field as $\eta(x)$, rewritten to emphasise that it corresponds to the physical Higgs field. Eq. 1.22 can the be substituted into the expanded covariant derivative term of the Lagrangian in Eq. 1.19, and expressed in terms of the physical $W^\pm$ and Z bosons:

$$\begin{aligned}(D_\mu\phi)^\dagger(D_\mu\phi) = &\frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{4}g_{\mathrm{W}}^2(W_\mu^+ W^{\mu-})(v+h)^2 \\ &+ \frac{g_{\mathrm{W}}^2 + g'^2}{8}(Z_\mu Z^\mu)(v+h)^2 \\ &+ \frac{\mu^2}{2}(v+h)^2 - \frac{\lambda}{16}(v+h)^4,\end{aligned} \tag{1.23}$$

from which the mass terms of the electroweak bosons can be read off:

$$\begin{aligned} m_W &= \frac{g_{\mathrm{W}} v}{2}, \\ m_Z &= \frac{v\sqrt{g_{\mathrm{W}}^2 + g'^2}}{2} = \frac{m_{\mathrm{W}}}{\cos\theta_{\mathrm{W}}}, \\ m_\gamma &= 0. \end{aligned}$$

Additionally, Eq. 1.23 includes terms for the trilinear couplings between the elec-

troweak vector bosons and an additional scalar *Higgs boson* introduced by the Higgs field, h, self-coupling terms for the Higgs boson, and the Higgs mass term

$$m_H = \sqrt{2|\mu|^2}.$$

The spontaneous breaking of the SU(2)$_L$×U(1)$_Y$ symmetry also provides a mechanism for explaining fermion masses. This is accomplished by introducing a term coupling the fermions to the Higgs field into the Standard Model Lagrangian, and spontaneously breaking SU(2)$_L$×U(1)$_Y$ symmetry of the Higgs field as in Eq. 1.21. The resulting Lagrangian term is given as:

$$\mathcal{L}_{\text{Yukawa}} = \sum_f \left(-\frac{y_f}{\sqrt{2}} v \bar{\psi}_f \psi_f - \frac{y_f}{\sqrt{2}} h \bar{\psi}_f \psi_f \right),$$

where the first term has the correct form to represent the mass term for fermion f if the *Yukawa coupling* parameters y_f are chosen such that

$$y_f = \sqrt{2}\frac{m_f}{v}.$$

This concludes the overview of the Standard Model, a theory that describes the fundamental nature of the universe with 18 free parameters:

- 9 related to the masses of the quarks and charged leptons;
- 4 related to the CKM matrix, comprising 3 mixing angles and 1 complex phase;
- 3 couplings to the strong and electroweak forces;

- 2 related to the Higgs mass and Higgs vacuum expectation value.

The Standard Model, as described here, has been crafted, tested and updated over the course of many decades and, so far, has stood up robustly to experimentation. However, it is an incomplete theory, with several avenues for new discoveries remaining, including:

- Dark matter and/or supersymmetry;
- The hierarchy problem;
- CP violation in the strong sector and the matter/antimatter asymmetry of the universe;
- Neutrino mixing, masses and Dirac vs Majorana nature;
- The incorporation of a description of gravity into the model.

1.5 W boson production in association with jets

A sufficiently robust foundation of the Standard Model has now been established such that the production of W bosons in association with inclusive jets, hereafter referred to as W+jets can now be discussed. W+jets provides challenging conditions for QCD models, and therefore is an important benchmark process; the W boson production requires knowledge of the electroweak sector, while the associated jet can provide a probe of calculations' QCD modelling.

At LHC energies, W+jets differential cross-sections can cover several orders of magnitude and can reach up to several TeV in jet energy. Additionally, as shown in Figure 1.5, W+jets has the second-highest production cross-section at the LHC after inelastic pp scattering, meaning the large statistics can extend the sensitivity of differential W+jets cross-sections up to large jet energies and rapidities. W+jets production data can also provide powerful constraints for measuring PDFs. Furthermore, due to this large cross-section and the presence of leptons, jets and missing transverse energy from the neutrino in the final-state, W+jets is an important background for Higgs measurements and searches for physics beyond the Standard Model.

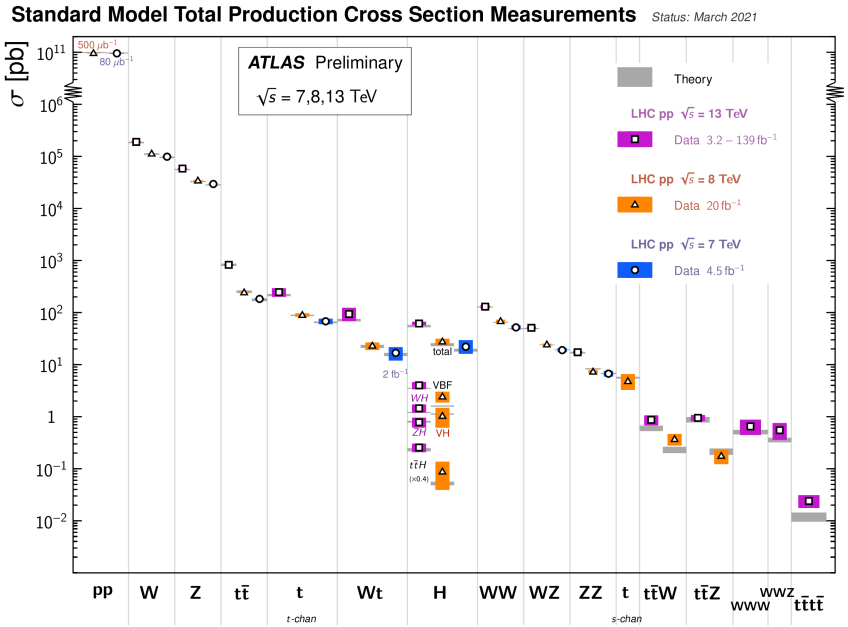


Figure 1.5: Summary of production cross-sections at the LHC. Figure from [20].

At the lowest order, W+jets is produced via quark-gluon fusion or initial-state gluon radiation, as illustrated in Figure 1.6, both of which provide direct sensitivity to the gluon. In the analysis which will be presented in Chapters 5-7, W+jets production cross-sections

will be measured differentially in variables sensitive to the higher-order corrections of several state-of-the-art predictions, to which the measured cross-sections are compared, to benchmark advancements in fixed-order QCD calculations and matching and merging schemes.

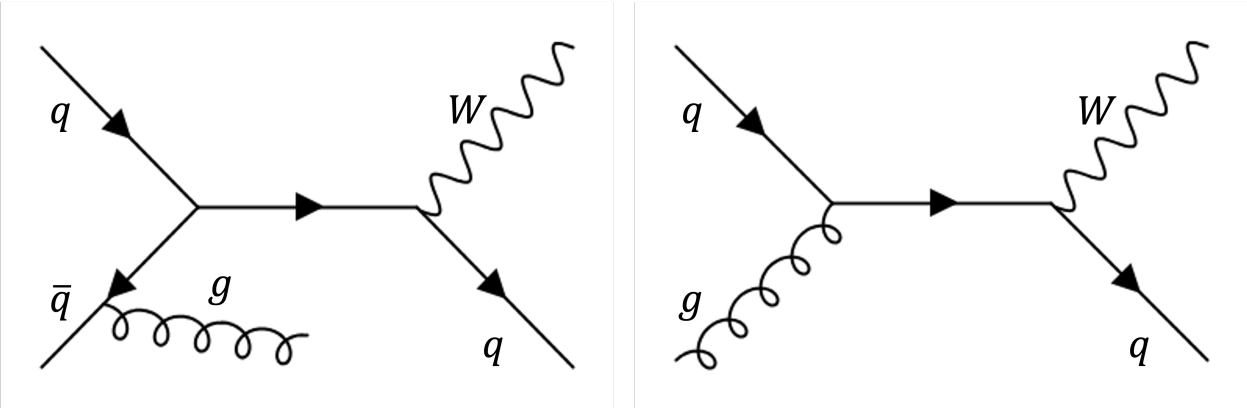


Figure 1.6: Feynman diagrams of two W+jets production modes at lowest order in QCD; Initial-state radiation (left) and quark-gluon fusion(right).

Despite the larger branching fraction for $W \rightarrow qq'$, the measurement presented in this thesis is made in the $W \rightarrow l\nu$ decay channel as this provides the cleanest detector signature, and minimises the contribution from systematic uncertainties related to jet reconstruction. The ratio of the charge-dependent W^+/W^- differential cross-sections will also be presented; this ratio measurement allows many systematic uncertainties to cancel, increasing the precision of the test of the models.

Since gluons couple equally to u and d quarks, the charge-dependent cross-section measurements also provide sensitivity to the PDFs of valence quarks in the range $10^{-3} \leq x \lesssim 0.3$ [21] via the quark-gluon fusion production mechanism. The initial-state radiation production mode also provides PDF-constraining information for the sea quarks; the transition from $\bar{u} \rightarrow d$ has a different cross-section to $\bar{u} \rightarrow s$ since the CKM matrix elements

$V_{ud}^* \neq V_{us}^*$, hence, the relative contribution of the different down-type sea quarks to the proton is accessible via the W+jets production cross-section.

Chapter 2

The LHC and the ATLAS detector

"One ring to rule them all, one ring to find them..."

— J.R.R Tolkien, 'Lord of the Rings'

The goal of experimental high energy physics is to test fundamental theories of subatomic matter, and its interactions, through scattering experiments in high-energy regimes. There are two phases involved in performing such experiments - each interaction has parent particles which must be prepared by an accelerator, and products which are observed by detectors. This thesis will focus on data from the Large Hadron Collider (described in Section 2.1), where collisions are prepared by accelerating protons and heavy ions to very high energies. This increases the energy available for the formation of the collision products and allows smaller length scales to be probed, as observed by the ATLAS detector, described in Section 2.2.

2.1 The Large Hadron Collider

Figure 2.1: An aerial image of the area covered by the 27 km long LHC ring, with the outline of the underground accelerator superimposed as a yellow line [22].

The Large Hadron Collider (LHC) [23] is the world's largest and highest energy particle accelerator. It takes the form of a 27 km long synchrotron ring under the French-Swiss border approximately 100 m underground, occupying the tunnel formerly used by the Large Electron Positron (LEP) collider. The LHC primarily accelerates two beams of protons (although it also has the capability to accelerate heavy ions) to several TeV before colliding them at four interaction points, where the four main LHC experiments are located, to study the products of the collisions and their subsequent decays.

The LHC acceleration chain is shown in Figure 2.2. The LHC's protons are sourced from a hydrogen bottle, where the hydrogen atoms are ionised to H^- ions in an electric field. Since 2020, the LINAC 4 (previously the LINAC 2) is used to accelerate the H^- ions to 160 MeV. The ions are then injected into the Proton Synchrotron Booster (PSB,

referred to simply as ‘Booster’ in Figure 2.2) where they are stripped of two electrons to create protons. The PSB is composed of four synchrotron rings where the protons are boosted to 1.4 GeV and gathered together into ‘bunches’.

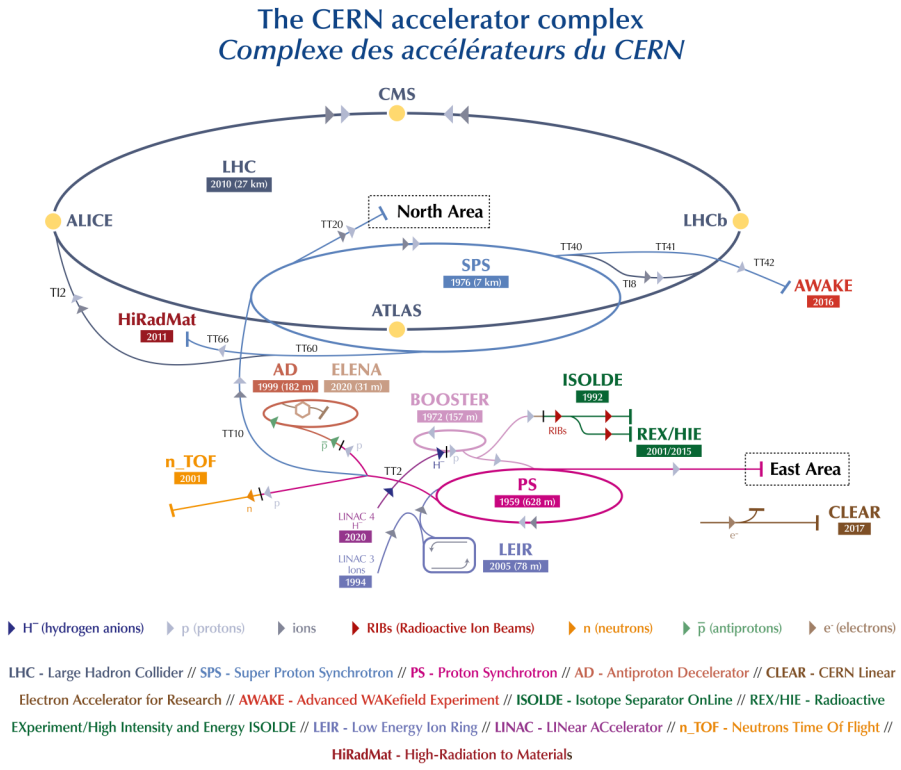


Figure 2.2: The CERN accelerator complex feeding the Large Hadron Collider (dark blue line). Graphic courtesy of [24].

For heavy ion collisions, the ions originate from a vapourised lead source, before being accelerated by the LINAC 3 and boosted by the Low Energy Ion Ring (LEIR) and subsequently following the same path as protons to the LHC.

The next stage in the acceleration chain is the Proton Synchrotron (PS), which increases proton energy to 25 GeV. Protons then continue to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Finally, the protons or heavy ions are injected into the LHC synchrotron rings’ beam pipes, where one beam circulates clockwise

and the other circulates anticlockwise. As of the beginning of LHC Run 3 in 2022, maximum beam energy for protons is 6.8 TeV in each beam for a total centre-of-mass energy of 13.6 TeV.

The LHC uses radio frequency cavities to boost the protons/ heavy ion beams, which are bent in a circle using 8.3 T superconducting dipole magnets and are focused using quadrupole magnets. The proton beams are composed of 2544 bunches of 10^{11} protons each, separated in time by 25 ns and circulate in groups called bunch trains. After circulating for up to several hours, the bunches are brought together to collide at one of four interaction points on the LHC ring, where the four main LHC experiments are located:

- ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment) [25] focuses on heavy-ion collisions to study a phase of matter called the quark-gluon plasma, which provides information on the strong force in environments of extreme energy densities with many decay products from the heavy-ions collisions.
- LHCb (**L**arge **H**adron **C**ollider **B**eauty) [26] specialises in the study of the differences between matter and antimatter, focusing particularly on hadrons composed of heavy-flavour quarks. These hadrons are produced mainly in the forward region, leading to the distinctive asymmetric shape of the LHCb detector.
- The CMS (**C**ompact **M**uon **S**olenoid) [27] and ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) [28] experiments are general purpose high energy physics detectors.

As this document focuses on data collected by the ATLAS experiment, this detector will be described in Section 2.2.

2.1.1 Cross-sections, luminosity and pile-up

A typical observable for measurements in high energy physics is the cross-section, a measure of probability of a particular interaction occurring. This is generally denoted by σ and measured in units of area, usually the barn b = 100 fm^2.

Another important quantity in high energy physics is luminosity. Instantaneous luminosity is a measure of real-time beam intensity, and for two Gaussian beams colliding head-on can be calculated from [29]:

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y},$$

where $N_{1,2}$ are the number of protons in bunch 1 and bunch 2 respectively, f is the revolution frequency, N_b is the number of bunches and $\sigma_{x,y}$ are the beam widths in the x and y directions respectively. In practice, it is more typical to refer to the time-integrated luminosity $L = \int \mathcal{L} dt$. This 'integrated luminosity' is treated as a proxy for the amount of data collected as the number of recorded events and can be expressed as:

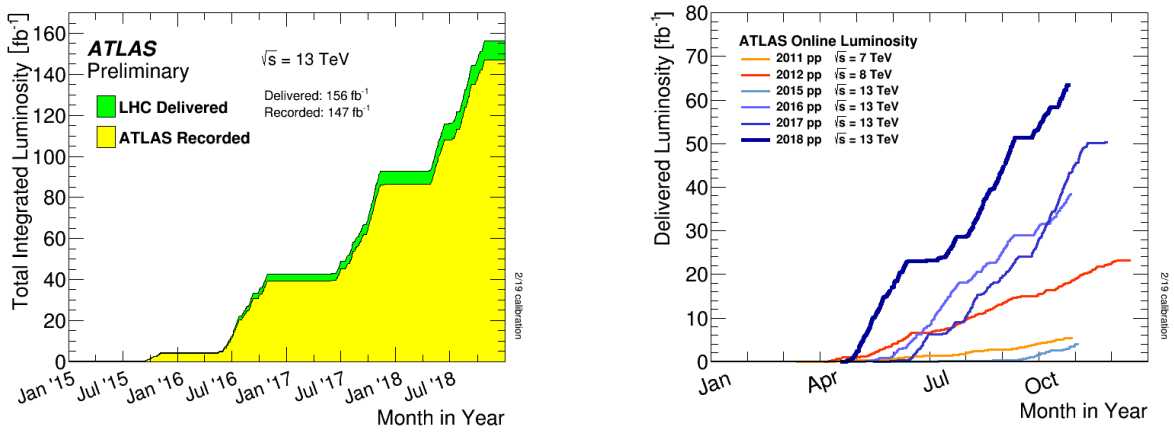
$$N = \sigma L. \tag{2.1}$$

Higher luminosity conditions allow lower cross-section processes to be probed by experiments, however, this comes at the expense of large amounts of pile-up. Pile-up is additional pp interactions or soft QCD activity other than the process of interest, which can affect the reconstruction of event kinematics and create additional backgrounds. A more in-depth discussion of pile-up sources, classification and strategies for mitigating

its effects is in Chapter 4. Pile-up is generally quantified by $\langle\mu\rangle$, the average number of interactions per bunch crossing.

An additional key quantity in high energy physics experiments is the centre-of-mass energy, $\sqrt{s}$, which is the energy of the system of two colliding particle in the centre-of-mass frame and corresponds to the maximum energy available for the products created by the collision. This is calculated as $\sqrt{s} = \sqrt{E_1^2 - |p_1^2|} + \sqrt{E_2^2 - |p_2^2|}$ in the centre-of-mass frame. During Run 1, the LHC ran at a $\sqrt{s}$ of 7 TeV and 8 TeV. Following an upgrade to the LHC magnets, this was raised to 13 TeV for the Run 2 and subsequently 13.6 TeV for Run 3. The High-Luminosity LHC upgrade is expected to achieve a centre-of-mass energy of 14 TeV.

This thesis will focus on pp collisions at a centre-of-mass energy of 13 TeV taken by ATLAS during the first two years of LHC Run 2, which took place from 2015-2018 and collected an integrated luminosity of 140.1 $\pm$ 1.2 (i.e. an uncertainty of 0.83%) fb^{-1} [30]. The integrated luminosity delivered by the LHC during Run 2 compared to that collected by ATLAS is shown in Figure 2.3a, and the former is broken down by year for both Run 1 and Run 2 in Figure 2.3b. The 2015 + 2016 dataset analysed in this thesis has an integrated luminosity of 36.2 fb^{-1} .



(a) (b)

Figure 2.3: (a) The integrated luminosity delivered to ATLAS' Interaction Point 1 during LHC Run 2 (green) and the integrated luminosity recorded by ATLAS (yellow as a function of time) [31]. (b) The integrated luminosity delivered to ATLAS during LHC Run 1 and Run 2, broken down per year [32].

2.2 The ATLAS experiment

A Toroidal **L**HC **A**pparatu**S** (ATLAS) [28] [32] is a general-purpose particle physics experiment located at the LHC Interaction Point (IP) 1. It is designed to carry out a broad physics programme. The ATLAS detector is cylindrical in shape and has almost 4π solid angle coverage around the collision point, with the aim of reconstructing every particle produced in an event so that the Missing Transverse Energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) from non-interacting particles can be calculated. The detector, of dimensions 46 m in length and 25 m in diameter and weighing approximately 7000 t, is located in a cavern at IP 1 approximately 90 m underground. A cut-away diagram of the ATLAS detector as of the beginning of LHC Run 3 is shown in Figure 2.4.

The design of ATLAS was influenced by the physics goals at the beginning of LHC operation, specifically Higgs, Dark Matter and Supersymmetry searches. To be able to dis-

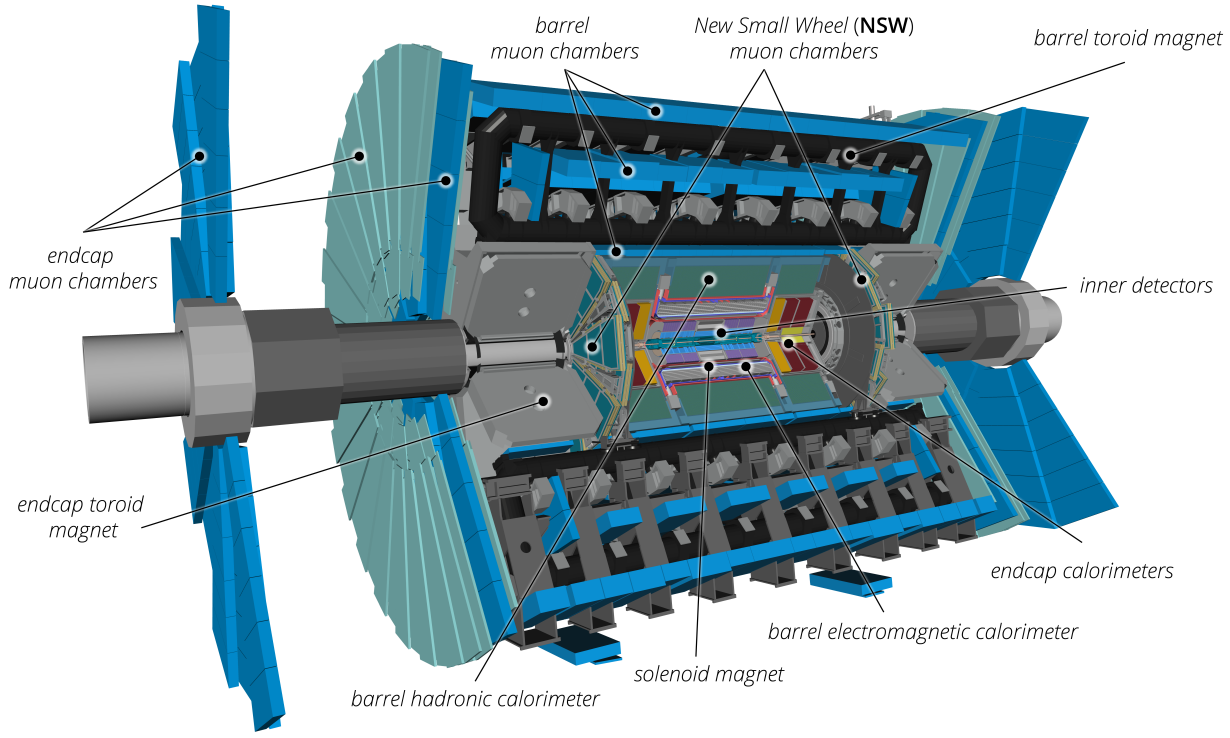


Figure 2.4: Cut-away diagram of the ATLAS detector configuration as of the start of LHC Run 3, with labels indicating the major subsystems [32].

tinguish these low cross-section processes from much higher cross-section backgrounds produced by the LHC, the ATLAS experiment was designed with excellent particle identification, in addition to displaced vertex, and missing transverse energy reconstruction, as key criteria. Additional key design considerations were:

- High detector granularity
- Fast and radiation hard electronics and sensors
- Excellent tracking and calorimetry
- Efficient triggering, including at low transverse momenta, with sufficient background rejection such that an acceptable trigger rate can be maintained.

The overall ATLAS detector is made up of nested individual subsystems, each with a complimentary role in particle identification and reconstruction. From the centre outwards, these are; a tracking system called the Inner-Detector (ID) [33] [34], electromagnetic sampling calorimeters (ECAL) [35], hadronic sampling calorimeters (HCAL) [35] [36] and a muon spectrometer (MS) [28]. This combination of subsystems was selected so all types of stable products of a collision can be unambiguously reconstructed, while particles unstable on the scale of the detector can be reconstructed based on their own decay products. All components in these subsystems must be able to perform robustly and reliably in the high radiation environment at Point 1 and be sufficiently fast to contend with 40 MHz LHC bunch crossing frequencies.

The ATLAS experiment employs a right-handed coordinate system with an origin at the interaction point in the detector's centre and a z-axis running along the beam pipe. The x-axis joins the IP to the centre of the LHC ring, while the y-axis points directly upwards towards the surface. In the transverse plane, defined by the azimuthal angle around the z-axis, the cylindrical coordinates r, ϕ are commonly used. The pseudorapidity (η) can be defined in terms of the polar angle θ in the x, z plane, where $\eta = \ln\tan(\frac{\theta}{2})$. The ATLAS detector has full coverage in ϕ and coverage in η up to $|\eta| < 4.9$. The rapidity can be expressed as $y = \frac{1}{2}\ln(\frac{E+p_z}{E-pz})$ where E is energy and p_z is the momentum component in the z direction. Since the longitudinal boost of particles is not generally known, as partons within the proton carry an unknown fraction of the proton's momentum, it is typically advantageous to use y or η in place of θ, as differences in y, Δy, are Lorentz invariant and, for massless particles, $y = \eta$. Angular distances are measured in

terms of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. As the initial momentum along the z-axis is not known, it is standard practice to use transverse momentum $p_{\mathrm{T}} = p\sin\theta$ and transverse energy $E_{\mathrm{T}} = E\sin\theta$.

The ATLAS detector is supported by a hybrid system of four large superconducting magnets [37] 22 m in diameter and 26 m in length which provide a magnetic field for momentum measurements in the ID and MS. The magnets are aligned such that there is no field in the direction of the beam. The system is made up of both solenoid and toroid magnets. The solenoid [38] [39] magnet produces a 2 T axial field aligned with the beam for the ID and is designed with a minimised radiative thickness so as not to affect the kinematics of particles as they pass through via electromagnetic interactions. In addition to the solenoid, there are three toroid magnets for the MS – one for the barrel [40] and two for the endcaps [41] with 0.5 T and 1.0 T magnetic fields respectively. All these magnets are supported by a cryostat system to maintain the 4.6 K temperature required for cooling.

The ATLAS detector is accompanied by three additional detectors designed to provide coverage in the forward region. In increasing order of distance from the IP these are:

- The **LUminosity measurement using Cerenkov Integrating Detector** (LUCID) [42], replaced by the upgraded LUCID-2 [43] during the LHC Long Shutdown 1, detects inelastic pp scattering in the forward region and is located $\pm$ 17m from the interaction point. LUCID is the main relative luminosity monitor for ATLAS, and is the only ATLAS subdetector for online luminosity and beam condition monitoring. LUCID is also capable of measuring both integrated and instantaneous luminosity.

- The **Zero Degree Calorimeter** (ZDC) [44] is located $\pm$ 140 m from the IP and is embedded in the Target Absorber Neutral just beyond where the LHC beam pipe diverges into two separate pipes. Its purpose is to detect forward neutrons from heavy ion collisions.

- The **Absolute Luminosity for ATLAS** (ALFA) [42] detector is located $\pm$ 240 m from the IP. ALFA measures the absolute luminosity delivered to ATLAS via the small angle elastic scattering cross-section, which can be connected to the total cross-section and therefore luminosity via the Optical Theorem. It is composed of scintillating fibre trackers inside Roman pots - a type a detector where the volume (pot) is separated from the accelerator vacuum by a thin window and can be moved very close to the beam. Particles detected by ALFA have scattering angles which are smaller than the beam divergence (β^*), necessitating special high β^* runs with reduced beam emittance. These special runs can also be used for LUCID calibration.

2.2.1 Inner-Detector

The ATLAS ID [33] [34] provides high-precision measurements of the trajectory and momentum of charged particles produced at the ATLAS Interaction Point. These measurements are performed based on the reconstruction of tracks of connected hits in the detector elements. The ID is the ATLAS subsystem located closest to the beam pipe and is made up of three independent but complementary subdetectors; Pixel [45], Semi-Conductor Tracker [46] and the Transition Radiation Tracker [47]. A cut-away diagram

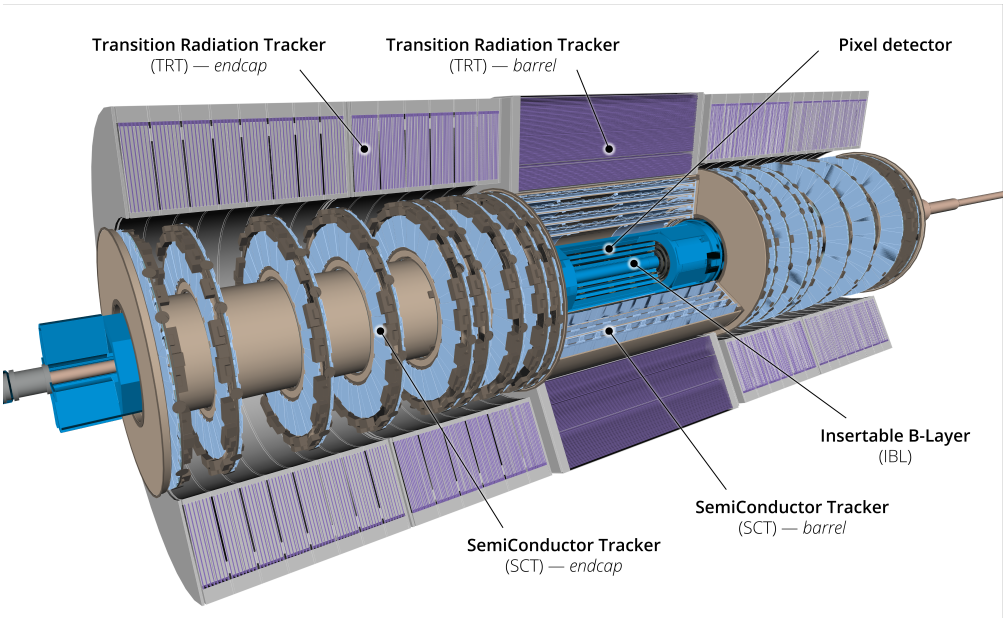


Figure 2.5: Cut-away diagram of the ATLAS Inner Detector configuration as of the start of LHC Run 3, with labels indicating the major subsystems [32].

of the ATLAS ID as of the beginning of LHC Run 3 is shown in Figure 2.5.

The ID is designed to provide excellent momentum resolution, track reconstruction efficiency and primary and secondary vertex reconstruction efficiency. To achieve this, it must have a high granularity, particularly close to the bean pipe. Due to this proximity to the beam pipe, its components must be able to tolerate a high track density and a radiation hard environment. Finally, it must be constructed from the minimum amount of material possible, to prevent multiple scattering.

The ID is served by a 2 T solenoid magnet [38] [39], which provides the magnetic field required to measure the charged particles' momentum, calculated from the radius of curvature of its path r:

$$r = p/Bq, \tag{2.2}$$

where p is the particle's momentum, B is the magnitude of the magnetic field and q is the particle's charge.

The ID has hermetic coverage in ϕ and acceptance in pseudorapidity up to $|\eta| < 2.5$. It is cylindrical in shape: 3512 mm in length and has a radius of 1150 mm. The Pixel and SCT are cooled to between -5 and -10 °C to maintain an acceptable level of noise in the radiation hard environment close to the beam pipe while the TRT is at room temperature. Combined, the ID subsystems have a total transverse momentum resolution [28] of:

$$\frac{\sigma_{p\mathrm{T}}}{p_{\mathrm{T}}} = 0.036\%\, p_{\mathrm{T}} \oplus 1.3\%.$$

The ID primarily serves to reconstruct tracks, which can be used as analysis objects themselves, but also are vital for electron, jet and muon reconstruction. Tracks are reconstructed [48] in ATLAS by performing a fit to individual hits in elements of the ID, tracing hits together into tracks. Typically at least three hits are required to make a track. For further discussion of track reconstruction in ATLAS, see Section 3.1.1.

The individual subsystems making up the ID are illustrated in detail in Figure 2.6 and will be discussed further in the subsections below.

Insertable B-Layer & Pixel

The Insertable B-Layer (IBL) [49], is the ID subdetector closest to the beamline with the highest granularity. It is comprised of 8 million reverse biased silicon diode sensors, with a 50 x 250 μm^2 pitch. In this context, 'pitch' refers to the distance between neighbouring

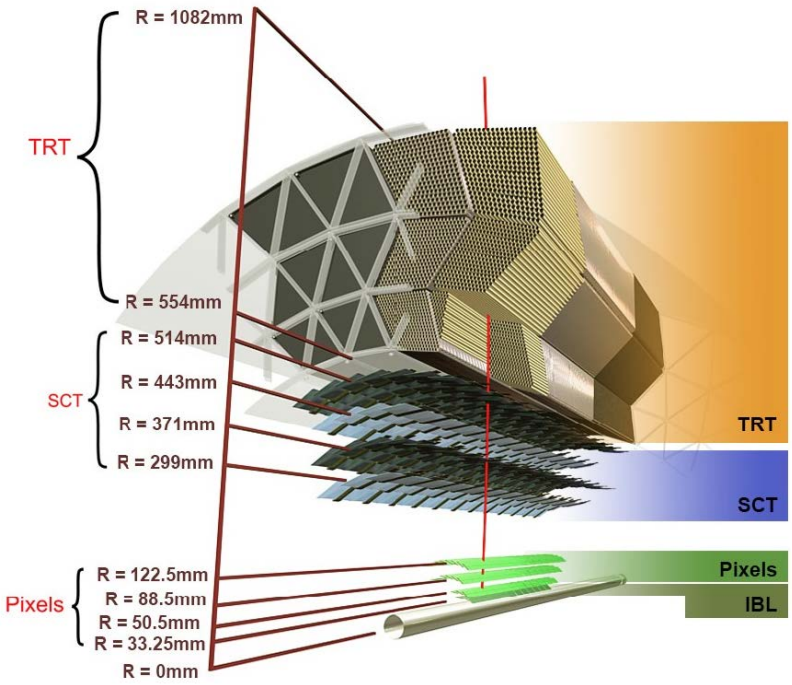


Figure 2.6: A detailed cross-section of the ATLAS Inner Detector configuration as of the start of LHC Run 3, with the major subsystems colour-coded and distance measurement indicated [32].

sensors. Choice of pitch necessitates a trade-off between tracking resolution and the number of readout channels required, which impacts factors such as detector noise and complexity. When a charged particle passes through the diode, an electron-hole pair is created, which drift towards the surface due to a potential difference placed across the pixel, creating electrical current. At the surface the current is collected, and a hit/ no hit decision is read out.

The IBL was added for LHC Run 2 to provide an removable extra layer of pixels closest to the beamline. The purpose of this upgrade was primarily for better primary and secondary vertexing for heavy flavour tagging, leading to the name ‘B-Layer’. Additional benefits of the IBL include higher small radius resolution, and improved tracking performance. The IBL is located 33 mm from the beam pipe.

Moving outwards from the beamline, the next ID subdetector is the silicon Pixel detector [45], which is comprised of three barrel layers and three endcap layers and also plays a crucial role in reconstructing both primary and secondary interaction vertices. It is composed of 1744 identical modules, 250μm thick. Almost all modules have a pixel pitch of 50×400 μm^2, which is dictated by the pitch of the front-end electronics. Each pixel module has an intrinsic accuracy of 10 μm in the r, ϕ plane and and 115 μm in z. In total, there are approximately 80 million individual pixels bump bonded through to readout integrated circuit channels in the Pixel detector.

Semi-Conductor Tracker

Moving radially outwards, the next subdetector in the ID is the Semi-Conductor Tracker (SCT) [46]. The SCT is made of silicon microstrip modules, which operate similarly to the pixels but are elongated in the z direction. At the radius at which the SCT is located, this is a worthwhile trade-off, as continuing the use of pixels would be prohibitively expensive.

The SCT is comprised of four cylindrical overlapping layers in the barrel and nine discs at each end cap. There are a total of 6 million 6.4 cm long strips with 80 μm pitch and 285 μm thickness across 4088 modules, with 768 active strips per module. In each barrel module, two layers of strips are overlaid at a 40 mrad angle. This provides eight hits for any track traversing this region, in addition to the ability to measure hits in two coordinates. The endcap is made up of strips arranged radially, and an additional set of stereo strips offset at an angle of 40 mrad. The endcap strips also have a pitch of approximately 80

μm. This gives a per-module intrinsic accuracy for the SCT of 17 μm in the r, ϕ plane and 580 μm in the z direction.

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) [47] is the furthest ID subdetector from the interaction point, extending out to a radial distance of 1 m. The TRT is made of 4 mm diameter polyimide drift (straw) tubes, 144 cm/ 37 cm long in the barrel / endcap, providing tracking information in the r, ϕ plane only with a positional resolution of 130 μm. Each tube is filled with a mixture of oxygen, xenon, argon and carbon dioxide gas, and contain a 31 μm thick gold-plated tungsten 'straw' anode wire, with the wall of the tube acting as the cathode. When a charged particle passes through a straw, the gas inside is ionised and electron-hole pairs drift under the potential difference inside the straw, creating an electrical current which is collected and read out. In the barrel, there are 73 layers of straws arranged parallel to the beam axis. Each straw has an intrinsic accuracy of 130 μm. In the endcap, 160 layers of straws are arranged radially in wheels. In total, the TRT system has coverage up to $|\eta| < 2.0$ and involves approximately 351000 individual straws.

In between the straws, polypropylene fibres and foils provide abrupt changes in refractive index, causing particles to emit different amounts of transition radiation x-ray photons depending on their Lorentz boost. This provides information on Particle ID (PID), as the masses of particles with the same momentum can be distinguished, with lighter particles emitting more radiation. In particular, this PID system allows electrons and pions to be

distinguished. The transition photons are measured via the photoelectrons they create inside the tubes.

The dimensions of the three ID subsystems described in the sections above are summarised in Figure 2.7.

Item		Radial extension (mm)	Length (mm)
Overall Inner Detector (ID) envelope		$0 < r < 1150$	$0 < \|z\| < 3512$
Beampipe		$23.5 < r < 30$	
Pixel (Insertable *B*-Layer (IBL) included)	Overall envelope	$31 < r < 242$	$0 < \|z\| < 3092$
4 cylindrical layers	Sensitive barrel	$33.5 < r < 122.5$	$0 < \|z\| < 400.5$
2×3 discs	Sensitive endcap	$88.8 < r < 149.6$	$495 < \|z\| < 650$
Semiconductor Tracker (SCT)	Overall envelope	$255 < r < 549$ (barrel)	$0 < \|z\| < 805$
		$251 < r < 610$ (endcap)	$810 < \|z\| < 2797$
4 cylindrical layers	Sensitive barrel	$299 < r < 514$	$0 < \|z\| < 749$
2×9 discs	Sensitive endcap	$275 < r < 560$	$839 < \|z\| < 2735$
Transition Radiation Tracker (TRT)	Overall envelope	$554 < r < 1082$ (barrel)	$0 < \|z\| < 780$
		$617 < r < 1106$ (endcap)	$827 < \|z\| < 2744$
73 straw planes	Sensitive barrel	$563 < r < 1066$	$0 < \|z\| < 712$
160 straw planes	Sensitive endcap	$644 < r < 1004$	$848 < \|z\| < 2710$

Figure 2.7: A summary table of the dimensions of the components of the ATLAS Inner-Detector [32].

2.2.2 Calorimeters

ATLAS measures the energy of particles that interact both electromagnetically and hadronically using several types of sampling calorimeter technology [35] [36]. Sampling calorimeters work by alternating layers of 'active' material, which sample particles' energy, and 'passive' layers of dense, high Z material. Upon entering the calorimeter, a primary particle is absorbed by the passive layer and triggers a cascade of secondary particle 'showers'. The physical process triggering the shower depends on both the type of incident

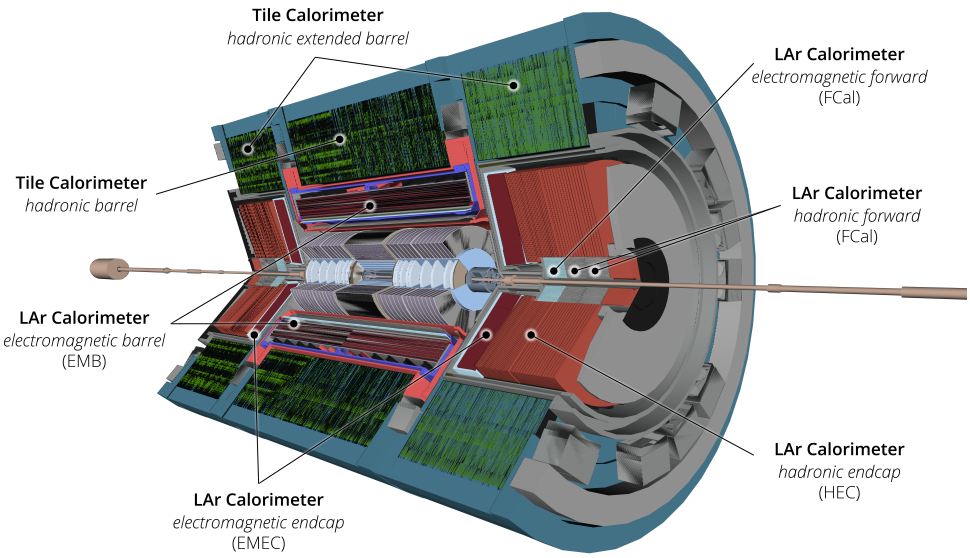


Figure 2.8: Cut-away diagram of the ATLAS calorimeter configuration as of the start of LHC Run 3, with labels indicating the major subsystems [32].

particle and the calorimeter technology employed - examples include bremsstrahlung for the case of electrons and pair production for the case of photons. The showers then interact with material in the active layer to produce a detectable signal. This signal is collected between layers, before the secondary particles enter a new passive layer and trigger further showers. This process continues until secondary particles no longer have sufficient energy to produce new showers, after which the primary particle is considered to have been fully absorbed. The total magnitude of the shower's measured signal is proportional to the energy of the original incident particle. Because of this, calorimeters aim to fully stop any particle for which they measure energy, and hence most SM particles are stopped in the ATLAS calorimeters. A cut-away diagram of the ATLAS calorimeter system as of the beginning of LHC Run 3 is shown in Figure 2.8.

The ATLAS calorimeters are hermetic in ϕ and have an acceptance up to $|\eta| < 4.9$

so that most showers fall within the calorimeters' acceptance, a feature which is also beneficial for $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The calorimeters are segmented in the transverse plane to provide positional information, and in z to give information of shower shape which can aid particle identification. The segmentations have a finer granularity in the region covered by the ID, and become coarser in the forward region where only jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ can be reconstructed.

Because calorimeters must completely absorb all showers triggered by a primary particle, depth is an important design consideration. For electromagnetic calorimeters, this is quantified using radiation length X_0, the mean distance in the material for a high-energy electron to have its energy reduced by a factor of $\frac{1}{e}$. For hadronic calorimeters, the depth measure is the interaction length λ, which is the mean distance travelled by a hadronically interacting particle before undergoing a hadronic interaction with the material.

The general expression for energy resolution (σ_E) in the ATLAS calorimeters [28] can be described using the equation:

$$\frac{\sigma_E}{E} = \frac{N}{E} \oplus \frac{S}{\sqrt{E}} \oplus C \tag{2.3}$$

where E represents energy and N, S, C are parameters specific to individual types of calorimeter technology, and are usually derived from fits to calibration data [50]. The first term in Eq. 2.3 is a noise term, arising from electronics, soft QCD and pile-up and significant in the low-energy regime. The second term is a stochastic term arising from sampling statistics and shower shape evolution and dominates the resolution up to several

hundred GeV. The final term corrects fluctuations that are independent of energy, including detector effects and non-uniformities. For the ATLAS electromagnetic calorimeters, typical resolution is approximately $\frac{\sigma_E}{E} \approx \frac{10\%}{\sqrt{E}} + 0.7\%$. For the hadronic calorimeters, which must contend with more complex hadronic showers, typical resolution is approximately $\frac{\sigma_E}{E} \approx \frac{50\%}{\sqrt{E}} + 3.4\% \oplus \frac{1\%}{E}$ in the barrel and endcap regions and $\frac{\sigma_E}{E} \approx \frac{100\%}{\sqrt{E}} + 10\%$ in the forward calorimeters.

The specifics of the individual subsystems making up the ATLAS calorimeters will now be described below.

Electromagnetic Calorimeter

The ATLAS Electromagnetic calorimeter (ECAL) is designed to measure energy from electromagnetically interacting particles i.e. electrons, photons, and a portion of jets' energy. The ECAL is a lead – liquid argon (LAr) sampling calorimeter [35], with 1024 layers of lead in an accordion shaped structure acting as the passive material. Liquid argon fills the space between each lead layer, chosen because it is both stable and radiation hard. The signal is gathered by Kapton electrodes stationed in between each layer.

The ECAL has complete symmetry in ϕ and is divided in to two parts, a barrel section for $|\eta| < 1.475$ and two endcaps covering the range $1.375 < |\eta| < 3.2$, each with their own cryostat system for maintaining the argon in a liquid state. The **LAr Electromagnetic Barrel** (EMB) [51] is designed to have a radiation length of $X_0 > 22$ and is divided into two half barrels with a small gap at z=0. The **LAr Electromagnetic Endcaps** (EMECs)

[52] have a radiation length of $X_0 > 24$ and are made up of two coaxial wheels – an outer wheel covering $1.375 < |\eta| < 2.5$ and an inner wheel covering region $2.5 < |\eta| < 3.2$.

In the central region $|\eta| < 2.5$ covered by the ID, the ECAL is three layers deep, with two layers in the remaining regions. The innermost layer is divided into 4 mm strips in η, which is sufficient granularity to resolve individual photons generated by neutral pions. The remaining layers are made up of coarser 'towers' in η, ϕ of dimensions 0.025 × 0.025 and 0.05 × 0.025 respectively for the measurement of shower shape. The region $|\eta| < 1.8$ is further supported by a presampler [53] composed of one active LAr layer of thickness between 0.5 – 1.1 cm, to correct for energy lost by particles before they reach the ECAL.

Hadronic Calorimeter

The ATLAS Hadronic Calorimeter (HCAL) employs multiple different technologies to measure the energy of hadronically interacting particles. In total, it is approximately 9.7/ 10 interaction lengths deep in the barrel/ endcap region(s) respectively [28]. This depth is sufficient to reconstruct high-energy jets with acceptable resolution and to reduce the level of jet punch-through to the muon system to below the level of muons produced by the non-prompt decay of heavy flavour jets.

Extending radially outwards from the ECAL the first layer of the HCAL encountered in the barrel region is hadronic **Tile** calorimeter [36], a sampling calorimeter with 64 modules made of steel passive layers and plastic scintillating tiles. The Tile calorimeter consists of three layers in the barrel region ($|\eta| < 1.0$) and two layers in its extended barrels

($0.8 < |\eta| < 1.0$), and is segmented into η, ϕ towers of dimensions 0.1 $\times$ 0.1 and 0.2 $\times$ 0.1 for the first two layers and outermost layer respectively. The photon signal from the scintillating tiles is read out using wavelength shifting optical fibres and measured using photomultiplier tubes.

In further forward regions, two different types of calorimeter technology are employed. The **LAr Hadronic Endcap Calorimeter** (HEC) [35] [54] covers the range $1.5 < |\eta| < 3.2$ and is made up of two wheels per endcap. The HEC is composed of active LAr layers and passive copper layers to optimise resolution and heat removal. The HEC is segmented into towers of dimensions 0.1 × 0.1 up to $|\eta| < 2.5$, after which towers have dimensions 0.2 x 0.2.

Finally, the **LAr Forward Calorimeter** (FCal) [55] covers the most forward region $3.1 < |\eta| < 4.9$, and uses liquid argon as its active material. For its passive layers, this calorimeter employs copper in its first layer and tungsten in subsequent layers to provide containment for the hadronic showers.

2.2.3 Muon Spectrometer

The Muon Spectrometer (MS) [28] is the outermost subdetector in ATLAS, and is designed specifically to measure muons, the only interacting Standard Model objects expected to survive the calorimeter systems. The muon system provides tracking coverage up to $|\eta| < 2.7$ alongside triggering up to $|\eta| < 2.4$. The system is comprised of superconducting air core toroid magnets, precision muon trackers and a fast muon trigger system.

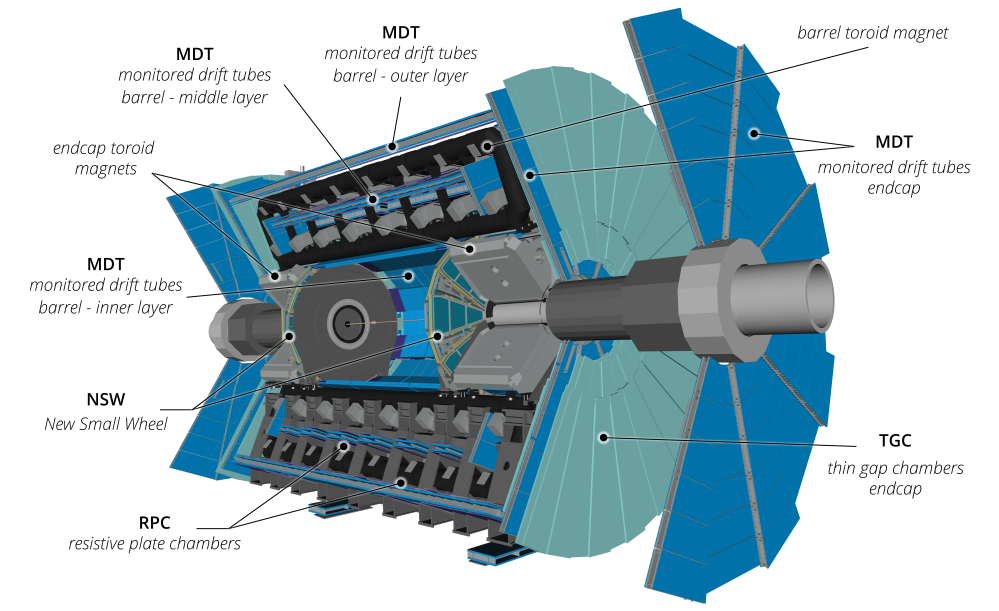


Figure 2.9: Cut-away diagram of the ATLAS muon spectrometer configuration as of the start of LHC Run 3, with labels indicating the major subsystems [32].

The system must provide muon p_{T} measurements over a broad spectrum, with typical momentum resolution of $\frac{\sigma_{p\mathrm{T}}}{p_{\mathrm{T}}} = 10\%$ for 1 TeV tracks. A cut-away diagram of the ATLAS MS as of the beginning of LHC Run 3 is shown in Figure 2.9.

Precision Muon Trackers

The precision muon trackers [28] are designed to measure muon momentum and charge ID via bending in the r, z plane. They also supply the muon's positional coordinate in the bending plane, with the other azimuthal coordinate provided mainly by the muon trigger system. The bending is supplied by superconducting toroid magnets with 0.5 T and 1 T fields in the barrel/ endcap regions respectively. In the barrel region, the tracking chambers are arranged in three concentric cylinders centred around the beam at radii of approximately 5 m, 7.5 m, and 10 m. In the end-caps the chambers are arranged in

wheels perpendicular to the z-axis. The wheels are positioned at distances between z = 7.4 m - 21.5 m from the interaction point. The alignment of the chambers must be very precisely known (to $< 30\mu$m) in order to achieve the desired sagitta resolution.

The main technology employed in the muon trackers are 1098 **Monitored Drift Tubes** (MDTs), which provide spatial resolution of 35 μm per chamber up to $|\eta| < 2.7$. Muons passing through the MDTs ionise the gas inside the tubes, a mix of 93% argon and 7% carbon-dioxide chosen for its robust ageing properties. The resulting electrons and ions drift towards a tungsten rhenium anode held at 3 kV at the centre of the tubes. In the barrel region, the MDTs are arranged in 16 overlapping sections with a gap at z=0 due to cabling for the solenoid magnet, with further gaps arising in some sectors due to mechanical supports.

In further forward regions $2.0 < |\eta| < 2.7$ where the large muon flux impacts the trackers' radiation tolerance and reconstruction efficiency, the first layer of MDTs is replaced by a system of two disks of multiwire proportional chambers called **Cathode Strip Chambers** (CSCs) with eight chambers each. The cathodes in these chambers are segmented into orthogonal strips, one parallel to the wire and the other perpendicular. The CSCs have a resolution of 60 μm in the bending plane compared to 80 μm per MDT tube. Additionally, the split CSC cathodes provide the azimuthal coordinate with a resolution of 5 mm. The CSCs have a drift time of 40 ns compared to the MDTs' 700 ns, leading to better timing resolution and muon reconstruction performance in the high occupancy region.

Muon Trigger System

In addition to precision tracking, the ATLAS MS also includes hardware for triggering. The muon trigger system [28] is designed to trigger on muon hits in different layers, providing accurate bunch crossing identification requiring a total response time of $<$ 25 ns. The muon trigger system can also measure both the η and ϕ spatial coordinate, providing the additional coordinate for the muon not supplied by the MDTs.

Two separate subsystems make up the muon trigger system. In the barrel region, three cylindrical layers of **Resistive Plate Chambers** (RPCs) provide triggering on muon tracks with p_{T} thresholds ranging from 6 GeV to 35 GeV. These chambers are made up of plates positioned 2 mm apart with an electric field of 4.9 kV/mm between them, and have a response time of 1.5 ns. These chambers are read out by capacitive coupling to the metallic strips on the outside of the plates.

In the endcaps up to $|\eta| < 2.4$, 3-4 layers of **Thin Gap Chamber** (TGC) multiwire proportional chambers provide triggering with a response time of 4 ns. The TGCs are made of two parallel plates with a carbon-dioxide and n-pentane gas mixture in between, with a wire-to-cathode distance smaller than the wire-to-wire distance. Muons passing through the chambers ionise the gas and cause avalanches of electrons which are read out as signal.

During the ATLAS Phase-I upgrades during the LHC Long Shutdown 2, the innermost layers of the MS endcaps commonly referred to as the 'small wheels' were replaced in advance of LHC Run 3. The '**New Small Wheels**' (NSWs) [56], occupy the same position

and cover the same range $1.3 < |\eta| < 2.7$ as the previous system. Two new technologies, **Small-Strip TGCs** (sTGCs) and MICRO-MEsh GASeous detectors (**Micromegas**), provide sufficiently fast response time for integration into the ATLAS Level 1 trigger (Section 2.2.4). The NSWs provide improved p_{T} resolution for the trigger alongside increasing the background rejection, allowing for lower muon p_{T} trigger thresholds while maintaining an acceptable trigger rate.

2.2.4 ATLAS Trigger System

With collisions in ATLAS occurring every 25 ns, it would be impossible to save every event without quickly exceeding the available computing resources, bandwidth and data storage constraints. The solution put in place by ATLAS, and many other particle physics experiments, is to implement a 'trigger' system which saves only events with interesting signatures and high energy objects. By prioritising saving such events, trigger systems are essential for returning a dataset dense in interesting physics and making efficient use of resources.

The ATLAS trigger system [57] [58] [59] has two main layers; Level 1 (L1) and the High-Level Trigger (HLT), where each subsequent layer refines decisions and implements additional selections with respect to the previous layer, reducing the rate of events passing each time. During each ATLAS run, a 'trigger menu' is provided, which is a list of available trigger chains (combinations of L1 and HLT selections) that can be enabled by the ATLAS Central Trigger Processor. The set of triggers enabled depends on the goal of a particular

run. Certain chains in a trigger menu may be 'prescaled', a technique which optimises the use of the available bandwidth by saving only a specified fraction of randomly selected events passing a given chain. This technique is typically used with low threshold trigger chains, or to adapt on the fly to changing detector and beam conditions.

The ATLAS **L1 trigger** consists of custom-built electronic hardware, and is capable of making decisions in 2.5 μs, reducing the event rate to 100 kHz. The L1 trigger uses coarse granularity information, and generally passes events with a high p_{T} hit in either the calorimeters or muon system. The L1 trigger also defines Regions of Interest (RoIs) in η, ϕ based on these hits, which are passed onto the HLT.

The **High Level Trigger** is software-based and operates as similarly to the ATLAS offline event selection as possible to refine the decisions made by the L1 trigger. The HLT takes the RoIs as inputs and adds the full precision and granularity available from the subdetectors, either in this region only or for the full event depending on the trigger chain. During Run 2, the average event processing time by the HLT was 400 ms, reducing the rate to 3 kHz. Compared to the L1, the HLT has the capability to apply more complex procedures like b-tagging or $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction in its decisions.

Finally, events passing the HLT are subject to the full ATLAS offline event reconstruction procedure. Data recorded by ATLAS, along with Monte-Carlo simulations produced by ATLAS are initially stored at CERN's Tier 0 of the Worldwide LHC Computing Grid before being distributed around the world. When this data is analysed, object reconstruction, calibration and initial preselections are typically performed using the internal ATLAS

'Athena' software. Final data processing and plot production is generally done with ROOT or Python.

The result of all the subsystems described in Section 2.2 operating in tandem is shown in the example in Figure 2.10, of an event recorded by the ATLAS detector on 11 May 2023 during the first collisions of stable 6.8 TeV beams of 2400 bunches of protons [60].

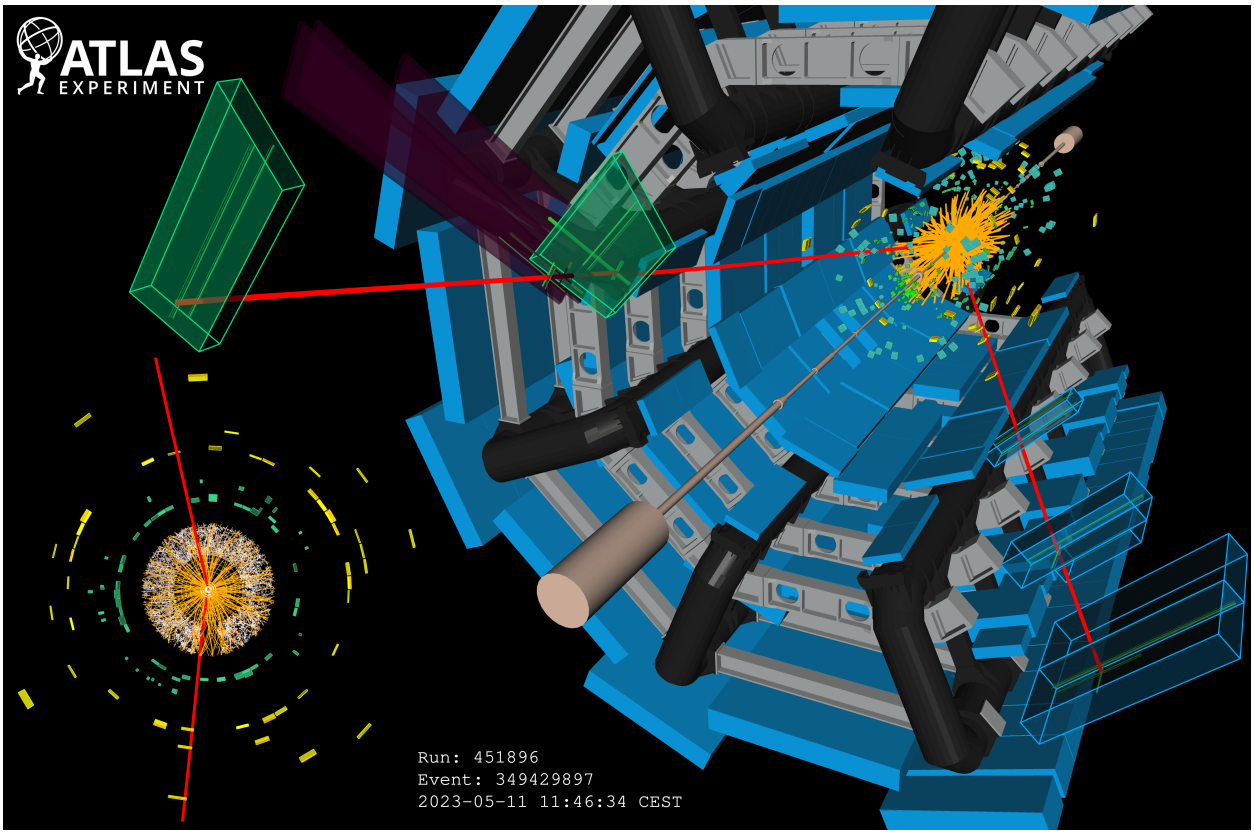


Figure 2.10: A transverse plane view of an event recorded by the ATLAS detector on 11 May 2023 during the first collisions of stable 6.8 TeV beams of 2400 bunches of protons [60]. Starting from the centre Pixel hits (white dots) are shown, followed by hits in the SCT (yellow strips), and TRT (white/red dots). These hits are fit together to reconstruct charged particle tracks (orange lines). Moving radially outwards, energy deposits in the barrels of the electromagnetic LAr calorimeter (green squares) and the hadronic Tile hadronic calorimeter (yellow/orange squares) are shown. Finally, the tracks of two reconstructed muons from the Interaction Point our to the Muon Spectrometer (red lines) are shown.

Chapter 3

Object reconstruction in ATLAS

"If the shoe fits, walk in it 'til your high heels break"

— Taylor Swift, 'Long Story Short'

The ATLAS detector is designed such that each type of particle produced in a pp collision results in a distinct set of signals recorded in each subsystem described in Chapter 2. Subsequently, these signals are processed to infer the type of particle or 'object' which deposited them, in a procedure called 'object reconstruction'. In ATLAS, 'low-level objects', or collections of detector deposits, are initially reconstructed and serve as the basis for the rebuilding of 'high-level objects', representing an event's collision products. Prior to their use in physics analysis, these collections of particles are subject to quality-control requirements, and energy or transverse momentum calibration. ATLAS maintains specialised performance groups, each with expertise in different objects, dedicated to optimising the reconstruction process, performing calibrations and evaluating the associated

uncertainties.

3.1 Low-level objects

The first stage of the object reconstruction procedure in ATLAS is the initial processing of the signals read out by individual subsystems of the detector, resulting in the rebuilding of 'low-level objects'. These objects themselves do not represent individual particles produced by a collision, but rather a particle's interaction with the material of the detector; for example, a charged particle's trajectory in a magnetic field, or electromagnetic or hadronic showers in the calorimeter. Ultimately, each type of reconstructed particle is built from different combinations of these low-level objects. In this section, the following low-level objects will be discussed; tracks, vertices and calorimeter clusters.

3.1.1 Tracks

ID tracks

'Tracks' represent a charged particle's trajectory through a magnetic field, from which the charge and momentum of the particle can be inferred (see Eq. 2.2). Track reconstruction in the ATLAS Inner-Detector is performed in two stages: first, 'inside out', then 'outside in' [61].

The inside out procedure begins with preprocessing 'clusters' of hits in the silicon Pixel

and SCT trackers, created by combining groups of pixels or strips with common edges or corners in which the deposited signal is above a certain threshold [48]. These clusters are then translated into three dimensional 'space-points', and represent the point at which a charged particle intersected the active material.

Track candidates are seeded from sets of three space-points in either the Pixel or SCT, with loose selection criteria applied, which are compatible the the trajectory of a charged particle track. A fourth 'confirmation space-point', one which would be compatible with the candidate track's curvature were it to replace either the innermost or outermost space-point in the seed, is subsequently included. From this updated seed a 'search road' of detector modules compatible with the track candidate's trajectory is defined, where compatible clusters are expected to be found. The seeds are then extended along the search road in both directions by incorporating these additional clusters using a Kalman filter [62] to produce the final collection of track candidates.

Overlaps between track candidates, and the rejection of fake tracks from incorrect combinations of clusters, are then resolved using the 'ambiguity solver'. Track candidates are scored on the basis of quality and number of shared clusters with other nearby candidates, and lower quality tracks with many shared clusters are rejected. The estimated position and uncertainties of the remaining track candidates are then updated using a neural network. The resulting collection of track candidates are refit using a global χ^2 approach to return a collection of high-precision tracks. Finally, these tracks are extended into the TRT using the same road search - Kalman filter - χ^2 fit procedure described above.

The inside out approach is the primary method by which the high precision tracks of primary particles produced in pp interactions are reconstructed. However, to increase the acceptance of particles produced further from the beam line (e.g. conversion photons), the outside in procedure is subsequently applied - a secondary back-tracking pass originating from the TRT using only hits that have not yet been assigned to tracks. The outside in procedure is seeded by segments of TRT hits located in 'Regions of Interest', as determined by EM calorimeter deposits. In regions close by the segment, silicon track seeds are constructed using two space-points, which are subsequently extended into track candidates using the same procedure as the primary track reconstruction before being (re)extended back into the TRT. Unused TRT segments are reconstructed as TRT-Standalone tracks, which, in addition to "special" tracking passes also performed by ATLAS, are not used in the work presented in this document and will therefore not be discussed further.

The tracks used in the pile-up tagging and W+jets sections of this thesis are required to pass the 'Tight Primary' quality selections as recommended by the ATLAS tracking performance group, requiring:

- $p_{\mathrm{T}} >$ 500 MeV;
- $|\eta| <$ 2.5;
- Number of Pixel+SCT hits $\geq$ 9 for $|\eta| \leq$ 1.65, $\geq$ 11 for $|\eta| >$ 1.65;
- 0 Pixel holes (a missing track in an active sensor), and ≤ 2 SCT holes;
- ≥ 1 hit in either the IBL or b-layer, if hits are expected in both.

Muon spectrometer tracks

Muon-spectrometer tracks are initially reconstructed using localised short, straight-line track segments in different stations. These segments are combined into three-dimensional initial track candidates based on the position of the interaction point and the parabolic bending of a muon in the magnetic field using precision measurements in the bending plane from the MDTs and an additional coordinate from the muon triggers [63]. Muon tracks are then reconstructed using a χ^2 fit of the muon's path through the magnetic field, accounting for interactions with the detector material and possible misalignment between different chambers. Based on the trajectory determined by the fit, the set of hits included in the track is updated and the fit is repeated. Low-quality tracks which share a large proportion on hits with high-quality tracks are removed, and the final set of tracks is returned by performing a final fit accounting for energy-loss in the detector and constraints imposed on the IP.

3.1.2 Vertices

Using good quality ID tracks reconstructed as described in Section 3.1.1, it is possible to reconstruct the primary vertices at which pp interactions occur. A seed position for the primary vertex is selected based on the beam spot position in the transverse plane and the position of closest approach of the selected tracks to the beam spot. The tracks and the seed vertex position are then subject to an iterative fit procedure, where, in each iteration, incompatible tracks are removed. This procedure is repeated with the tracks

deemed 'incompatible' with the first reconstructed vertex until all other primary vertices in the event have been reconstructed. The event's hard-scatter primary vertex, PV_0, is determined to be the primary vertex with the largest associated sum of squared transverse momentum Σp_{T}^2.

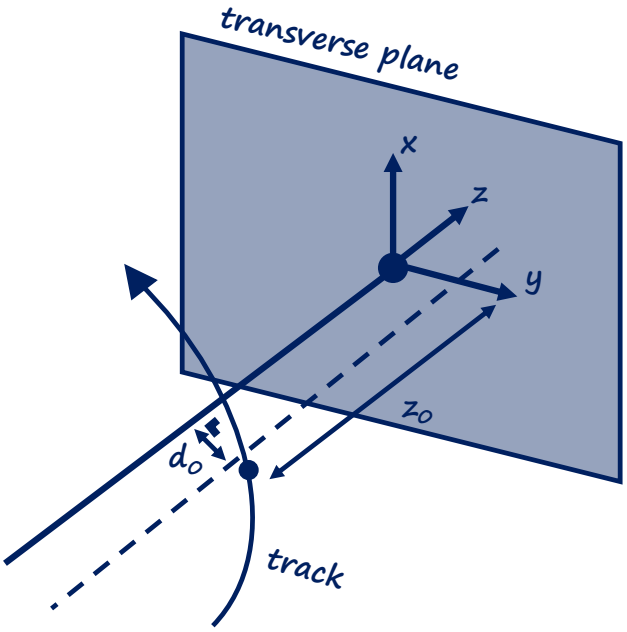


Figure 3.1: A sketch of the impact parameters on which many ATLAS object reconstruction procedures place selections.

Reliable reconstruction of PV_0 is required for the determination of particles' impact parameters, selections on which are frequently used to remove particles produced by light hadron decays (see Section 5.3), and to identify jets originating from heavy-flavour quarks (see Section 3.2.4). These variables are illustrated in Figure 3.1; d_0 is the distance of closest approach in the transverse plane of a particle to the beamline, and z_0 is the distance along the z axis from PV_0 to the position at which d_0 is measured.

3.1.3 Calorimeter clusters

'Calorimeter clusters' are collections of nearby calorimeter cells in both lateral and longitudinal directions containing significant deposits of energy from the showering of electromagnetically and hadronically interacting particles . The shorthand 'calo-clusters' or 'topoclustes' is frequently used to refer to calorimeter clusters.

'Topoclustering' [64] [65] algorithms reconstruct showers in ATLAS by iteratively combining neighbouring calorimeter cells measuring an energy E which is significantly greater than the expected level of noise σ from the detector electronics and pile-up. Topoclusters are seeded by calorimeter cells which have energy $E > 4\sigma$. Neighbouring cells are then included in the cluster if $E > \sigma$. Finally, all adjacent cells to the cluster are included if $E > 0$. If the topocluster is found to have more than one local maximum, the cluster is split; see Section 3.2.3 for further detail. Topoclustering algorithms are efficient at noise suppression, particularly for large clusters, and are therefore used for jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

For either type of clustering algorithm, the resulting clusters are initially calibrated to the approximate electromagnetic (EM) scale. The high-level objects they are used to reconstruct are subsequently fully calibrated based on type of particle. Generally, during the calibration procedure for reconstructed particles built using clusters, the particle's energy will be corrected for dead areas of the detector included in the cluster, and for any shower energy deposited outside of the reconstructed cluster.

3.2 High-level objects

'High-level objects', built using combinations of the low-level objects described in Section 3.1, represent the particles produced in a collision reconstructed from the signals left through interaction with the detector. Once reconstructed, these objects are subject to additional processing; applying selections to reduce backgrounds, calibrating object kinematics, or 'tagging' the object with additional information. This section summarises the reconstruction procedure for the objects used in the pile-up jet tagging studies described in Chapter 4 and in the muon channel of the W+jets analysis detailed in Chapters 5-7, however; this is not an exhaustive list of all high-level objects reconstructed by ATLAS.

3.2.1 Muons

Muons act as minimum-ionising particles as they pass through ATLAS, leaving signatures in every detector subsystem [63]. ATLAS muon reconstruction, which can be performed up to $|\eta| < 2.7$, relies primarily on tracking and hit information from the ID and MS, with some additional information from the calorimeters used to assist the determination of track parameters and to correct for large energy losses.

Several muon reconstruction strategies are employed by ATLAS, with each corresponding to a particular 'type' of muon:

- Combined (CB) muons are reconstructed by matching ID tracks to MS tracks and performing a combined fit of the matched tracks. Calorimeter information is added

to account for energy losses as the muon passes through the detector. Based on the outcome of the combined track fit, the set of MS hits included in the track is updated, and the combined track fit is repeated. In cases where the reconstructed muon falls outside the $|\eta| < 2.5$ acceptance of the ID, the MS track may instead be combined with shorter track segments made up of Pixel and SCT hits;

- Inside-out (IO) combined muons are built from ID tracks extrapolated to the MS, where at least three loosely aligned hits are matched. This method is beneficial in regions where there are gaps in the MS coverage due to cabling or mechanical supports, and for low p_{T} muons which may not reach the middle stations of the MS;

  – The muon channel of the W+jets analysis, presented in Chapters 5-7, uses CB and IO muons;

- Muon-spectrometer extrapolated (ME) muons are used when an MS track cannot be matched to an ID track, and instead the MS track is extrapolated directly to the beamline. This type of muon reconstruction is useful in regions where there is MS coverage but no ID coverage i.e. $2.5 < |\eta| < 2.7$;

- Segment-tagged (ST) muons require that an ID track extrapolated to the MS is matched, with tight angular constraints, to a hit in at least one MS segment. This reconstruction method is beneficial for low p_{T} muons;

- Calorimeter-tagged (CT) muons are reconstructed from ID tracks extrapolated through the calorimeters and matched to deposits consistent with a minimum-ionising particle. This approach yields better reconstruction efficiency in regions of limited MS

coverage.

Once muon candidates have been reconstructed, a collection of high quality muons must be identified for use in physics analysis. As individual analyses have distinct requirements, ATLAS defines a collection of 'Working Points' (WP) which each have different criteria and efficiencies for muon identification and rejection of backgrounds due to non-prompt muons from the in-flight semileptonic decay of light hadrons. The muon identification WPs are defined with respect to the number of hits in the ID and MS, properties of the track fits, and on the q/p compatibility of the measurements in the two subsystems, defined as:

$$q/p\,\text{compatibility} = \frac{|q/p_{\text{ID}} - q/p_{\text{MS}}|}{\sqrt{\sigma^2(q/p_{\text{ID}}) + \sigma^2(q/p_{\text{MS}})}},$$

where $q/p_{\text{ID,MS}}$ is the ratio of charge q and momentum p in the ID and MS respectively and $\sigma(q/p_{\text{ID,MS}})$ are the associated uncertainties.

In order of increasing purity and decreasing efficiency, the three 'standard' muon identification WPs are: `Loose`, `Medium` and `Tight`, where the selections defining tighter WPs are subsets of those for looser WPs. The muon identification efficiencies for both prompt muons and non-prompt muons from light-flavour hadron decays using these standard WPs are shown in Figure 3.2 as a function of η and p_{T} . Additional WPs are defined for use in extreme regions of phase space. For the `Medium` WP used in the muon channel of the W+jets analysis, CB and IO muons within $|\eta| < 2.5$ must have at least two stations where the muon has at least three hits ('precision stations') in the MDTs or CSCs and a

q/p compatibility of less than 7. In the region $2.5 < |\eta| < 2.7$, ME muons are also included and must have at least three precision stations.

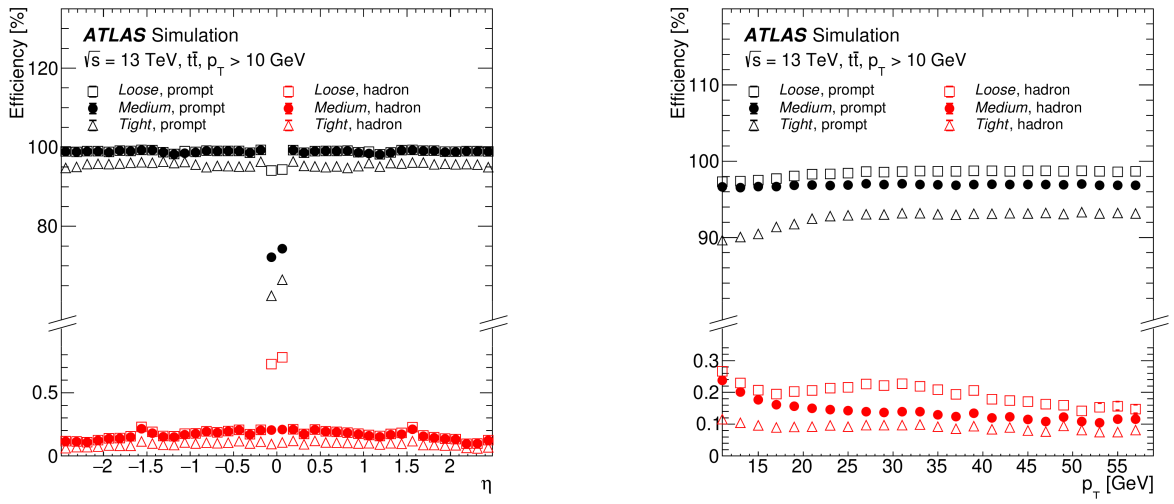


Figure 3.2: The muon identification efficiency for prompt muons and non-prompt muons from light-hadron decays of the `Loose`, `Medium` and `Tight` identification Working-Points as a function of muon pseudorapidity (left) and transverse momentum (right). Figure from [63].

To distinguish prompt muons from additional non-prompt backgrounds from heavy flavour hadron decays, a set of isolation selections is imposed, as prompt muons are less likely to have significant additional activity in their vicinity. Isolation variables can be either track or calorimeter based, with combinations of the two generally resulting in better performance. Both types of variables measure the amount of transverse momentum or transverse energy in a cone centred around the muon (with the muon's p_{T} or energy removed). Track based isolation variables are calculated by taking the scalar sum of the p_{T} of tracks associated with the primary vertex in a ΔR cone around the muon, where ΔR is either 0.2 or min(10GeV/p_{T}^{μ}, 0.3), labelled $p_{\mathrm{T}}^{\mathrm{cone20}}$ and $p_{\mathrm{T}}^{\mathrm{varcone30}}$ respectively. Calorimeter-based isolation calculates the sum of E_{T}, corrected for deposits from the muon and from

pile-up, in topological clusters located in a ΔR = 0.2 cone centred around the muon's trajectory through the calorimeters, and is labelled $E_{\mathrm{T}}^{\mathrm{topocone20}}$.

Similarly to muon identification, several isolation WPs are defined which balance prompt muon acceptance with background rejection, built from combinations of selections on the isolation variables. The '`PFlowTightFixedRad` WP applied to muons used in the W+jets analysis muon channel requires tracks to have $p_{\mathrm{T}} >$ 500 MeV and imposes the following selections on the isolation variables:

$$\begin{aligned}(p_{\mathrm{T}}^{\mathrm{varcone30}} + 0.4 \cdot E_{\mathrm{T}}^{\mathrm{topocone20}}) < 0.16 \cdot p_{\mathrm{T}}^{\mu}, \qquad (p_{\mathrm{T}}^{\mu} < 50\mathrm{GeV}),\\ (p_{\mathrm{T}}^{\mathrm{cone20}} + 0.4 \cdot E_{\mathrm{T}}^{\mathrm{topocone20}}) < 0.16 \cdot p_{\mathrm{T}}^{\mu}, \qquad (p_{\mathrm{T}}^{\mu} > 50\mathrm{GeV}).\end{aligned}$$

Differences in muon reconstruction, identification, isolation and track-to-vertex association efficiencies between data and MC must be accounted for and corrected. These efficiencies are measured in samples of $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ decays for the $3\,\mathrm{GeV} < \mathrm{p}_{\mathrm{T}}^{\mu} < 15\,\mathrm{GeV}$ and $p_{\mathrm{T}}^{\mu} > 15\,\mathrm{GeV}$ regimes respectively. Efficiencies are measured using the Tag and Probe method [63] in the region $|\eta| <$ 2.5. In this method, one of the final state dimuon pair, the *tag*, is subject to stringent identification and trigger criteria while the other *probe* muon is used to test the properties of a particular algorithm or WP. The efficiency of a particular criterion X is then calculated from the ratio of the number of probes P that are ΔR matched to a muon reconstructed or identified according to X to the total number of probes:

$$\epsilon(X|P) = \frac{N_P^X}{N_P^{\text{all}}}.$$

The level of agreement between efficiencies calculated in data and MC is quantified using scale factors (SF)

$$\text{SF} = \frac{\epsilon^{\text{Data}}}{\epsilon^{\text{MC}}},$$

which are deployed at analysis-level to correct for differences between data and MC. Finally, mismeasurement of the muon momentum scale and resolution, and MS alignment must also be calibrated. This procedure is described in detail in [66].

3.2.2 Electrons

Electrons are reconstructed in ATLAS by combining information from the ID and the electromagnetic calorimeters. In the central region $|\eta| < 2.5$, electron candidates can be identified via a track entering the ECal and depositing a calorimeter cluster in its vicinity. In the forward region, only calo-clusters are available for electron reconstruction, however; rejection of backgrounds from photons or pions is still possible in this region through the use of calorimeter variables such as electromagnetic shower shape.

Electron reconstruction is seeded by calorimeter topoclusters, which are reconstructed as described in Section 3.1.3. Only energy from cells in the EM calorimeter is considered, the the exception of the 'transition region' $1.37 \leq |\eta| \leq 1.63$ between the calorimeter barrel

and endcaps, where measurements from the calorimeter presampler and the scintillator between the calorimeter cryostats are also included [67]. This is referred to as the 'EM energy' of the cluster, while the ratio of a cluster's EM energy to its total energy is known as its 'EM fraction' f_{EM}. Only clusters passing preselections $f_{\mathrm{EM}} > 0.5$ and and EM energy $>$ 400 GeV are retained for electron reconstruction.

The tracks used for electron reconstruction are initially reconstructed as described in Section 3.1.1. Additionally, any candidate tracks which fail the initial track fit are refit accounting for electron energy loss in the detector via bremsstrahlung. This set of track candidates is subsequently matched to the barycentre of the seed clusters and refit again using an optimised Gaussian-sum filter [68] to better account for charged particle energy loss in material. The track-cluster matching procedure is then repeated with more stringent requirements.

Subsequently, collections of matched tracks and topoclusters are used to create electron candidate 'superclusters'. In the first stage, a set of seed clusters are identified, which are required to have $E_{\mathrm{T}} > 1$ GeV, and a matched track with ≥ 4 hits in the silicon detectors. Subsequently, the algorithm attempts to identify a set of 'satellite clusters' to be added to the seed cluster to recover energy losses in the detector. Initially, unused clusters within $\Delta\eta \times \Delta\phi$ = 0.075 $\times$ 0.125 of the barycentre of the seed clusters are considered, as these likely originate from secondary showers from the same primary electron. Furthermore, clusters within $\Delta\eta \times \Delta\phi$ = 0.125 $\times$ 0.300 are also considered if their 'best-matched' track is the same track matched to the seed cluster - such satellite clusters are likely the result of bremsstrahlung radiation or topocluster splittings. Finally, the set of reconstructed su-

perclusters are assigned calorimeter cells, are restricted to a maximum width of 0.075 or 0.125 in the η direction in the barrel and endcap region respectively to limit sensitivity to pile-up noise, and have their energy and position calibrated to produce the collection of reconstructed electrons used for analysis.

Similarly to muons, to ensure high-quality electrons are reconstructed for physics analyses, candidate electrons are subject to identification selections to increase the purity of prompt electrons and reduced backgrounds due to photon conversions or pion decays. This is performed using a likelihood (LH) based discriminant [67] [69]. This discriminant is calculated using an Multi-Variate Analysis of several input variables including; hits in each layer of the ID, ECal shower shape and energy density, impact parameters with respect to the IP and the track-cluster matching. Several WPs corresponding to values of the LH discriminant are defined, corresponding to trade-offs between signal purity and electron identification efficiency.

Isolation working points are also defined in order to reduce backgrounds from semileptonic heavy quark decays and from photon conversions. Similarly to muons, isolation variables are defined with respect to the degree of track activity or energy deposits, corrected for the electron energy, pile-up and energy leakage, within a cone centred around the electron, from which electron isolation WPs are derived.

Finally, mismeasurements of the electron kinematics must be corrected, in addition to differences in detector responses for electron reconstruction, identification and isolation between data and MC. The procedure for the former is described in [70], and the latter is

corrected via multiplicative scale-factors evaluated in $J/\psi, Z \to ee$ samples as described in [69].

3.2.3 Jets

During LHC Run 1, hadronic jets were rebuilt in ATLAS using information from either solely the calorimeter or solely the tracker, with the majority of analyses using calorimeter topoclusters [65] as inputs to jet clustering algorithms. During Run 2, an alternative approach was introduced, where jets are seeded by collections of matched tracks and topological clusters called Particle Flow Objects (PFOs) [71].

The combination of ID and calorimeter measurements versus the use of calorimeter information alone provides several benefits for jet reconstruction, including:

- The ID has an inverse transverse momentum resolution of $\frac{\sigma_{p\mathrm{T}}}{p_{\mathrm{T}}} = 0.036\% \cdot p_{\mathrm{T}} \oplus 1.3\%$, whereas the calorimeter energy resolution for single charged pions in the central region is $\frac{\sigma_E}{E} = \frac{50\%}{E} \oplus \frac{1\%}{\sqrt{E}} \oplus 3.4\%$. Therefore, for low energy charged particles, the tracker has a significantly better resolution than the calorimeter;
  - Additionally, the tracker also has better angular resolution for single charged particles;
- The tracker has track-to-vertex association capabilities, meaning that hadronic activity originating from in-time pile-up vertices (see Chapter 4) can be rejected by requiring that PFlow jets be associated with the hard-scatter vertex;

- Low-p_{T} charged particles can be clustered into a jet using tracking information before they are swept out of the jet cone by the magnetic field;

- The calorimeter provides the ability to reconstruct neutral particles, coverage in the ATLAS forward region and better energy resolution for highly-energetic particles.

The ATLAS PFlow algorithm [71] uses tracks and positive-energy topoclusters as inputs, and returns PFOs comprised of matched tracks and clusters. For every charged particle for which the track information is used, the corresponding energy must be subtracted from cells in the matched cluster to prevent the double-counting of transverse energy/ momentum. An overview of the PFlow algorithm is presented in Figure [71], with each step summarised below:

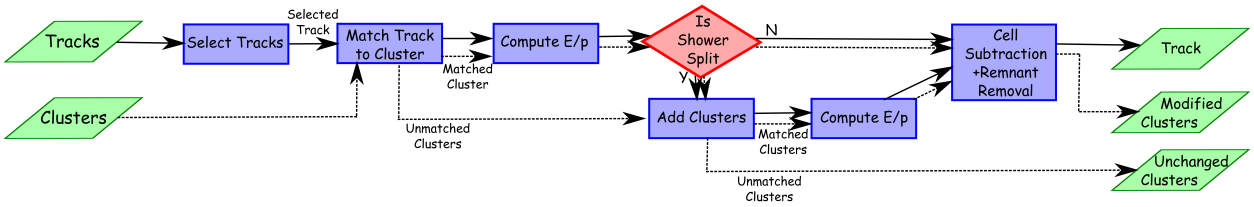


Figure 3.3: An overview of the Particle Flow algorithm which reconstructs the Particle Flow Objects used as inputs to the jet clustering algorithms ATLAS employs to reconstruct jets. Figure from [71]

.

1. The set of all well-reconstructed tracks in event are selected; tracks with at least nine hits in the silicon detectors, no missing Pixel hits, $|\eta| < 2.5$ and $0.5 < p_{\mathrm{T}} < 40$ GeV. The upper bound on track p_{T} is included as energetic particles often have significant nearby hadronic activity. As this makes removing the associated clusters challenging, it is therefore more accurate to account for energetic particles using the

calorimeter. Tracks previously associated with either an electron or muon are also removed.

2. Each selected track is matched to the barycentre of a single reconstructed topocluster;

3. For successfully matched tracks, the expected energy deposited in the calorimeter by the particle that created the track, E/p, is calculated based on the matched cluster position and track momentum;

4. For each track-cluster system, the likelihood that all the particle's energy was deposited the single matched cluster is evaluated. On this basis, more clusters are added to the system until the whole shower energy is recovered;

5. The expected energy deposit from track is subtracted cell-by-cell in each matched cluster;

    - If the remaining energy in the matched clusters is consistent with expected fluctuations from the hadronic shower, the clusters are completely removed.

This procedure returns a set of Particle Flow Objects for each event, containing matched track and energy-subtracted cluster collections generated by charged particles, in addition to unmatched, unmodified clusters from neutral particles. Before use in ATLAS jet reconstruction, tracks in the PFOs are required to match the hard-scatter primary vertex within $|z_0 \sin\theta| < 2$ mm to suppress pile-up, and clusters are required to have positive energy. PFOs meeting these requirements are then used as inputs to the anti-k_{T} [72] jet-clustering algorithm with a radius parameter of 0.4.

Anti-k_{T} is a conceptually-simple, Infra-Red and Collinear (IRC) safe algorithm for hadronic jet recombination, which results in jets with regular boundaries which are not influenced by soft radiation emissions. It can be derived from a generalisation of sequential combination clustering algorithms, of which k_{T} [73] and Cambridge/Aachen [74] are also IRC-safe examples; however, the boundaries of jets clustered using these algorithms can be irregular as the result of soft radiation emissions - see Figure 3.4.

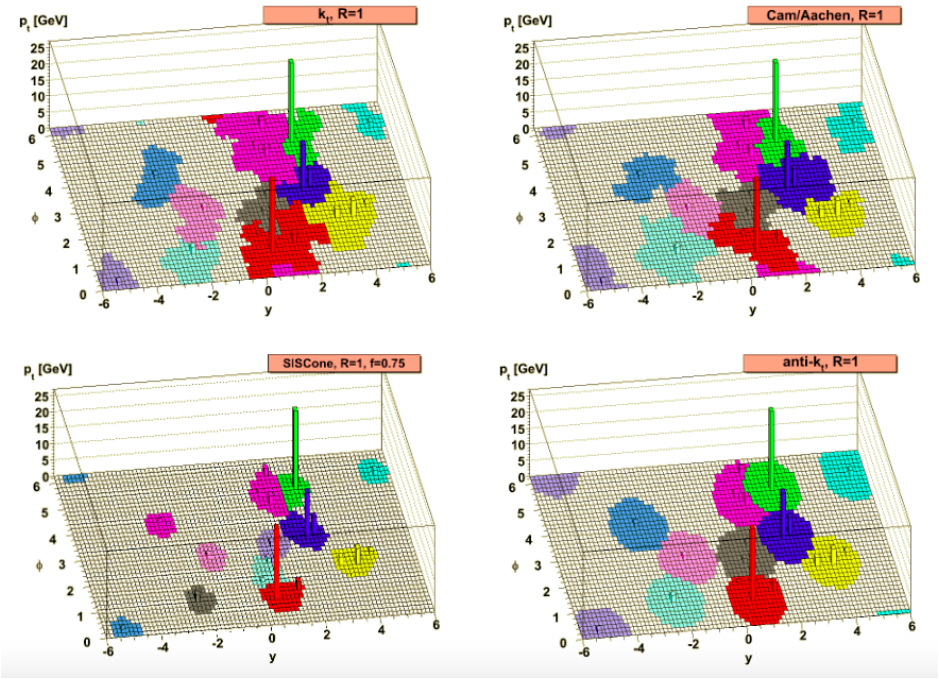


Figure 3.4: The jet shapes generated by a variety of jet clustering algorithms, including anti-k_{T} , in a parton-level event generated by Herwig version 6.5 [75]. Figure from [72].

The generalised case [72] of the sequential combination algorithm can be expressed by defining the distance measures:

$$d_{ij} = \min(k_{\mathrm{Ti}}^{2p}, k_{\mathrm{Tj}}^{2p}) \frac{\Delta_{ij}^2}{R^2},$$
$$d_{i,\mathrm{B}} = k_{\mathrm{Ti}}^{2p},$$

where d_{ij} is the distance between inputs i, $j \in$ (Topoclusters, PFOs, etc., depending on the use case), $d_{i,\mathrm{B}}$ is the distance between input i and the beam B. k_{Ti} is the transverse momentum of input i and $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is defined with respect to the inputs' rapidities y and azimuthal angles ϕ. The distance metric also depends on two parameters: the jet cone R, and p, which governs the relative power of the energy scale versus the geometrical scale. p=1 corresponds to the k_{T} algorithm, p=0 recovers the Cambridge-Aachen algorithm and $p = -1$ corresponds to anti-k_{T} . In the anti-k_{T} case, the distance metric d_{ij} becomes:

$$d_{ij} = \min(\frac{1}{k_{\mathrm{Ti}}^2}, \frac{1}{k_{\mathrm{Tj}}^2})\frac{\Delta_{ij}^2}{R^2}. \tag{3.1}$$

Jet recombination is then performed by iteratively identifying the smaller of the distance metrics. If $d_{ij} < d_{i,\mathrm{B}}$, i and j are combined into a single 'pseudojet' which is retained in the set of inputs. If $d_{ij} > d_{i,\mathrm{B}}$, i is labelled a jet and removed from the list of inputs. The factors of $1/k_{\mathrm{T}}$ in Eq. 3.1 ensure that hard particles are prioritised over a similarly separated soft particles, and importantly, soft emissions do not modify the shape of the final jet, which is almost circular.

The energy response of reconstructed anti-k_{T} jets is subsequently calibrated to the ATLAS Jet Energy Scale (JES) [76], to restore the measured reconstructed jet energy to the particle-level energy scale. This calibration sequence involves multiple stages, which are illustrated in Figure 3.5 and summarised below:

1. **Origin correction:** The four-momentum of the jet is recalculated with respect to the

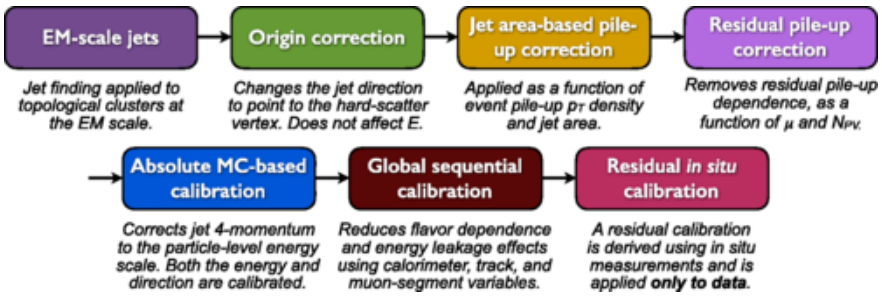


Figure 3.5: A summary of the ATLAS Jet Energy Scale calibration sequence. Figure from [76].

hard-scatter primary vertex, rather than the origin of the detector. This procedure improves the jet η resolution;

2. : **Pile-up subtraction:** Excess p_{T} originating from pile-up is removed from the reconstructed jet in two stages. **'Area-based pile-up subtraction'** first subtracts the pile-up contribution to the p_{T} of each jet based the the jet area A and the event's average pile-up p_{T} density ρ in the η, ϕ plane:

$$p_{\mathrm{T}}^{\mathrm{corr},1} = p_{\mathrm{T}}^{\mathrm{reco}} - \rho \times A.$$

Subsequently, a **residual pile-up correction** is applied for the dependence exhibited by ρ on the number of primary vertices N_{PV} and on the mean number of interactions per bunch crossing μ, which are sensitive to in-time and out-of-time pile-up and are discussed in further detail in Section 4.1:

$$p_{\mathrm{T}}^{\mathrm{corr2}} = p_{\mathrm{T}}^{\mathrm{corr1}} - \alpha \times (N_{\mathrm{PV}} - 1) - \beta \times \mu,$$

where the parameters α and β are derived from a fit to MC.

3. The **absolute JES calibration** corrects the jet four-momentum to the particle-level

energy scale using a jet energy response derived from a Gaussian fit to the core of the $E^{\mathrm{reco}}/E^{\mathrm{truth}}$ distribution for truth jets in Pythia dijet MC [77]. A numerical inversion is then used to derive a transformation from E^{reco} to E^{truth} [78]. A correction for η biases in the calibration, due to differences in calorimeter granularity at different pseudorapidities and transition regions between calorimeter technologies, is also applied at this stage;

4. The reconstructed four-momentum and associated uncertainties are further improved at the **Global Sequential Calibration** stage, by correcting for residual dependencies on jet flavour and longitudinal and transverse shower shape. Jets initiated by gluons are often softer with more particles spread over a wider area, whereas quark-initiated jets contain fewer hadrons with a higher fraction of the jet p_{T} each, which are more likely to punch through the HCal to the muon-spectrometer;

5. Finally the ***in-situ* calibrations** are performed to correct jets in data for differences between jet response in data and MC. This is performed using well-measured reference objects which balance a jet's p_{T} to directly measure the JES, including Z bosons and photons. This stage also includes the 'η-intercalibration' extrapolation to the forward region of the jet energy response derived in well-measured central jets. For each *in-situ* calibration, a correction factor c is derived from the double ratio of jet response $\mathcal{R}$ in data and MC:

$$c = \frac{\mathcal{R}_{\mathrm{data}}}{\mathcal{R}_{\mathrm{MC}}} = \frac{\langle p_{\mathrm{T,probe}}/p_{\mathrm{T,ref}}\rangle_{\mathrm{data}}}{\langle p_{\mathrm{T,probe}}/p_{\mathrm{T,ref}}\rangle_{\mathrm{MC}}} \,.$$

In regions of jet p_{T} where the jet energy resolution in data is better than in MC, the

simulation is subsequently smeared until its average resolution matches that of data. This smearing is based on the Jet Energy Resolution (JER) measurement, which takes the general form:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C,$$

where the noise N, stochastic S and constant C terms have previously been discussed with respect to the calorimeter energy resolution Eq. 2.3, and are determined using balancing techniques against well-measured reference objects.

Finally, reconstructed jets are subject to pile-up tagging procedures, to identify and remove any jets not originating from the primary hard-scatter vertex. These procedures are discussed in detail in Chapter 4.

3.2.4 Jet Flavour Tagging

In order to study a significant number of processes in the ATLAS physics programme, the ability to distinguish which quark flavour initiated a particular reconstructed jet is required. To accomplish this, ATLAS employs 'flavour tagging' algorithms. These algorithms operate on the premise that jets initiated by heavier quarks which therefore have longer lifetimes, subsequently decay at Secondary Vertices (SV) which are located at a distance greater than the spatial resolution of the Inner-Detector, meaning that it is possible to reconstruct them. ATLAS flavour-tagging algorithms typically categorise jets into one of three categories - b-jets, c-jets or light flavour jets, the latter including jets that originate

from the hadronisation of u, d, s quarks and gluons.

ATLAS b-jet tagging algorithms ('b-taggers'), such as the DL1r algorithm used in the W+jets analysis are usually machine-learning based and generally involve two stages - low-level and high-level taggers [79]. Low-level taggers are algorithms designed to reconstruct the characteristic features of b-jets. They either exploit the larger impact parameters associated with b hadron decays, or explicitly reconstruct displaced secondary vertices.

The outputs of the low-level taggers are subsequently combined and used as inputs to deep-learning classifiers which return the likelihood that the jet originated from a particular quark flavour. The DL1r algorithm is a member of the fully-connected multi-layer feed-forward neural network DL1r series introduced at the beginning of Run 2, trained on 70% $t\bar{t}$ events and 30% $Z' \rightarrow q\bar{q}$ events [80]. DL1r returns a multidimensional output: The probabilities that a jet originates from a b, c or light jet. These probabilities are used to calculate the discriminant D_{DL1r} for labelling an individual jet as a b-jet:

$$D_{\mathrm{DL1r}} = \ln(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\mathrm{light}}}),$$

where p_b, p_c and p_{light} are the probabilities of a b, c or light flavour jet, and f_c being the background hypothesis effective c-jet fraction. As with all taggers, DL1r is calibrated to correct for data/MC differences - this is performed separately for each flavour of jet, in data, and is extrapolated p_{T} values beyond the coverage of the data using MC.

3.2.5 Missing Transverse Energy

In ATLAS pp interactions, all momentum of the initial colliding protons is aligned along the beam pipe, with zero initial momentum component in the transverse plane. Therefore, by conservation of momentum, all reconstructed objects the final state must have their vector $\mathbf{p}_{\mathrm{T}}$ sum to zero, with a non-zero sum indicating the presence of non-interacting particles in the event, for example SM neutrinos or weakly-interacting BSM particles. In ATLAS, this missing transverse energy is referred to interchangeably as $E_{\mathrm{T}}^{\mathrm{miss}}$ and calculated according to [81]:

$$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} = \underbrace{-\sum_{\substack{\text{selected}\\\text{electrons}}} \boldsymbol{p}_{\mathrm{T}}^{e} - \sum_{\substack{\text{selected}\\\text{photons}}} \boldsymbol{p}_{\mathrm{T}}^{\gamma} - \sum_{\substack{\text{selected}\\\text{taus}}} \boldsymbol{p}_{\mathrm{T}}^{\tau} - \sum_{\substack{\text{selected}\\\text{muons}}} \boldsymbol{p}_{\mathrm{T}}^{\mu} - \sum_{\substack{\text{selected}\\\text{jets}}} \boldsymbol{p}_{\mathrm{T}}^{\mathrm{jet}}}_{\text{hard term}} \underbrace{- \sum_{\substack{\text{unused}\\\text{tracks}}} \boldsymbol{p}_{\mathrm{T}}^{\mathrm{track}}}_{\text{soft term}} . \tag{3.2}$$

Each object in the 'hard term' in Eq. 3.2 is reconstructed and calibrated using the same procedure as those same objects reconstructed for use in physics analyses. The order of the objects included in the hard term is chosen deliberately, such that objects with better background rejection efficiency are reconstructed first. Lower priority objects which share calorimeter signals with higher priority objects are removed, so that no energy deposit is double-counted. Jets are also subject to pile-up removal and a dedicated overlap-removal procedure with muons, discussed in further detail in Section 4.4. Additionally, the same selections applied to objects in a given analysis should also be applied to the corresponding objects entering the Eq. 3.2 hard term.

The ‘unused tracks’ in the soft term refer to ID tracks originating from the primary vertex, but which have not been associated with or calibrated as any particular object. These signals must still be included, however, so that soft event activity is accounted for in the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

Large values of $E_{\mathrm{T}}^{\mathrm{miss}}$ in data are an important signature in a variety of new physics searches. Additional observables related to the $E_{\mathrm{T}}^{\mathrm{miss}}$ include: The scalar $E_{\mathrm{T}}^{\mathrm{miss}} = |\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}| = \sqrt{(E_x^{\mathrm{miss}})^2 + (E_y^{\mathrm{miss}})^2}$, which indicates the overall scale for the hardness of the event in the transverse plane, and $\phi^{\mathrm{miss}} = \arctan(E_y^{\mathrm{miss}}/E_x^{\mathrm{miss}})$, which corresponds to azimuthal angle of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the transverse plane.

Chapter 4

Pile-up tagging in the ATLAS central region

"So many bright lights to cast a shadow"

— My Chemical Romance, 'Famous Last Words'

4.1 An introduction to pile-up

At a hadron collider, partons in colliding protons other than those involved in the primary interaction create pile-up - collisions or energy deposits in the detector, that do not primarily originate from the same pp interaction as the event's primary hard-scatter.

There are two types of energy deposits that contribute to pile-up - QCD pile-up and stochastic pile-up. QCD pile-up is caused by additional single pp interactions, from either

the same or different bunch crossing as the hard-scatter, which are expected to conserve momentum in the ATLAS transverse plane. In contrast, stochastic pileup combines energy deposits from a combination of different sources and therefore momentum is not expected to be conserved. In general, QCD pile-up is more likely to occur at higher p_{T} , pseudorapidity and less harsh pile-up conditions. Jets formed from stochastic pile-up lack the distinctive dense energy core of jets that originate from a single pp interaction such as hard scatter or QCD pile-up jets. Instead energy is distributed more uniformly throughout the cone of stochastic pile-up jets. [82].

While the reconstruction and calibration of jets from the hard-scatter, as implemented in the ATLAS software, accounts for these additional energy deposits in a pile-up subtraction procedure, this 'jet-area subtraction' assumes a uniform pile-up density across the ATLAS calorimeter, and local fluctuations in pile-up density can cause additional pile-up jets to be reconstructed [76]. Figure 4.1 illustrates an event containing one hard-scatter jet, one QCD pile-up jet and one stochastic pile-up jet. Most of the deposits clustered into the QCD pile-up jet contain particles from a QCD pile-up interaction whereas the stochastic pile-up jet contains particles from several interactions without one dominant source. As illustrated in Figure 4.1, however, both types of pile-up jets can contain particles which are clustered into the jet from more than one source. Indeed, the binary classification of pile-up jets can be somewhat arbitrary, particularly in intense pile-up conditions.

Pile-up can be further classified on the basis of timing. In-time pile-up originates from collisions occurring in the same bunch crossing as the event's hard scatter, whereas out-of-time pile-up refers to energy deposits from previous or following bunch crossings taking

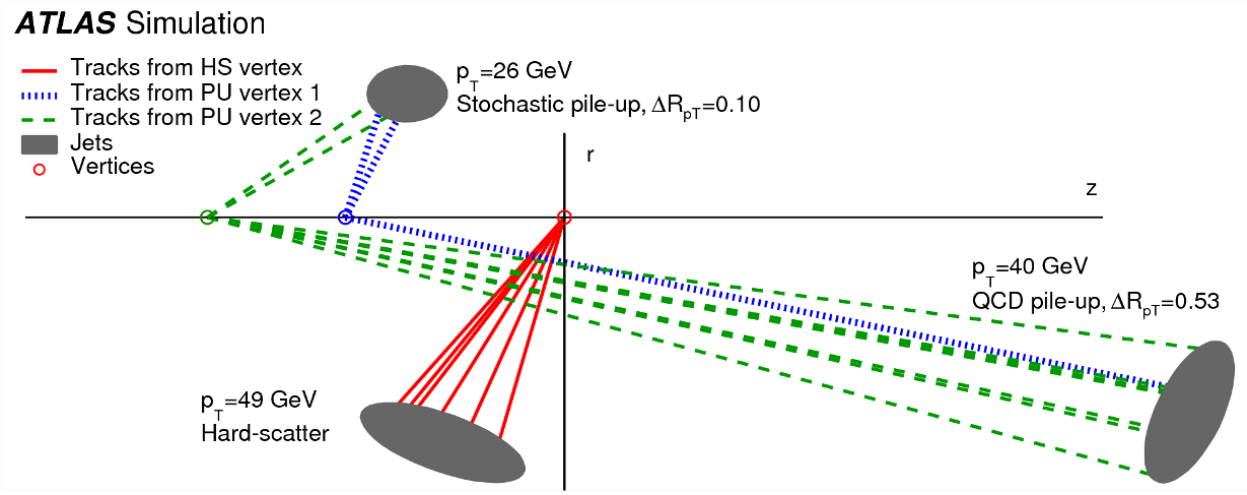


Figure 4.1: Diagram of simulated hard-scatter, QCD pile-up and stochastic pile-up jets in the ATLAS $r - z$ plane [82]. Also shown are the jets' transverse momentum p_{T} and, for the pile-up jets, difference in $R_{p\mathrm{T}}$ (defined in Section 4.3) from the primary vertex.

place 25 ns in the past or future respectively. As with the QCD / stochastic distinction above, this classification is not strictly binary as signals from multiple sources can be clustered into a single jet. In general, QCD pile-up jets are more likely to originate from in-time pile-up interactions while stochastic pile-up jets are more likely to come from out-of-time deposits.

Pile-up tagging, the identification and removal of pile-up jets, is a crucial aspect of the ATLAS jet reconstruction, and will become increasingly important as the level of pile-up increases over time. Figure 4.2 illustrates the time evolution of pile-up in ATLAS, quantified by the 'actual number of interactions per bunch crossing' μ, which is equivalent to the mean of the Poisson distribution of the number of interactions per crossing for each bunch and calculated from:

$$\mu = \frac{\mathcal{L}\sigma_{\mathrm{inel}}}{f_{\mathrm{r}} n_{\mathrm{b}}}$$

where $\mathcal{L}$ is the instantaneous luminosity, σ_{inel} is the inelastic cross-section, n_{b} is the number of colliding bunches and f_{r} is the LHC revolution frequency [83]. Figure 4.2 shows that the 'average number of interactions per bunch crossing' $\langle\mu\rangle$, determined by integrating μ over all colliding bunches, is increasing over time, from $\langle\mu\rangle = 33.7$ during Run 2 to $\langle\mu\rangle = 44.6$ during Run 3 up to 5th June 2023. These levels are expected to further increase into the future, with a $\langle\mu\rangle \approx 65$ planned for the remaining high-intensity pp runs of Run 3 and an expected mean number of interactions per bunch-crossing of at least $\langle\mu\rangle \approx 130$ planned for the High-Luminosity LHC upgrade [84].

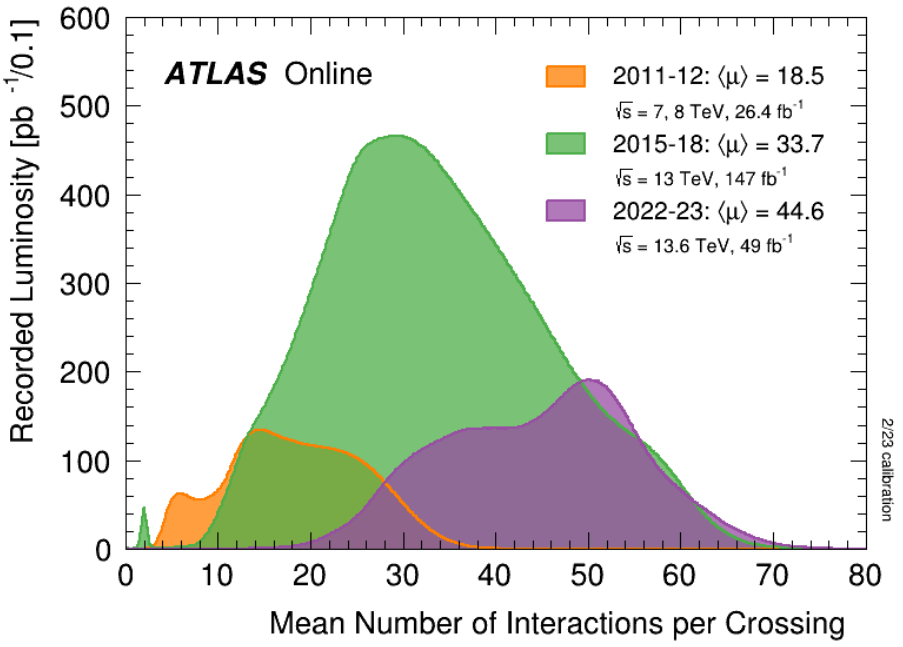


Figure 4.2: The evolution over time of the recorded luminosity of pile-up in collisions delivered to ATLAS, quantified by the mean number of interactions per bunch crossing. Orange indicates collisions collected during the 2011-2012 years of Run 1, green indicates luminosity recorded during Run 2 and purple corresponds to Run 3 up to 5th June 2023 [85].

4.1.1 A brief history of pile-up tagging in ATLAS

Several variables have been implemented by ATLAS for the purposes of pile-up tagging over the course of LHC Run 1 and Run 2, and are described in detail in [83]. Jet Vertex Fraction (JVF) may be considered the most straightforward of such variables. Defined for each jet with respect to each primary vertex (both hard-scatter and QCD pile-up) in the event, the JVF variable is used to quantify the likelihood that jet$_i$ originated from primary vertex PV$_j$ by calculating the ratio of the scalar p_{T} sum of all tracks matched to jet$_i$ that originate from PV$_j$ to the total scalar p_{T} sum of all tracks matched to jet$_i$:

$$\mathrm{JVF}(\mathrm{jet}_i, \mathrm{PV}_j) = \frac{\sum_m p_{\mathrm{T}}(\mathrm{track}_m^{\mathrm{jet}_i}, \mathrm{PV}_j)}{\sum_n \sum_l p_{\mathrm{T}}(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_n))}, \tag{4.1}$$

where the index m runs over all tracks matched to jet$_i$ originating from PV$_j$, n runs over all primary vertices in the event and l runs over all tracks matched to jet$_i$ originating from PV$_n$.

However, while JVF selections can be an effective [83] method for rejecting pile-up in moderate luminosity conditions, Eq. 4.1 is dependent on the number of primary vertices in an event (N_{PV}), introducing limitations in intense pile-up conditions. To correct for this, two new variables were defined for use as inputs to a new pile-up tagger.

Firstly, corrJVF, a variable similar to JVF with the N_{PV} dependence corrected, is defined for each jet:

$$\mathrm{corrJVF}(\mathrm{jet}_i) = \frac{\sum_m p_{\mathrm{T}}(\mathrm{track}_m^{\mathrm{jet}_i}, \mathrm{PV}_0)}{\sum_l p_{\mathrm{T}}(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_0) + \frac{\sum_{n\geq 1}\sum_l p_{\mathrm{T}}(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_n)}{k \cdot n_{\mathrm{track}}^{\mathrm{PU}}}}, \tag{4.2}$$

where PV_0 corresponds to the event's primary hard-scatter vertex. The term $k \cdot n_{\mathrm{track}}^{\mathrm{PU}}$ is a correction for the increase the average scalar transverse momentum originating from pile-up activity with increasing number of pile-up tracks per event $n_{\mathrm{track}}^{\mathrm{PU}}$, where k is a parameter to be determined.

Secondly, a new variable $R_{p\mathrm{T}}$ is defined which depends only on the hard-scatter vertex. $R_{p\mathrm{T}}$ is the ratio of the scalar sum of tracks associated to jet_i that originate from the hard-scatter primary vertex, to the fully calibrated jet p_{T} :

$$R_{p\mathrm{T}}(\mathrm{jet}_i) = \frac{\sum_m p_{\mathrm{T}}(\mathrm{track}_m^{\mathrm{jet}_i}, \mathrm{PV}_0)}{p_{\mathrm{T}}^{\mathrm{jet}_i}}. \tag{4.3}$$

The distribution of $R_{p\mathrm{T}}$ for pile-up jets peaks at 0 then falls sharply, since comparatively few tracks from the hard-scatter primary vertex are expected to contribute to the reconstruction of pile-up jets.

Finally, the Jet-Vertex-Tagger (JVT) discriminant is constructed using corrJVF and $R_{p\mathrm{T}}$ as inputs to a k-nearest neighbour algorithm [86], with discriminant values close to 1 indicating a strong likelihood a jet originates from hard-scatter and values close to 0 indicating a jet may be pile-up. During the LHC Run 2, the JVT discriminant was the default method of pile-up tagging in the ATLAS central region ($|\eta| < 2.5$) [83]. However, at pseudorapidities greater than 2.5, which fall outside the coverage of the ATLAS inner-detector (the ATLAS forward region), the tracking information used by the JVT tagger is no longer available. In

this region, alternative methods must be used to identify and remove pile-up jets [82], an overview of the principles of which will be described in Section 4.2.

4.2 Momentum balancing for pile-up tagging

The momentum balance pile-up tagging method for QCD pile-up jets exploits topological information; the conservation of momentum in the ATLAS transverse plane at all interaction vertices. While the initial longitudinal momentum of the beam is unknown, its initial transverse momentum is zero. This implies that, at any interaction vertex, the vector sum of all transverse momentum produced by the collision linked to this vertex should be zero. One prominent example of the use of this idea in ATLAS event reconstruction is the calculation of the Missing Transverse Energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) [81], but the same principle can also be applied to the pp pile-up vertices which produce QCD pile-up jets.

The vector sum of transverse momentum linked to individual pile-up vertices can be used to ascertain whether a given jet more likely originates from the event's hard scatter interaction vertex or a pile-up vertex. If the transverse momentum of the jet is approximately (allowing for limitations such as the detector's jet energy and spatial resolution and the lack of a strictly binary distinction between QCD and stochastic pile-up jets) equal and opposite to the vector summed transverse momentum of one of the event's pile-up vertices, that jet likely originated from that vertex. This jet can then be identified ('tagged') as a pile-up jet, and removed from the event.

This principle has been employed in the forward region of the detector from Run 2 in the Forward Jet Vertex Tagger (fJVT) [82], as no techniques based on the jet's tracks can be used. This technique can also be impactful in the central region, as the topological information on which it relies is complementary to the track-based inputs to the default JVT tagger used in that region. However, in Run 2 this was not employed. The remainder of this chapter will describe the operation and performance of a pile-up tagging tool using jet balancing techniques in the ATLAS central region, and plans for its implementation in Run 3.

4.3 The Balance Vertex Tagger tool

To this end, a tool was developed that employs the balance technique for pile-up tagging in the central region. Two versions of the 'Balance Jet Vertex Tagger' (bJVT) tool will be discussed in the following sections - one 'preliminary' version used for development and initial performance studies, and a 'final' version which has been integrated into the central ATLAS 'Athena' software framework. The basic operation of the 'final' tool integrated into Athena is as follows:

- For all pile-up vertices in an event, the associated Particle Flow jets (as described in Section 3.2.3) are rebuilt. This is a necessary step to correctly capture the event's pile-up jets, as the Particle Flow algorithm only clusters tracks and topoclusters associated to a particular vertex into jets (by default, the primary hard-scatter vertex).

- For each pile-up vertex, the vertex's missing momentum is reconstructed by calculating the inverse vector sum of the transverse momentum of the jets associated to that vertex by the PFlow algorithm. Jets included in the sum must pass a minimum p_{T} requirement of 20 GeV corresponding to the minimum p_{T} to which PFlow jets are calibrated, be in the ATLAS central region, and not have any close-by hard-scatter jets.

  In addition, the compatibility between the p_{T} of potential jets to be included in the sum and the p_{T} of tracks associated to the vertex is considered. This is quantified by the generalised version of the $R_{p\mathrm{T}}$ definition given in Eq. 4.3 with respect to any primary vertex in the event:

$$R^{i}_{p\mathrm{T}} = \frac{\Sigma p_{\mathrm{T}}^{\mathrm{trk}}(\mathrm{PV}^{i})}{p_{\mathrm{T}}^{\mathrm{jet}}}, \tag{4.4}$$

  where the index i refers to the jet under consideration, trk stands for tracks, PV^{i} is jet i's primary vertex and $p_{\mathrm{T}}^{\mathrm{jet}}$ is the fully calibrated jet transverse momentum, including pile-up subtraction.

  For jets originating from a single pp hard scatter, the average value of $R_{p\mathrm{T}}$ with respect to that hard-scatter vertex is approximately 0.5 [82]. Jets that pass a minimum $R_{p\mathrm{T}}$ threshold of 0.1 have their total transverse momentum included in the vertex's missing momentum. For jets that fail to pass the threshold, primarily stochastic pile-up jets, only the charged component of their transverse momentum is included in the vertex missing momentum as this is the component of their p_{T} which can be reliably associated to the vertex.

- For each central jet in the event exceeding a minimum p_{T} threshold, its vector transverse momentum is projected onto the set of pile-up vertex missing momenta. The maximum value of the set of projections, normalised to the central jet's p_{T} , is returned as the bJVT tagger's discriminant. Larger bJVT discriminant values indicate a greater degree of balancing between the central jet and a pile-up vertex, and therefore a greater likelihood the central jet originates from pile-up. If there are no valid pile-up vertices in the event, or if a central jet has $p_{\mathrm{T}} < 20$ GeV and is therefore too soft to meet the minimum threshold for PFlow jet calibration, a value of 0 is returned instead.

The scope of a performance study on this version of the bJVT tool as implemented in the ATLAS internal Athena software framework has been defined and is described in Section 4.5 with initial samples being prepared at the time of writing. However, initial studies have been completed for the earlier 'preliminary' version and will be presented in Section 4.4. The differences between the bJVT tool as described above and the preliminary version are as follows:

- The preliminary version of the bJVT tool operated on jets reconstructed using calorimeter clusters only, instead of PFlow jets.
- The preliminary version of the tool used JVT and JVF selections to assign jets to QCD pile-up vertices, instead of rebuilding the jets originating from the vertices using the PFlow algorithm.
- JVT and JVF preselections were also applied to the central jets to which the tagger

was applied, in addition to the minimum threshold for jet p_{T} .

4.4 Performance of the preliminary bJVT tool

The performance of the preliminary version of the bJVT tool was assessed using a sample of 50,000 $Z \to \mu\mu$ events, a topology expected to have only small amounts of true missing transverse energy, generated using Powheg-Box [87] and Pythia 8.2 [88] for the parton shower with the AZNLO tune [89] with the CTEQ6L1 PDF set at NLO in QCD [90]. This sample included collections of in-time and out-of-time pile-up at truth level, which were incorporated by overlaying simulated minimum bias inelastic interactions generated using Pythia 8.2 [88] with the A3 tune [91] and the NNPDF2.3 PDF [92] set at LO in QCD.

To establish the ability of the bJVT tool to identify QCD pile-up jets, a sample of jets with a high concentration of pile-up jets must first be identified. The sample selected was the set of jets included in the missing transverse energy calculation in events with less than 1 GeV of truth $E_{\mathrm{T}}^{\mathrm{miss}}$ but 50 GeV of reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$. This collection of jets is slightly different from the set of 'standard' reconstructed jets, as an additional jet-muon overlap removal procedure is applied during $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. This procedure [1] removes jets that have significant overlap with a muon to prevent double-counting momentum in the case where a muon loses sufficient energy in the calorimeter to create deposits that are reconstructed as a jet [81]. JVT is also applied to jets included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ re-

[1] Overlap removal procedure: A jet is removed where, $p_{\mathrm{T,track}}^{\mu}/p_{\mathrm{T,track}}^{\mathrm{jet}} > 0.8$, $p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T,track}}^{\mu} > 2$, $N_{\mathrm{track}}^{PV} < 5$ where $p_{\mathrm{T,track}}^{\mu}$ is the p_{T} of the ID track associated with the muon, $p_{\mathrm{T,track}}^{\mathrm{jet}}$ is the sum of the p_{T} of all tracks associated to the jet, $p_{\mathrm{T}}^{\mathrm{jet}}$ is the fully calibrated jet p_{T} and N_{track}^{PV} are the number of tracks associated from the event's primary hard-scatter vertex.

construction, meaning that any central QCD pile-up jets present in this sample have been missed by the default tagger.

Because the ATLAS $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction depends on correctly identifying all visible objects from the event hard scatter, the incorrect inclusion of pile-up jets can lead to $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction inefficiencies such as those selected for here. Indeed, when a ΔR jet truth matching procedure, inspired by similar performance studies for the fJVT tagger [82]), was applied between truth hard-scatter and pile-up jets and the reconstructed jets included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation in these events, Figure 4.3 shows that QCD pile-up jets are a large fraction of the jets included in these events' $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The matching procedure used is described below:

1. Reconstructed jets are labelled as 'hard-scatter' if a truth hard-scatter jet with $p_{\mathrm{T}} >$ 10 GeV is located within $\Delta R <$0.3.

2. If no truth **hard-scatter** jet with $p_{\mathrm{T}} >$ 4 GeV is found within $\Delta R <$0.6, a reconstructed jet is determined to originate from pile-up. It can then be further classified by type of pile-up:

    - The reconstructed jet is labelled as QCD pile-up if a truth pile-up jet with $p_{\mathrm{T}} >$ 10 GeV is found within $\Delta R <$0.3. The QCD pile-up jets are subsequently further classified as **in-time QCD pile-up** of **out-of-time QCD pile-up** depending on the collection of truth pile-up jets to which they were matched.
    - The reconstructed jet is labelled as **stochastic pile-up** if no truth pile-up jet with $p_{\mathrm{T}} >$ 10 GeV is found within $\Delta R <$0.6.

- If no truth pile-up jet of $p_{\mathrm{T}} >$ 10 GeV is found within 0.3 $< \Delta R <$0.6, the reconstructed pile-up jet is not further classified and simply labelled '**ambiguous pile-up**'.

3. For reconstructed jets with $p_{\mathrm{T}} >$ 10 GeV within 0.3 $< \Delta R <$0.6 of a truth hard-scatter jet, and reconstructed jets with 4 GeV $< p_{\mathrm{T}} <$ 10 GeV within $\Delta R <$0.3 of a truth hard-scatter jet, it is determined that their origin cannot be unambiguously established. These jets are labelled as **ambiguous** jets.

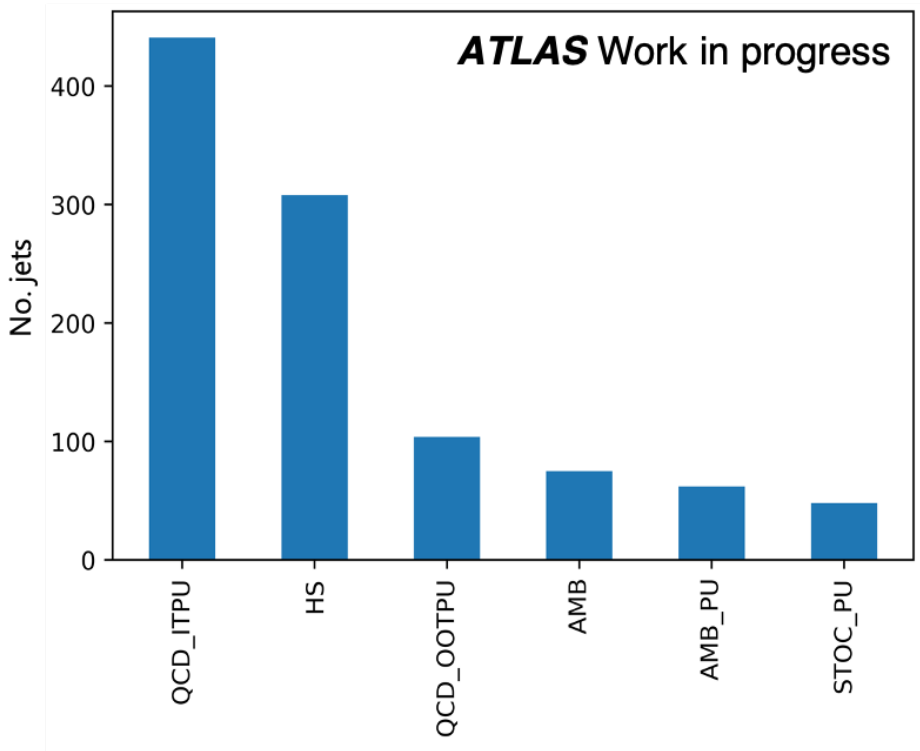


Figure 4.3: The truth origins of jets included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction of events in the sample described in Section 4.4, as determined by the ΔR jet matching scheme described in [82]. The meanings of the acronyms on the x-axis are as follows: $\mathrm{QCD\_ITPU}$: In-time QCD pile-up, HS: Hard-scatter, $\mathrm{QCD\_OOTPU}$: Out-of-time QCD pile-up, AMB: Ambiguous whether jet originates from hard-scatter or pile-up, $\mathrm{AMB\_PU}$: Ambiguous whether jet originates from QCD pile-up or stochastic pile-up, STOCH: Stochastic pile-up.

Figure 4.4 shows a ROC curve (a plot which demonstrates the performance of a binary

classifier by plotting the trade-off between its true positive rate and its false positive rate) for the preliminary version of the bJVT tool applied to this collection of jets. In this plot, the missing transverse momentum of the primary hard-scatter vertex was also calculated in the same manner as the pile-up vertices. A jet included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is classified as pile-up if the largest normalised projection of the jet's p_{T} onto the set of hard-scatter and pile-up vertex missing momenta is a pile-up vertex. The truth type of the jet is established using the same ΔR matching procedure as mentioned above.

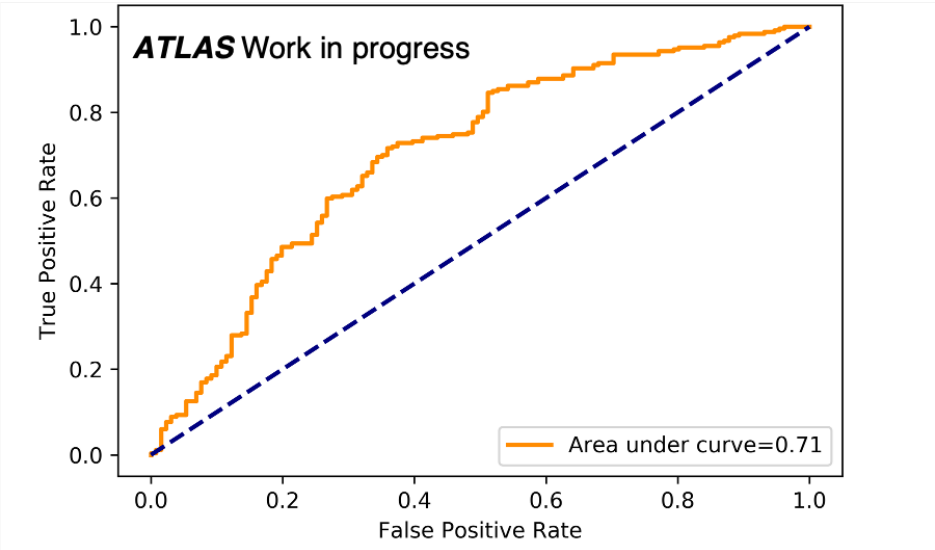


Figure 4.4: ROC curve of the preliminary bJVT tool used as a binary classifier to distinguish hard-scatter jets from QCD pile-up jets.

Figure 4.4 shows that the tagger performs above random (indicated by an area under the curve greater than 0.5) at classifying jets in this fashion. This demonstrates that the tool is effective at identifying QCD pile-up jets which have been missed by the JVT tagger, even in samples where other types of pile-up is present, although the impact of other classes of pile-up jets is not expected to be large given the composition of the sample as shown in Figure 4.3. This is an important feature, since, as discussed in Section

4.1, the distinction between QCD and stochastic pile-up can at times be subjective. The composition of the sample used in Figure 4.4 also presents an insight on how ATLAS may further refine the momentum balancing pile-up tagging method, as some inefficiencies in the ROC curve may be due to insufficiently stringent stochastic pile-up jet rejection by the bJVT tool.

Figure 4.5 shows the distribution of bJVT scores for jets included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, classified by their truth type as designated by the ΔR matching scheme. The separation in scores between hard-scatter and pile-up jets indicates that the bJVT discriminant is sensitive to the differences between hard-scatter and pile-up jets.

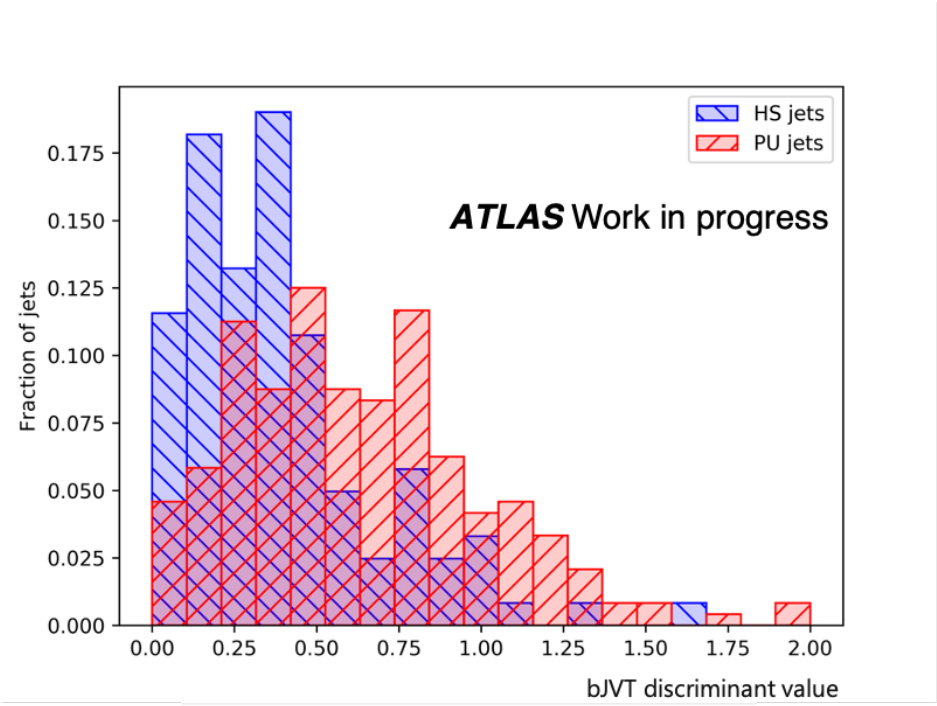


Figure 4.5: Distribution of bJVT scores for hard scatter jets (blue) and pile-up jets (red).

Finally, the missing transverse energy for the events in this sample was recalculated, rejecting jets with bJVT scores above a threshold of 0.7. This initial threshold was selected

to match that used by the 'Loose' working-point of the fJVT tagger, however, the study outlined in Section 4.5 will establish working-points for the bJVT tool and their associated thresholds. Figure 4.6 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution before and after the recalculation. With the application of the preliminary bJVT tagger, the amount of misreconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ in the sample decreased substantially.

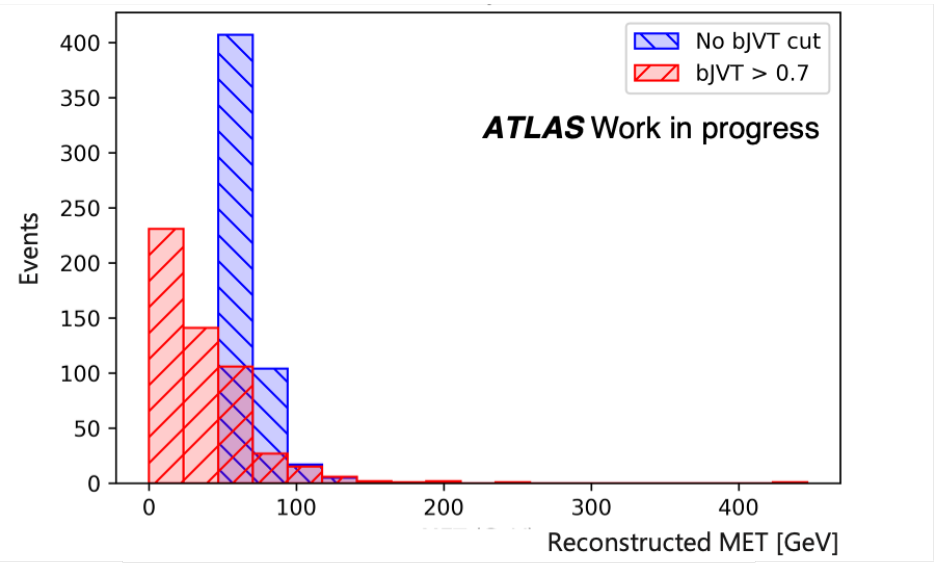


Figure 4.6: Reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ for the sample of events described in Section 4.4 (blue) and the sample events where the $E_{\mathrm{T}}^{\mathrm{miss}}$ has been recalculated with an upper threshold of bJVT = 0.7 for jets included in the reconstruction (red).

A small number of events had an increased amount of misreconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ after the application of the tool. Proposed solutions include imposing an upper p_{T} limit on jets to which the bJVT tool is applied, as is currently done for JVT and fJVT, or a cut on the $E_{\mathrm{T}}^{\mathrm{miss}}$ significance when the tool is used for $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. $E_{\mathrm{T}}^{\mathrm{miss}}$ significance $\mathcal{S}$ quantifies the level of belief that the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructed is real. The approximation of $\mathcal{S}$ currently most commonly used in ATLAS is the object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ significance [93]:

$$\mathcal{S}(E_{\mathrm{T}}^{\mathrm{miss}}) = \frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sqrt{\sigma_L^2(1-\rho_{\mathrm{LT}}^2)}},$$

where σ_L is the longitudinal component of the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and ρ_{LT}^2 is the correlation between the longitudinal and transverse components of the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution.

4.5 Outlook

This chapter has presented the first application of the momentum balancing pile-up tagging technique in ATLAS outside of the forward region. That it would also be effective in the central region of ATLAS is non-trivial, given the different detector systems and physics activity present in the two regions. Additionally, it is not a given that applying additional topologically-based pile-up tagging would provide any increase in pile-up tagging performance in the central region, given the highly efficient track-based JVT algorithm already in use.

In this chapter, it has been shown for the first time in ATLAS that a large proportion of central jets in events with large amounts of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ originate from QCD pile-up, and have been missed by the default JVT tagger. Additionally, it has been shown for the first time that a momentum balancing technique, as implemented in the bJVT tool, is effective at identifying such jets, even with additional types of pile-up present, and that the discriminant produced by this tool provides separation between hard-scatter and pile-up jets. Finally, it has been proven that the application of this technique can successfully reduce the amount of misreconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ in events in the central region with large

amounts of fake $E_{\mathrm{T}}^{\mathrm{miss}}$.

This first investigation of the momentum balancing pile-up tagging technique in the ATLAS central region provides a new avenue for performance improvements in both pile-up tagging and $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, with these studies having already spawned new investigations in the relevant ATLAS performance group. Studies are now ongoing in the group to reoptimise the tool for new ATLAS software releases, measure the technique's efficiency in new topologies, and to incorporate the tool into a combined central pile-up tagger through the inclusion of the bJVT discriminant as an input variable to the neural-network based central pile-up tagger in use for LHC Run 3.

Chapter 5

W+jets analysis introduction

"It's a dangerous business, Frodo, going out your door. You step onto the road, and if you don't keep your feet, there's no knowing where you might be swept off to."

— J.R.R Tolkien, 'Lord of the Rings'

The goal of this analysis is to measure the cross-sections for W boson production in association with jets for the first time at $\sqrt{s}$=13 TeV in ATLAS. It aims to measure the cross-sections differentially in variables sensitive to the QCD aspects of Monte Carlo (MC) modelling, while also providing PDF constraining information, for charge independent W production, separately for W^+ and W^- charged bosons, and for different jet multiplicities, to compare against the current state-of-the-art SM predictions to investigate regimes where improved modelling is required.

ATLAS has previously published two measurements of W+jets production at lower centre-of-mass energies. The 7 TeV ATLAS W+jets measurement [94], published in 2015,

used 4.6 fb^{-1} of data to produce differential cross-sections in the $W \to e\nu$ decay channel, the $W \to \mu\nu$ decay channel and a combination of the two channels. The 8 TeV ATLAS W+jets publication [21] utilised 20.2 fb^{-1} of data to measure W+jets production differential cross-sections in the $W \to e\nu$ channel only.

The CMS experiment has produced two analyses involving W boson production with associated jets at 13 TeV - a measurement of W + inclusive jets production cross-sections using 2.2 fb^{-1} of LHC Run 2 data [95], and a measurement of W + dijet production with an integrated luminosity of 32.9 fb^{-1} [96]. The present analysis will include 36.2 fb^{-1} of data collected by ATLAS in 2015 and 2016 during LHC Run 2, which will be the most luminosity of any W + inclusive jets analysis.

This W+jets measurement aims to produce unfolded differential cross-sections for two decay channels of the W boson: $W \to e\nu$ and $W \to \mu\nu$, in addition to a combination of the measured unfolded cross-sections in each channel into a $W \to l\nu$ channel. The author's research and this document's focus is the $W \to \mu\nu$ channel, henceforth referred to as the 'muon channel'. This will be the first W+jets muon channel cross-section measurement in ATLAS since the 7 TeV analysis in 2015 [94]. At the most basic level, this process is identified through a single, high p_{T} muon, missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) indicating the presence of the non-interacting neutrino, and at least one high p_{T} jet. The charge of the W is identified through the charge of its decay muon, for the purposes of the charge-dependent distributions.

5.1 Distributions to be produced

In this measurement, differential-cross section distributions, both W charge-independent and charge-dependent, are produced as functions of jet multiplicity N_{jet} , in addition to variables sensitive to higher-order terms in the theoretical predictions and the PDFs:

- The transverse momentum of the event's most energetic ('leading') jet, $p_{\text{T}}^{\text{leading}}$;
- The rapidity of the leading jet, y^{leading} ;
- The scalar sum of the transverse momentum of all jets in the event, $S_{\text{T}} = \sum_i p_{\text{t}}^{i,\text{jet}}$;
- The scalar sum of the transverse momentum of all jets in the event, the muon and the neutrino, $H_{\text{T}} = \sum_i p_{\text{t}}^{i,\text{jet}} + p_{\text{T}}^{\mu} + E_{\text{T}}^{\text{miss}}$;
- The transverse momentum of the W boson, p_{T}^{W} . This is defined with respect to the W decay products as $p_{\text{T}}^{W} = \sqrt{(p_{\text{T}}^{\mu})^2 + (E_{\text{T}}^{\text{miss}})^2 + 2p_{\text{T}}^{\mu}E_{\text{T}}^{\text{miss}}\cos(\Delta\phi(p_{\text{T}}^{\mu}, E_{\text{T}}^{\text{miss}}))}$.

A binning scheme is chosen for these distributions such that granularity is maximised while bin-to-bin migrations and statistical uncertainties are kept at a minimum. The metrics 'purity' and 'stability' can be used to quantify the degree of migrations between bins as truth-level interactions are reconstructed by the ATLAS detector. The purity ($\mathcal{P}$) of bin i is defined as:

$$\mathcal{P}_i = \frac{N(T_i \cup R_i)}{N(R_i)},$$

while its counterpart stability ($\mathcal{S}$) is defined as:

$$\mathcal{S}_i = \frac{N(T_i \cup R_i)}{N(T_i)},$$

where $N(R_i)$ is the number of events reconstructed in bin i, $N(T_i)$ is the number of events generated in bin i and $N(T_i \cup R_i)$ is the number of events both generated and reconstructed in bin i.

To evaluate the trade-off between granularity and migrations, the following scheme is applied to the $p_{\mathrm{T}}^{\mathrm{leading}}$, S_{T} , H_{T} and p_{T}^{W} distributions:

1. An initial choice of a 10 GeV width for all bins was made, corresponding to approximately 1.4σ of the resolution for ParticleFlow jets with a p_{T} of 30$-$40 GeV [71].

2. Bins with purity varying more than 20$-$30% from the surrounding bins are rebinned.

3. Bins for which the statistical uncertainty was $\geq$20% were then also rebinned, as this impedes the ability to easily identify the QCD effects this analysis aims to test.

4. Finally, small optimisations were made by hand in regions where significant binning artefacts in purity or statistical uncertainty remained.

For the y^{leading} distribution, a bin width of 0.2 was chosen .

The resulting binning scheme is the one implemented in all Figures presented for this analysis [1].

The purity, stability and statistical uncertainties in each distribution as a result of this

[1] With the exception of the efficiencies used to model the multijet background; this will be discussed in further detail in Section 6.2.3

scheme, measured using the Sherpa version 2.2.11 [97] $W \to \mu\nu$ MC signal sample (described in Section 5.2 with the truth and reconstruction level analysis selections, described in 5.3 and Section 7.2 respectively, are shown in Figure 5.1.

5.2 Signal MC samples

After unfolding the measurement, the distributions described in Section 5.1 in data are compared to truth-level $W \to \mu\nu$ cross-section predictions created by two different state-of-the-art Monte Carlos, with the aim of identifying regions of phase space where the generators describe the data well, and regions where they do not. In addition, these signal samples serve various other purposes throughout the analysis such as correcting for detector effects in unfolding, which will be described in following chapters. A thorough description of the generator set-ups employed to generate these signal samples will be presented later in this section, following a brief summary of the MC simulation in ATLAS.

ATLAS uses Monte Carlo simulations to produce samples of events based on a calculated cross-section prediction for a given process. The cross-sections for hard-scatter interactions are modelled in a fixed-order matrix element calculation, which can be expressed by perturbation theory as a sum in increasing powers of the strong coupling constant α_{S} , and include Born level predictions with higher order real and virtual corrections. Partonic-level calculations are convolved with a particular set of Parton Distribution Functions (PDFs) to give particle-level predictions according to the factorisation theorem described in Chapter 1.

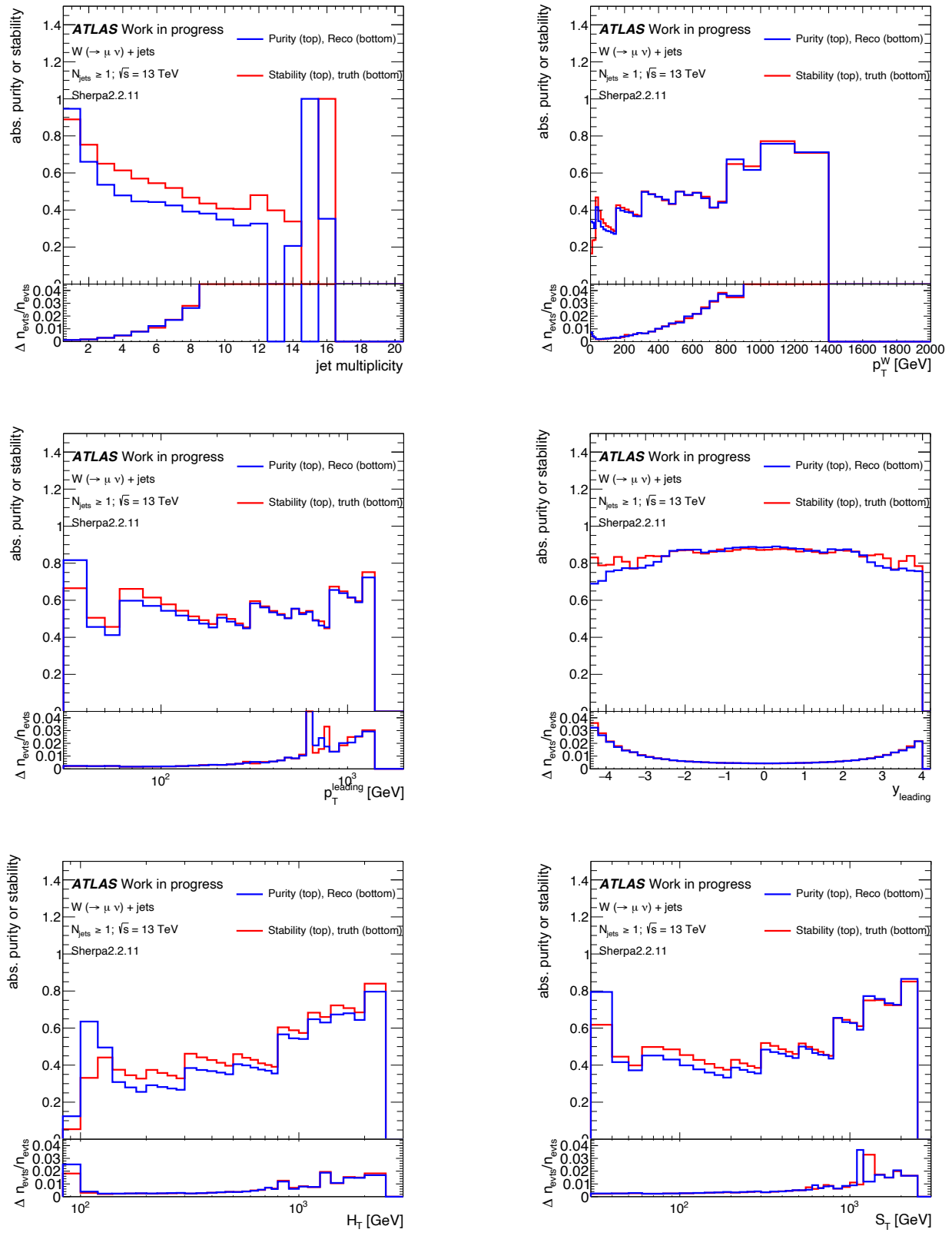


Figure 5.1: Top panels: Bin purity (blue) and stability (red) of the final binnings of the distributions to be unfolded, calculated using the Sherpa 2.2.11 $W \to \mu\nu$ signal MC sample. Bottom panels: Relative statistical uncertainty of the distributions at reconstruction level (blue) truth and level (red).

Non-perturbative soft and collinear emissions such as Initial State Radiation (ISR) and Final State Radiation (FSR) are modelled during the parton shower stage, where hard-scatter partons are “dressed” with softer showers of algorithmically generated emissions up to a certain Q^2 cutoff. To avoid double counting particles in the transition between hard and soft regimes, generators will employ one of a range of matching, the subtraction of a fixed-order parton shower calculation from a higher-order matrix element calculation to prevent double-counting, or merging, the combination of individual calculations of different parton multiplicities of interest, schemes. After this, physically observable hadrons are created from the generated partons during the hadronisation step, where algorithms are typically derived from phenomenological models as the hadronisation occurs in a non-perturbative regime. Examples include the Lund string fragmentation model implemented in Pythia [98] and the Sherpa implementation of the cluster hadronisation model [99]. Finally, the underlying event is simulated by incorporating models for processes including multi-parton interactions and effects of the beam remnants.

Once a ‘truth-level’ event has been simulated, the effects of the ATLAS detector must also be modelled to create a ‘reconstruction-level’ simulated event which can be compared to reconstructed data. The first step is a simulation of the interaction of the generated particles with detector material. In the following digitisation stage, the detector response to the particles is computed, and the effects of pile-up are overlaid. Finally, in the reconstruction stage, the events are passed through the same reconstruction chain as events in data, including a simulation of the ATLAS trigger and reconstruction algorithms.

Two multi-leg $W \to \mu\nu$ signal MC generators are considered for comparison to data.

The generator setups are described in detail in [100], and are summarised below:

- **Sherpa v2.2.11** [97] simulates the full event at next-to-leading-order (NLO) up to two jets and a further three jets at leading-order (LO) for a maximum of five jets calculated in the matrix element using Comix [101] and the Hessian NNPDF3.0NNLO [102] [103] PDF set. Virtual QCD corrections to the matrix element at NLO are calculated using OpenLoops [104][105][106][107]. This calculation uses the 5 flavour number scheme, with b and c quarks treated as massless in the matrix element and massive in the parton shower. QED radiation is included in the simulation using the Sherpa implementation of the Yennie-Frautschi-Suura algorithm [108][109]. Additional jet multiplicities are added using the parton shower which employs the p_{T}-ordered Cantani-Seymour dipole parton shower scheme [110], and matched to the matrix element using a colour-exact version of the MC@NLO algorithm [111] with a cutoff Q_{cut}=20 GeV. Events with different jet multiplicities are merged into an inclusive sample using the CKKW matching procedure [112] [113], extended to NLO accuracy using the MEPS@NLO prescription [114]. Hadronisation is provided using the cluster hadronisation model [99]. This sample will subsequently be referred to as 'Sherpa 2.2.11' for the remainder of the document.

- The second generator is **MadGraph5_aMC@NLO v2.6.5** [115] for matrix element calculation at NLO up to three jets using the MadLoop [116] module and the NNPDF3.1NNLO PDF set [117] supplemented with the LUXqed photon PDFs [118] with α_{S} =0.118. The p_{T}-ordered parton shower, and hadronisation based on Lund string fragmentation [98], are subsequently performed by Pythia v8.240 [77] using

the A14 [119] set of tuned parameters and the NNPDF2.3LO PDF set [92] with α_{S} =0.130. QED radiation is also added by Pythia 8. Events with different jet multiplicities are merged into an inclusive sample using the FxFx NLO merging procedure [120], with the merging scale, Q_{cut}=20 GeV. This calculation employs the 5 flavour number scheme, with b and c quarks treated as massless in the matrix element and massive in the shower. This sample will subsequently be referred to as 'MGPy8_FxFx' for the remainder of this document.

Both signal samples are normalised to the NNLO generator cross-sections. Similar descriptions of the MC samples used to model background processes are provided in Section 6.1.

5.3 Event selection

Object and event selections are applied to both data and MC samples in the Signal Region with the intent to increase signal yield, while also reducing background contamination as much as possible. In this section, these selections will be summarised. The reconstruction procedure for the objects discussed in this section has previously been described in Chapter 3.

The preselections applied to data entering the W+jets analysis chain are summarised in Table 5.1. The data analysed was collected by the ATLAS detector during the first two years of the LHC Run 2 (2015, 2016). Data events are required to pass a number of

selections recommended by ATLAS performance groups. The event must be considered 'Good for physics' i.e. it exists on ATLAS's 'Good Runs List', which stores 36.2 fb^{-1} of good quality data from the 2015/16 period. It is also required that there are no noise bursts in the ATLAS Liquid Argon and Tile calorimeters, and that the event's primary vertex is well reconstructed.

Data preselections	
Good for physics	Event on 'Good Runs List'
Data quality	No calorimeter noise bursts
Trigger	Event passes unprescaled single lepton triggers

Table 5.1: Preselections applied to data in the Signal Region.

The triggers used in the Signal Region of this analysis, broken down by year as the ATLAS trigger menu evolves over time, is as follows:

- **2015:** `mu20_iloose_L1MU15` or `mu50`,
- **2016:** `mu26_ivarmedium` or `mu50`.

The trigger efficiencies as a function of muon p_{T} for passing one of the two 2016 triggers as measured in W+jets events in a subset of 2016, 2017 and 2018 data are shown in Figure 5.2 [121]. The trigger naming convention for all chains above is laid out below.

- The `mu` label refers to the type of trigger object, in this case a single muon.
- The following number, `20`, `26` or `50` indicates the p_{T} threshold in GeV for objects selected by the trigger.

- Any following labels beginning with 'i', `iloose` or `ivarmedium`, indicate isolation requirements placed on the object.

- Finally, any trailing labels beginning with 'L1' indicate any instance where the p_{T} threshold for the Level-1 seed of the high-level trigger differs from the default, which is `L1MU20` in the case of single-muon triggers in Run 2.

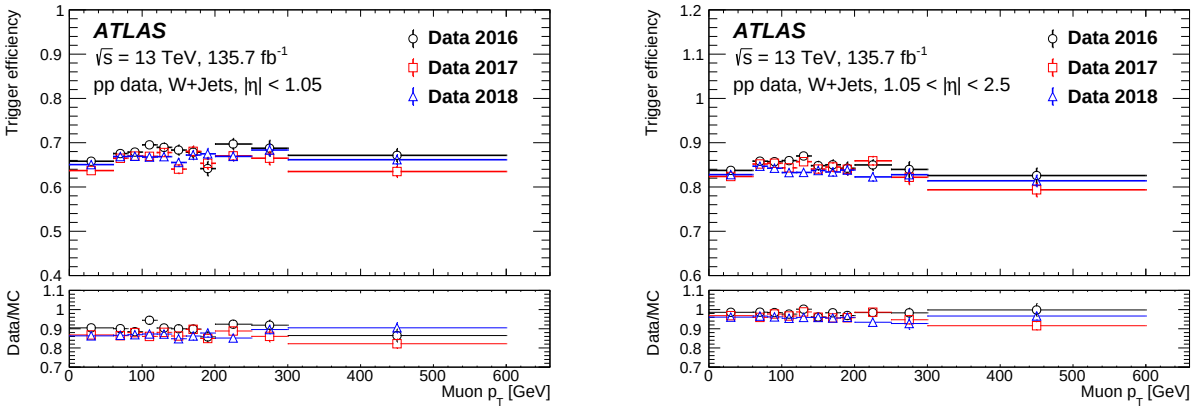


Figure 5.2: The trigger efficiencies as a function of muon p_{T} for the triggers applied to 2016 data in this analysis, `mu26_ivarmedium` or `mu50`, as measured in W+jets events in a subset of 2016, 2017 and 2018 data [121].

All of the triggers above are 'unprescaled', i.e. any event passing trigger thresholds is saved, subject to limitations such as detector dead-time. In addition to the triggers listed above, an alternative set of triggers with no isolation requirement is used to save events with 'Loose' muons, which will be described below and are required for the estimation of the multi-jet background. These triggers are 'prescaled', meaning only a specified fraction of events passing the thresholds are saved, usually in order to control the trigger rates. The prescaled triggers used for the selection of loose muons are given below:

- **2015:** `mu20_L1MU15`

- **2016:** `mu24`

The muon selections applied are summarised in Table 5.2. 'Tight' muons, used to identify the decay of the W boson, are required to have a transverse momentum of at least 28 GeV . This threshold is chosen to be at least 5% larger than the lowest trigger threshold, in line with ATLAS recommendations. They are required to have a maximum pseudorapidity of $|\eta| < 2.5$, within the coverage of the ATLAS inner-detector, as tracks are required for muon reconstruction as described in Section 3.2.1. The Medium muon identification criterion, the ATLAS default, is chosen to balance the selection efficiency with the misidentification rate and to reduce systematic uncertainties. A Tight isolation requirement is imposed to eliminate muons resulting from decays of hadrons containing heavy flavour quarks. Tight 'Track To Vertex Association' (TTVA) impact parameter selections ensure the muon originates from the primary vertex. The impact parameters considered are the transverse impact parameter significance $|d_0|_{\text{sig}} = |d_0|/\sigma_{d_0}$ where d_0 the distance of closest approach in the transverse plane to the beamline, and $|z_0 \sin\theta|$ where z_0 is the distance from the event primary vertex along the z axis of the position d_0 is measured. For a graphical depiction of the d_0 and z_0 variables, see Figure 3.1.

In addition, 'Loose' muons are defined with a p_{T} threshold of at least 10 GeV and no isolation requirements, for the purpose of a second muon veto and implementing data-driven multi-jet background estimation (see Section 6.2). The Loose muon definition is constructed so that the set of Tight muons are a subset of the Loose muons.

A set of electron identification selections are also defined, in order to veto events with

Muon criteria		
	Tight muon	**Loose muon**
Muon p_{T}	$p_{\mathrm{T}} >$ 28 GeV	$p_{\mathrm{T}} >$ 10 GeV
Muon pseudorapidity	$\lvert\eta\rvert <$ 2.5	
Identification	Medium	
Isolation	Tight	None
TTVA	$\lvert d_0\rvert_{\mathrm{sig}} < 3.0$	
	$\lvert z_0 \cdot \sin\theta\rvert <$ 0.5	

Table 5.2: Selections used for muon identification in the Signal Region. In addition to the 'Tight' muons used for signal identification, 'Loose' muons are also defined for a second muon event veto and the implementation of the multi-jet background estimation.

a sufficiently energetic electron. This definition is given in Table 5.3. These are required to have a p_{T} of at least 10 GeV and a pseudorapidity of less than $|\eta| < 2.4$, with electrons located in the 1.37$\leq |\eta| \leq$1.52 transition region between the ATLAS ECal barrel and endcaps excluded. Loose identification and isolation criteria are imposed, to increase the range of 'electron-like' objects which trigger the veto. The resulting loss in event selection efficiency is acceptable, as this analysis is not statistically limited. TTVA requirements are also imposed in line with internal ATLAS recommendations.

Electron criteria	
Electron p_{T}	$p_{\mathrm{T}} >$ 10 GeV
Electron pseudorapidity	$\lvert\eta\rvert <$ 2.4 No transition region 1.37$\leq \lvert\eta\rvert \leq$1.52
Identification	Loose
Isolation	Loose
TTVA	$\lvert d_0\rvert_{\mathrm{sig}} < 5$
	$\lvert z_0 \cdot \sin\theta\rvert <$ 0.5

Table 5.3: Selections used for electron identification for the opposite lepton flavour event veto.

The jets used for this measurement are reconstructed using calorimeter cells clustered

by the anti-k_t algorithm [72] with a radius $R = 0.4$ as the input to the ParticleFlow algorithm, and are calibrated at the electromagnetic energy scale. Jets must have a transverse momentum of at least 30 GeV, and an absolute rapidity of less than 4.4. As the $|y| < 4.4$ rapidity range includes both the central and forward regions of ATLAS, both the central Jet Vertex Tagger (JVT) [83] and the Forward Jet Vertex Tagger (fJVT) [82] must be applied to suppress pile-up. Additionally, ATLAS employs a set of jet quality selections to distinguish jets originating from hard-scatter collisions from those reconstructed from calorimeter noise and non-collision backgrounds. Events containing 'bad' jets that do not satisfy these 'jet cleaning' criteria are vetoed, using the high-efficiency 'Loose' working point defined in [122].

A b-jet veto (identified using the DL1r tagging algorithm [79] operating at the 85% efficiency working point) is applied to the N_{jet} distribution to events with at least 3 jets to reduce the impact of the top quark background. This decision was made to target the bins with the largest contamination of top quark events, while avoiding uncertainties due to the b-tagging in bins where the contribution is negligible.

Jet criteria	
Jet Collection	Antikt4EMPFlowJets
Jet p_{T}	$p_{\text{T}} >$ 30 GeV
Jet rapidity	$\|y\| <$ 4.4
Pile-up suppression	JVT Tight, fJVT Loose
Cleaning	'LooseBad': High efficiency Working Point
b-jet veto	Veto events with at least 3 jets with at least one b-tagged jet (DL1r 85% WP) (N_{jets} distributions only)

Table 5.4: Selections used for jet identification.

The objects described above are subject to an overlap removal procedure recom-

mended internally by ATLAS. As part of this procedure, jets found within $\Delta R < 0.2$ of a lepton (electron or muon) are removed, and leptons subsequently found within $\Delta R < 0.4$ of a jet are removed. When applying the overlap removal, muons are required to have $p_{\mathrm{T}} >$ 10 GeV. All other selections described above remain the same.

Events entering the analysis must fulfil the criteria set out in Table 5.5. The W boson decay is identified by the 'signal' muon selection, and backgrounds from leptonically decaying vector bosons where a lepton is either misidentified or missed are suppressed by the veto on events with either a second muon or one electron. The $E_{\mathrm{T}}^{\mathrm{miss}}$ selection is imposed to account for the neutrino from W decay and suppress the Z+jet background, while the cut on the transverse mass m_{T}^{W} is intended to suppress the diboson background. The transverse mass is defined as $m_{\mathrm{T}}^{W} = \sqrt{2p_{\mathrm{T}}^{\mu}E_{\mathrm{T}}^{\mathrm{miss}}(1-\cos(\Delta\phi(\mu, E_{\mathrm{T}}^{\mathrm{miss}})))}$. Finally, at least one jet is required.

Event selections	
Lepton selection	Exactly one Tight muon
Second same flavour lepton veto	No second Loose muon
Opposite flavour lepton veto	No Loose electron
Missing transverse momentum	$E_{\mathrm{T}}^{\mathrm{miss}}>$25 GeV
W transverse mass	$\mathrm{m}_{\mathrm{T}}^{W} >$ 60 GeV
N_{jet}	$N_{\mathrm{jet}} \geq 1$

Table 5.5: Muon channel event-level signal requirements.

5.4 Analysis chain overview

An overview of the procedure for measuring the unfolded differential cross-sections for W+jets production is as follows:

- The event selections described in Section 5.3 are applied to data and to both signal and background MC simulation samples. The events which pass these selections are then used to fill histograms of the distributions described in Section 5.1.
  - Simulated events are reweighted before being used to fill histograms. This is done to account for differences in detector response between data and MC, and to ensure the MC normalisation corresponds to the relevant data luminosity. This is described in more detail in Section 5.4.1.
- The background due to QCD multi-jet production is calculated using the data-driven 'Matrix Method'. This is described in Section 6.2.
- All background estimates and signal MC samples are combined to give a 'total Standard Model prediction' at reconstruction level, which is compared to data as shown in Section 9.1, to highlight the potential strengths or weaknesses of the signal models in different regions of phase space.
- Distributions are then scaled by bin width, to show the number of events per x axis unit, and backgrounds are subtracted from the data. The resulting distribution is then unfolded and compared to the truth-level $W \to \mu\nu$ signal predictions. These

are the final unfolded differential cross-sections which the analysis aims to produce. The unfolding procedure is discussed in further detail in Chapter 7.

– Experimental systematic uncertainties and theoretical uncertainties are evaluated by varying, for each uncertainty, the affected kinematic variable or event weight by that uncertainty, the repeating the analysis in full. The 1σ experimental and theoretical error bands on the final unfolded cross-sections are then the quadrature sum of the difference between each resulting systematically-varied distribution and the nominal distribution. This is discussed in more detail in Chapter 8.

5.4.1 MC reweighting

As discussed in Section 5.4, simulated events must be reweighted, to correct for differences in performance between data and simulation for several aspects of ATLAS reconstruction, and due to the need to correct the normalisation of the MC sample to the collected data luminosity.

Several key aspects of ATLAS detector response are challenging to accurately model in simulation, and MC samples must therefore be recalibrated to account for differences in performance with respect to data as a result of this mismodelling. In the case of W+jets, affected areas of reconstruction performance include lepton identification and isolation, jet reconstruction and calibration, pile-up estimations, trigger efficiencies and non-trivial variables such as JVT, fJVT and b-Tagging. This correction is accomplished through the

application of multiplicative 'scale factors' to the event weight for each area of detector performance affected by mismodelling. These scale factors are usually calculated in bins of p_{T} and rapidity from the ratio of efficiencies:

$$SF = \frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{MC}}},$$

where the efficiencies $\epsilon_{\mathrm{data,MC}}$ are derived separately in data and MC, and, using the example of a pile-up jet tagger for the purposes of illustration, can be defined as

$$\epsilon = \frac{N_{\mathrm{PUjet}}^{\mathrm{tagged}}}{N_{\mathrm{PUjet}}^{\mathrm{total}}}.$$

In addition to the application of scale factors, MC events must also be reweighted by their 'luminosity weight', an additional multiplicative factor, to correct for differences between the number of MC events generated in a given sample and the data luminosity collected. The luminosity weight w_{lumi} of a given MC sample is defined as

$$w_{\mathrm{lumi}} = \frac{L \times \sigma}{\sum w} \times k \times \epsilon_{\mathrm{filter}}, \tag{5.1}$$

where L is the integrated data luminosity to which the normalisation is being corrected. σ is the NNLO cross-section of the process being simulated, or beyond where applicable e.g. for $t\bar{t}$ production. This is calculated either directly, or at a lower order and corrected using a k-factor k. Some samples have certain slices filtered according to their b hadron and c hadron content at particle-level; in those cases, $\epsilon_{\mathrm{filter}}$ is the efficiency of the filter applied. Finally, $\sum w$ is the sum of the MC weights of all events generated as part of the

sample.

Chapter 6

Background estimation

"The world is a stage, but the play is badly cast"

— Oscar Wilde, 'Lord Arthur Savile's Crime'

While the Signal Region event selections outlined in Chapter 5 are chosen such that they maximise both the purity of $W \rightarrow \mu\nu$+jets signal events and background rejection efficiency, a number of background events remain which meet these criteria and enter the distributions produced. Background processes, or processes other than the signal which pass analysis selections, generally fall into one of two categories: **Reducible** and **irreducible**. Reducible backgrounds generally arise due to object misidentification during reconstruction and can be curtailed using targeted event selections, as the misidentified objects often have different kinematics than the bonafide signal objects. Irreducible backgrounds are more challenging to eliminate as they generally arise from processes with the same final state objects as the signal. In order to produce the total Standard Model pre-

diction to be subtracted from the measured spectra before unfolding, both the remaining reducible background events and irreducible backgrounds must be accurately accounted for and modelled.

Processes contributing to the W+jets background generally include an isolated, high-energy lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$, both either real or mimicked through mismeasurement. The $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement can be fulfilled either from neutrinos from heavy-flavour decays, or through mismeasurement of the deposited energy in an event. Efforts to reduce $E_{\mathrm{T}}^{\mathrm{miss}}$ mismeasurement in future ATLAS analyses through the identification and removal of pile-up jets have previously been discussed in Chapter 4. The final state jet may originate from a variety of physical sources, including pile-up, hadronic decay of vector bosons or τ leptons, and initial/ final state radiation.

Two methods of background estimation are employed in this analysis, depending on the source of the background. These are: Estimation from MC simulation, described in Section 6.1; and the data-driven multijet background estimation, described in Section 6.2.

6.1 MC estimated backgrounds

One of the most common methods of modelling backgrounds is by simulating the background processes using Monte-Carlo simulation. These simulations are subjected to the same Signal Region selections as the data and signal simulation, in addition to the same reweighting procedure as the signal simulation described in Section 5.4.1. The following

backgrounds to W+jets production in the muon channel are estimated using MC simulation:

- Z**+jets production:** This contributes when a muon is missed in a $Z \to \mu\mu$ decay, which can give rise to fake $E_{\mathrm{T}}^{\mathrm{miss}}$. This is modelled using Sherpa 2.2.11 [97] with up to two jets modelled at NLO and a further three calculated at LO in the matrix element, with higher jet multiplicities handled by the parton shower. The NNPDF3.0 PDF set [102] at NNLO is used. These samples are normalised to the NNLO generator cross-sections;

- **Top quark production:** This background is introduced through $t \to Wb$ decays, where the W decay mimics the W+jets signal. This background can arise either from single top decay or from the higher cross-section $t\bar{t}$, where one top decays hadronically or one lepton is missed. Top quark production is the dominant background at high jet multiplicities and is reduced through a veto on b-jets in the N_{jet} $\geq$3 bins of the jet multiplicity distribution, where this background dominates. The top background is modelled using PowhegBox version 2 [87, 87, 87, 123, 124], which calculates the matrix element at NLO. The NNPDF3.0NLO PDF is used for the matrix element. PowhegBox is interfaced to Pythia8 [88] set to the A14 tune [119] and using the NNPDF2.3 PDF set [92] at LO for parton shower, hadronization and underlying event. The $t\bar{t}$ samples have k-factors, provided centrally by ATLAS, applied to correct their normalisation to NNLO+NNLL. For single-top production, k-factors are applied to normalise the samples to the NLLO cross-sections;

- **Tau production:** $W \rightarrow \tau\nu$ production contributes to the muon channel when the tau decays leptonically to a muon, with both neutrinos being combined into a single $E_{\mathrm{T}}^{\mathrm{miss}}$ signal. $Z \rightarrow \tau\tau$ can also contribute if one of the $\tau \rightarrow l\nu$ leptons is missed or if the tau decays hadronically. Both of these processes are modelled using the same Sherpa 2.2.11 setup as for Z production described above. These samples are normalised to the NNLO generator cross-sections;

- **Diboson:** This can come from a number of processes – WW, WZ, ZZ where leptons are missed or misidentified, and/or if one boson decays hadronically. This background is suppressed relative to the others due to its lower cross-section. The matrix element is modelled at NLO with Powheg using the CT10 PDF set [125] at NLO and interfaced to Pythia8 using the CTEQ6L1 PDF set [90] and the AZNLO tune [89] for the parton shower. These samples are normalised to the NLO generator cross-sections;

In addition, a sample of QCD dijet production has been used for the purpose of comparison with the data-driven QCD multijet background estimate, which will be described in Section 6.2. This sample was produced at LO in QCD using Pythia 8 with the A14 tune and the NNPDF2.3 PDF set at LO.

6.2 Data-driven background

In the context of the muon channel, multijet production can mimic a W+jets signal from the production of a real, non-prompt muon from heavy quark decay, or the rarer case of fake muons from jet punch-through from the hadronic calorimeter to the muon spectrometer, both accompanied by $E_{\mathrm{T}}^{\mathrm{miss}}$. While the Signal Region event selections successfully eliminate a large number of multijet events. However, this process' large production cross-section ensures that this background must still be carefully assessed.

Non-prompt muons are those which do not originate from the collision's primary hard-scatter vertex. These can come from the leptonic decay of the original quark in heavy flavour jets, resulting in a muon embedded in the jet. The contribution from this background is reduced by the Tight muon isolation requirements and the ATLAS overlap removal procedure applied to Signal Region events, both described in Section 5.3.

Fake muons can arise from jet 'punch-through', where the energy of the jet is not entirely absorbed by the hadronic calorimeters. This allows some energy to reach the muon spectrometer where it can mimic a muon signal. This is a much rarer occurrence than the non-prompt case since, as described in Section 2.2.2, the ATLAS hadronic calorimeter is sufficiently deep ($\sim$10 hadronic interaction lengths) to almost completely absorb hadronic activity before reaching the muon spectrometer.

From this point, for the sake of brevity, both fake and non-prompt muons will be referred to collectively as 'fake muons'.

The multijet background (MJB) is often challenging to simulate to the desired degree of accuracy, as these simulations are quite sensitive to non-perturbative aspects of the modelling, and to the matching and merging methods employed. Therefore, in this analysis, the MJB is estimated using the data-driven 'matrix method' which requires two different sets of muon criteria to be defined:

- **Signal vs Control Regions:** These are orthogonal regions of phase space intended to be enriched in $W \to \mu\nu$+jets and MJB events respectively. The real muon efficiencies used in the MJB estimate will be calculated in the Signal Region, while the fake muon efficiencies will be calculated in the Control Region. A Control Region ideally has minimal contamination from signal, and similar fake composition and hadronic activity to the Signal Region. In the W+jets muon channel, as non-prompt muons from heavy flavour decays dominate the MJB, the Control Region is designed to target these fakes. Additionally, an optional Validation Region, orthogonal to both the Signal and Control Regions, may be defined for the purpose of validating efficiency calculations and MJB estimates. These three regions are distinguished on the basis of the muon transverse and longitudinal impact parameters, and will be discussed in greater detail in Section 6.2.1.

- **Tight vs Loose muons**: To calculate the real and fake efficiencies (in the Signal and Control Regions respectively), two collections of muons must be defined: Tight and Loose. In this analysis the same Tight and Loose selections as laid out in Table 5.2 for the signal muon selection and second muon veto are also employed for calculating MJB efficiencies, where the Tight and Loose selections are distinguished

on the basis of muon p_{T} and isolation.

	Signal	Validation	Control
Tight	$p_{\mathrm{T}} > 28$ GeV Isolation: Tight	$p_{\mathrm{T}} > 28$ GeV Isolation: Tight	$p_{\mathrm{T}} > 28$ GeV Isolation: Tight
	$0 \leq \lvert d_0 \rvert_{\mathrm{sig}} < 3$ $\lvert z_0 \cdot \sin\theta \rvert < 0.5$	$3 \leq \lvert d_0 \rvert_{\mathrm{sig}} < 5$ $\lvert z_0 \cdot \sin\theta \rvert < 0.5$	$d_0 \rvert_{\mathrm{sig}} > 5$ $\lvert z_0 \cdot \sin\theta \rvert < 0.5$
Loose	$p_{\mathrm{T}} > 10$ GeV Isolation: None	$p_{\mathrm{T}} > 10$ GeV Isolation: None	$p_{\mathrm{T}} > 10$ GeV Isolation: None
	$0 \leq \lvert d_0 \rvert_{\mathrm{sig}} < 3$ $\lvert z_0 \cdot \sin\theta \rvert$: None	$3 \leq \lvert d_0 \rvert_{\mathrm{sig}} < 5$ $\lvert z_0 \cdot \sin\theta \rvert$: None	$d_0 \rvert_{\mathrm{sig}} > 5$ $\lvert z_0 \cdot \sin\theta \rvert$: None

Table 6.1: A summary of the muon selections used in the MJB estimate where they differ from those laid out in Table 5.2, with the exception of those in the 'Signal' column, which lists a subset of the selections laid out in Table 5.2 and is included for completeness. Criteria coloured red arise from the definitions of the different regions in which the real and fake efficiencies are estimated separately, while criteria coloured blue are those which distinguish the Tight and Loose muon collections, which are both needed to calculate each of the efficiencies.

The above results in a total of six collections of muons used for the MJB estimate - two (Tight and Loose) for each of the three regions (Signal, Control and Validation). These are summarised in Table 6.1. The data-driven Matrix Method employed to estimate the MJB will be discussed in Section 6.2.2. The real and fake efficiencies calculated as inputs to the MJB estimate are discussed and presented in Section 6.2.3 and the MJB distributions derived are shown in Section 6.2.4 and compared to a sample of MC-modelled dijets in the Signal Region.

6.2.1 QCD multi-jet Control Region

As discussed in Section 6.2, the fake muons making up the QCD multijet background in the W+jets muon channel are expected to largely originate from leptonic decays of heavy flavour quarks. These non-prompt decays are generally displaced with respect to the ATLAS interaction point, and as such are expected to have larger impact parameters than muons from $W \to \mu\nu$+jets production. Therefore, the MJB Control Region in the W+jets muon channel is defined with respect to the muon impact parameter selections.

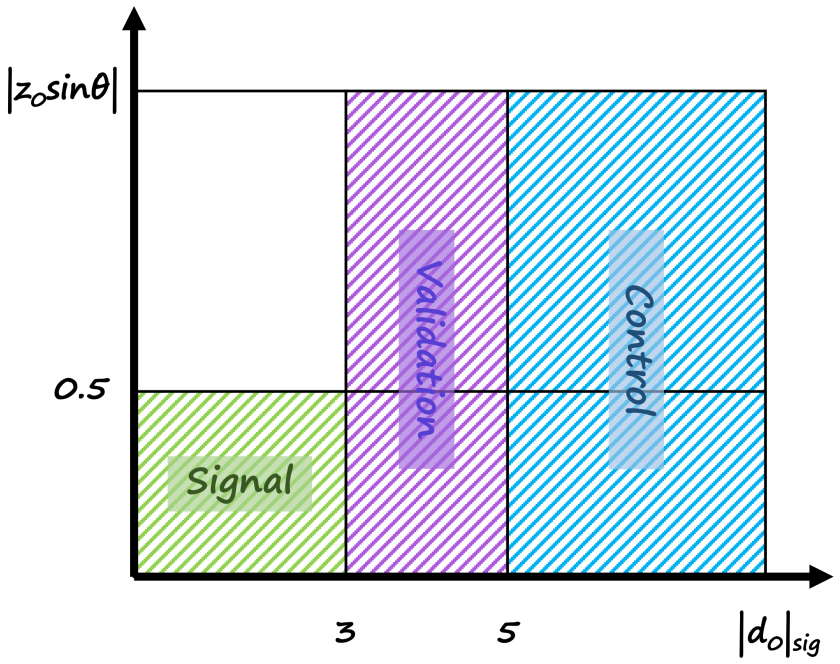


Figure 6.1: An illustration of the different muon impact parameter selections used to define the Signal Region (green), Validation Region (pink) and Control Region (blue) of the W+jets muon channel.

Figure 6.1 illustrates the different impact parameter selections used in the muon channel Signal Region and Control Region. As discussed in Section 5.3, in the Signal Region, muons are required to have $|d_0|_{\text{sig}} < 3.0$ and $|z_0 \cdot \sin\theta| < 0.5$. In the Control Region, $|d_0|_{\text{sig}}$

is required to be greater than 5.0 and the $|z_0 \cdot \sin\theta|$ selection is removed. All other Signal Region selection described in Section 5.3 are unchanged in the Control Region. The pink region in Figure 6.1, 3.0$< |d_0|_{\text{sig}} <$ 5.0, is orthogonal to both the Signal and Control Regions, and thus presents a viable candidate for a Validation Region in which to verify aspects of the data-driven MJB, which will be shown throughout this chapter.

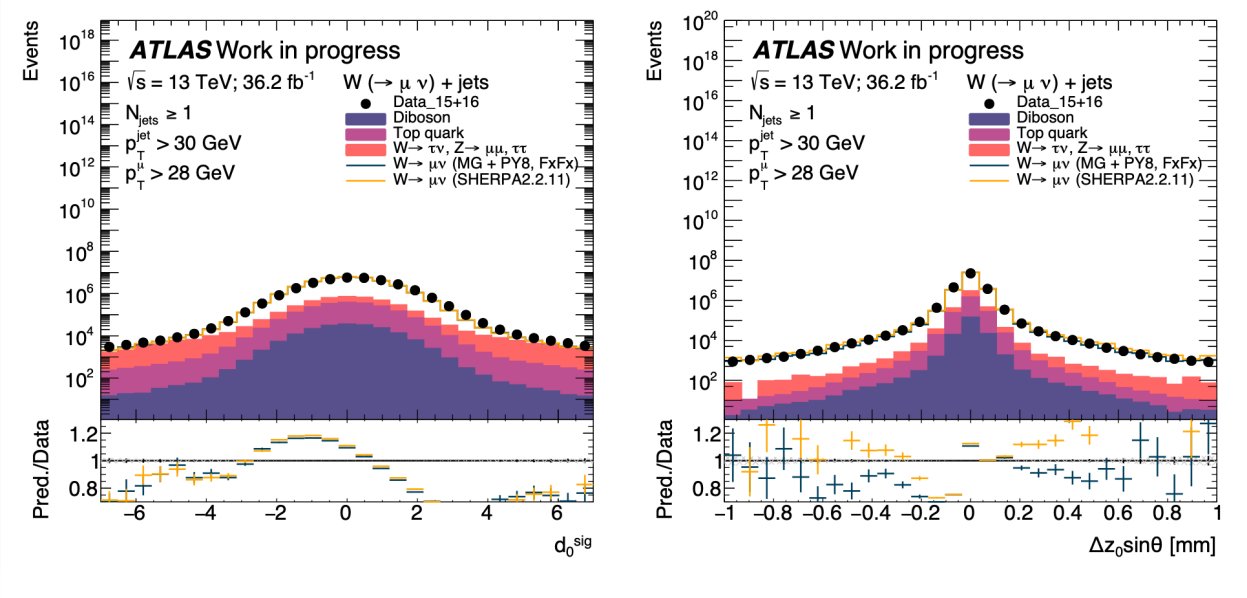


Figure 6.2: Data compared to the MC modelled $W \rightarrow \mu\nu$+jets and electroweak backgrounds in the $|d_0|_{\text{sig}}$ and $|z_0 \cdot \sin\theta|$ distributions, with all selections but $|d_0|_{\text{sig}}$ and $|z_0 \cdot \sin\theta|$ respectively applied.

Figure 6.2 shows data compared to the MC modelled electroweak backgrounds and the $W \rightarrow \mu\nu$ signal in the $|d_0|_{\text{sig}}$ and $|z_0 \cdot \sin\theta|$ distributions, with all selections but $|d_0|_{\text{sig}}$ and $|z_0 \cdot \sin\theta|$ respectively applied. Because the MJB estimate is not included in these plots, regions with very large underestimations of the predictions with respect to the data can be assumed to be enriched in multijet events. This is especially the case for the regions where $|d_0|_{\text{sig}} >$ 5.0, where it can also be observed the contamination from the W+jets signal is minimal. However, the overestimation of the data in the $-2.5 \leq d_0^{\text{sig}} \leq 0.5$ region,

alongside the large shape differences in the MC/data ratio in the regions $d_0^{\mathrm{sig}} < 0$ and $d_0^{\mathrm{sig}} > 0$, suggest that there may be some degree of mismodelling in this variable. This shape difference is particularly striking in the ranges which define the Validation Region; however, it can also be observed at the beginning of the range which defines the Control Region up to $|d_0|_{\mathrm{sig}} < 6$. Further multijet enrichment from the expansion into the region $|z_0 \cdot \sin\theta| >$ 0.5 is marginal, but the addition of this phase space to the Control Region may provide a small benefit. The use of either the Sherpa or Madgraph $W \to \mu\nu$+jets signal sample results in very similar signal+electroweak background predictions in the Control Region.

Figure 6.3 shows the data distributions for the six variables to be unfolded compared to the $W \to \mu\nu$ signal prediction and MC modelled electroweak backgrounds in the Control Region. No MJB estimate has yet been added, meaning, similarly to Figure 6.2, any large discrepancy between the data and signal+electroweak background prediction indicates a strong likelihood of MJB enrichment. Figure 6.3 shows that the Control Region is comprised of $\sim$50% MJB, however, this does decline with increasing p_{T} . Table 6.2 which expresses the signal+electroweak background prediction in the Control Region as a percentage of the number of data events in bins of inclusive jet multiplicity up to 8 jets, also demonstrates the MJB enrichment in the Control Region. Table 6.2 also confirms that the signal contamination in the Control Region is low, $\lesssim 10\%$ for all jet multiplicities for both generators, which is a requirement when designing a Control Region.

Figure 6.4 shows the data distributions for the six variables to be unfolded compared to the $W \to \mu\nu$ signal prediction and MC modelled electroweak backgrounds in

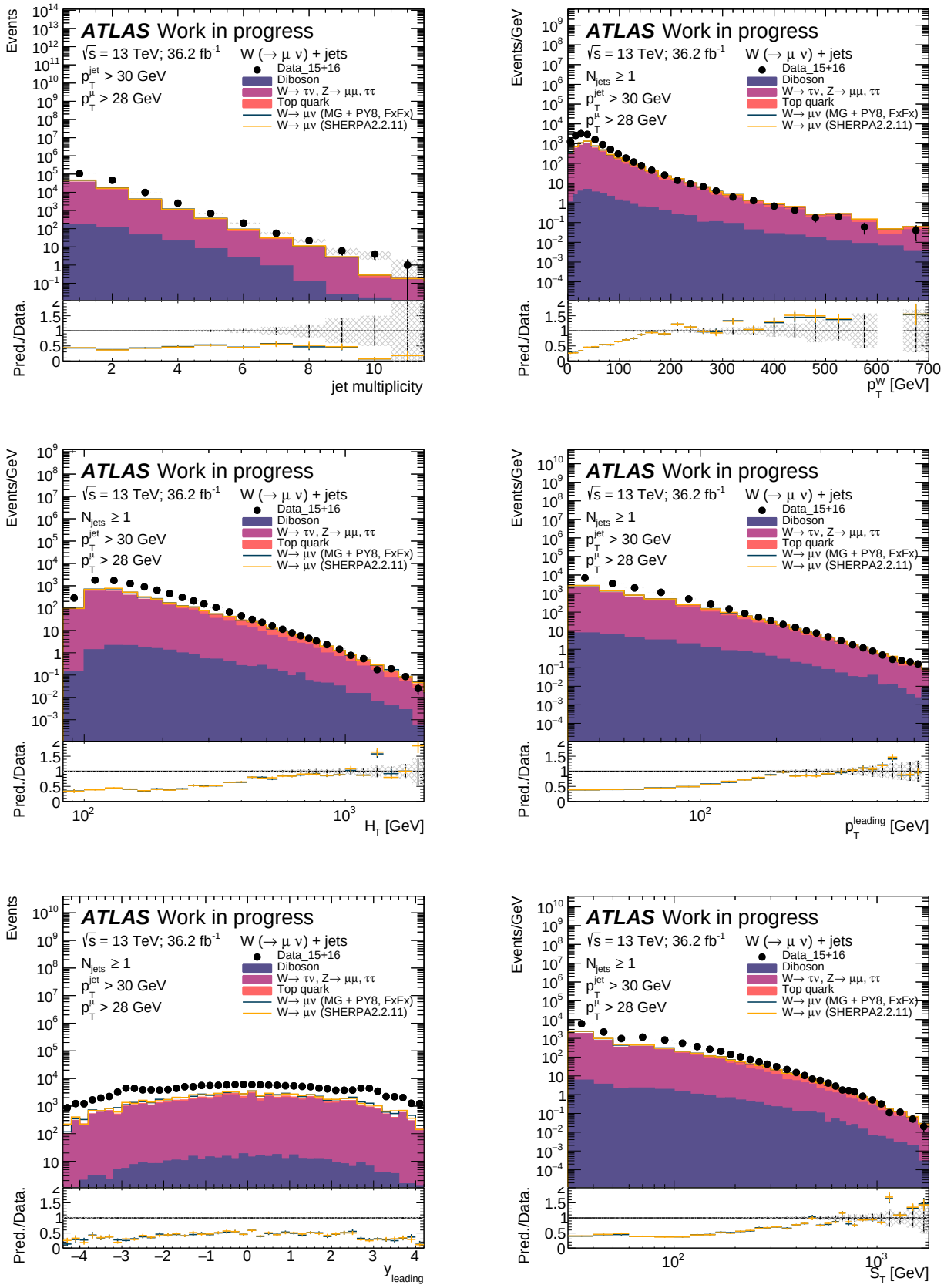


Figure 6.3: The data distributions for the six variables to be unfolded compared to the $W \to \mu\nu$ signal prediction and MC modelled backgrounds in the Control Region.

	$W \to \mu\nu$ (MG+PY8)	$W \to \mu\nu$ (SHERPA)	$W \to \tau\nu$, $Z \to \mu\mu, \tau\tau$	Top quark	Diboson
N_{jet}	%	%	%	%	%
$\geq$1	9.4	10.2	33.0	0.6	0.2
$\geq$2	6.0	4.7	28.5	3.1	0.2
$\geq$3	6.4	5.6	34.3	2.1	0.5
$\geq$4	5.3	3.5	37.4	4.4	0.8
$\geq$5	3.1	2.3	42.8	6.3	1.1
$\geq$6	2.7	1.7	31.8	10.2	1.3
$\geq$7	2.3	0.7	42.2	11.5	1.6
$\geq$8	0.2	5.0	39.5	6.8	0.6

Table 6.2: The MC-modelled composition of the Control Region for events with up to 8 jets. Percentages given are with respect to the number of data events.

the Validation Region, with no MJB added. The underestimation of the data by the signal+electroweak background prediction indicates an, approximately flat, $\sim$30% MJB contribution to the Validation Region. This is emphasised by the breakdown of signal+electroweak background prediction in the Validation Region as a percentage of the number of data events up to 8 jets in Table 6.3

	$W \to \mu\nu$ (MG+PY8)	$W \to \mu\nu$ (SHERPA)	$W \to \tau\nu$, $Z \to \mu\mu, \tau\tau$	Top quark	Diboson,
N_{jet}	%	%	%	%	%
$\geq$1	49.5	50.2	20.6	0.5	0.2
$\geq$2	39.0	40.2	25.9	4.3	0.5
$\geq$3	37.5	40.7	26.1	3.0	0.9
$\geq$4	32.8	35.8	26.8	6.0	1.4
$\geq$5	27.7	30.4	30.0	9.1	1.8
$\geq$6	20.1	41.1	36.7	10.4	2.0
$\geq$7	23.5	25.4	28.3	11.9	2.3
$\geq$8	17.9	29.6	33.2	7.0	0.5

Table 6.3: The MC-modelled composition of the Validation Region for events with up to 8 jets. Percentages given are with respect to the number of data events.

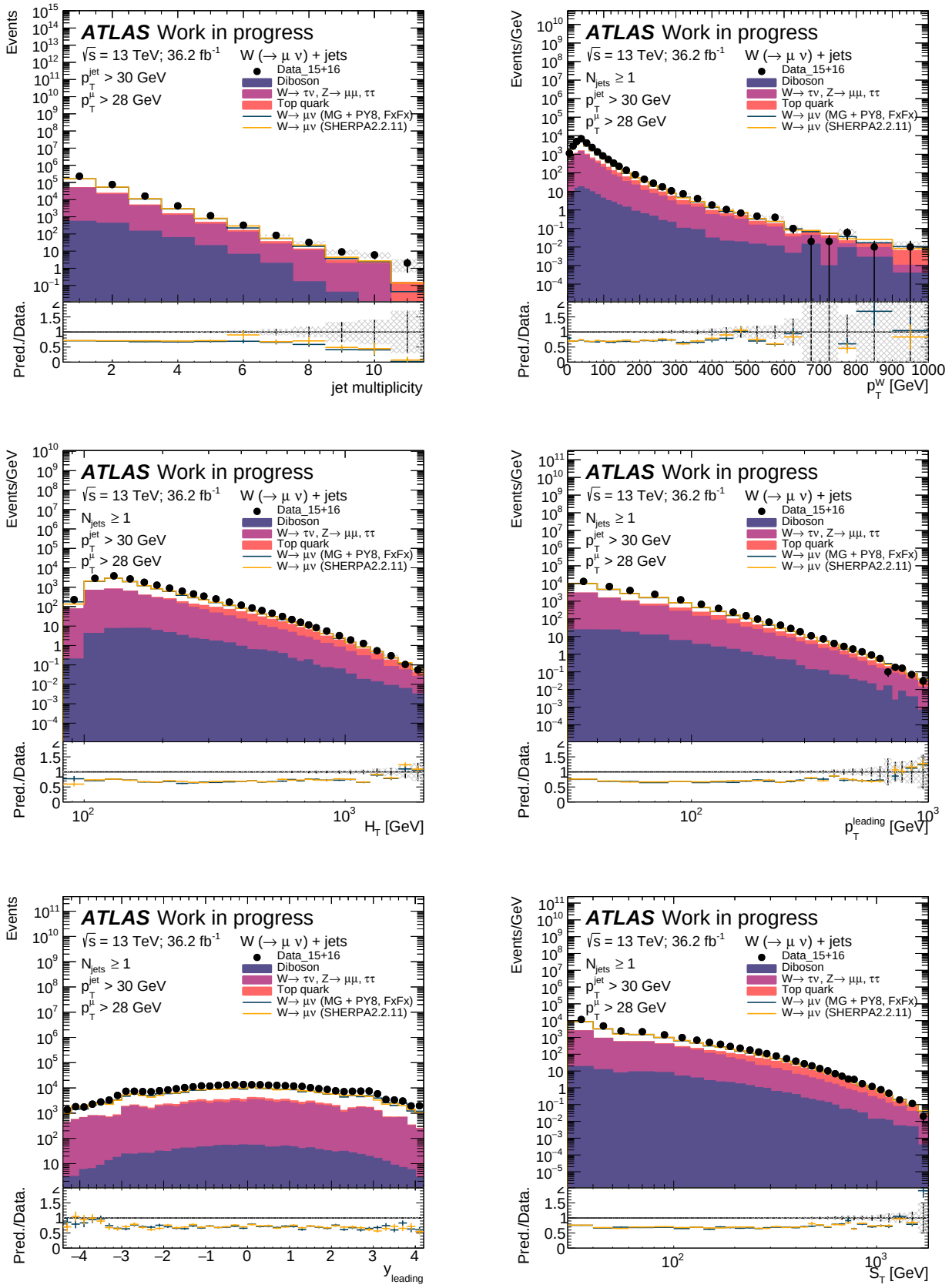


Figure 6.4: The data distributions for the six variables to be unfolded compared to the $W \to \mu\nu$ signal prediction and MC modelled backgrounds in the Validation Region.

6.2.2 The Matrix Method

The contamination of fake leptons originating from the multijet background in the W+jets Signal Region in both the electron and muon channels is estimated using the matrix method. The matrix method has several advantages as an approach to data-driven background estimation, being straightforward to express and, in principle, very precise. However, care must be taken with potential correlations between efficiencies and that contamination from additional processes in the Control Region are accounted for with accurate MC simulation, as the method cannot separate contributions from different processes. Typically used in ATLAS analyses involving top quarks [126], this measurement of W+jets production cross-sections is the first ATLAS V+jets analysis to employ the matrix method.

The matrix method presumes that, in the Signal Region, the total number of muons fulfilling the 'Tight' criteria set out in Section 5.3 (N^{t}) is composed of some number of real Tight muons ($N_{\mathrm{r}}^{\mathrm{t}}$) and some number of fake Tight muons ($N_{\mathrm{f}}^{\mathrm{t}}$). One can also loosen the muon identification criteria by removing the isolation requirement. In this case, the total number of Loose muons in the Signal Region (N^{l}) is composed of some number of real Loose muons ($N_{\mathrm{r}}^{\mathrm{l}}$) and some number of fake Loose muons ($N_{\mathrm{f}}^{\mathrm{l}}$),

$$\begin{aligned} N^{\mathrm{t}} &= N_{\mathrm{r}}^{\mathrm{t}} + N_{\mathrm{f}}^{\mathrm{t}}, \\ N^{\mathrm{l}} &= N_{\mathrm{r}}^{\mathrm{l}} + N_{\mathrm{f}}^{\mathrm{l}}. \end{aligned} \tag{6.1}$$

If, as in this analysis, the Tight selection is a subset of the Loose selection, a fake efficiency i.e. the fraction of Loose fake muons that also pass the Tight criterion can be

defined,

$$\epsilon_{\mathrm{f}} = \frac{N^{\mathrm{t}}_{\mathrm{f}}}{N^{\mathrm{l}}_{\mathrm{f}}}. \tag{6.2}$$

Similarly, a real efficiency ϵ_{r} can also be defined. Thus, Eq 6.1 can be written in terms of $N^{\mathrm{l}}_{\mathrm{r,f}}$, as:

$$\begin{aligned} N^{\mathrm{t}} &= \epsilon_{\mathrm{r}} N^{\mathrm{l}}_{\mathrm{r}} + \epsilon_{\mathrm{f}} N^{\mathrm{l}}_{\mathrm{f}}, \\ N^{\mathrm{l}} &= N^{\mathrm{l}}_{\mathrm{r}} + N^{\mathrm{l}}_{\mathrm{f}}, \end{aligned}$$

which can be expressed as the matrix equation

$$\begin{bmatrix} N^{\mathrm{t}} \\ N^{\mathrm{l}} \end{bmatrix} = \begin{bmatrix} \epsilon_{\mathrm{r}} & \epsilon_{\mathrm{f}} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} N^{\mathrm{l}}_{\mathrm{r}} \\ N^{\mathrm{l}}_{\mathrm{f}} \end{bmatrix}.$$

This can be inverted to give:

$$\begin{bmatrix} N^{\mathrm{l}}_{\mathrm{r}} \\ N^{\mathrm{l}}_{\mathrm{f}} \end{bmatrix} = \frac{1}{\epsilon_{\mathrm{r}} - \epsilon_{\mathrm{f}}} \begin{bmatrix} 1 & -\epsilon_{\mathrm{f}} \\ -1 & \epsilon_{\mathrm{r}} \end{bmatrix} \begin{bmatrix} N^{\mathrm{t}} \\ N^{\mathrm{l}} \end{bmatrix} \longrightarrow N^{\mathrm{l}}_{\mathrm{f}} = \frac{1}{\epsilon_{\mathrm{r}} - \epsilon_{\mathrm{f}}} (\epsilon_{\mathrm{r}} N^{\mathrm{l}} - N^{\mathrm{t}}).$$

Finally, substituting for $N^{\mathrm{l}}_{\mathrm{f}}$ using Eq 6.2 we obtain the expression for the number of fake muons in the Tight, i.e. Signal, selection

$$N^{\mathrm{t}}_{\mathrm{f}} = \frac{\epsilon_{\mathrm{f}}}{\epsilon_{\mathrm{r}} - \epsilon_{\mathrm{f}}} (\epsilon_{\mathrm{r}} N^{\mathrm{l}} - N^{\mathrm{t}}).$$

Several options are available for calculating the real and fake efficiencies. As the real

and fake efficiencies strongly depend on event characteristics, and thus vary from event to event, the approach taken in the muon channel is to parameterise the efficiencies in terms of the variables on which they depend and compute a single event-by-event weight, w_i, which is used to estimate the MJB in all distributions:

$$w_i = \frac{\epsilon_{\mathrm{f}}(\mathbf{x})}{\epsilon_{\mathrm{r}}(\mathbf{x}) - \epsilon_{\mathrm{f}}(\mathbf{x})}(\epsilon_{\mathrm{r}}(\mathbf{x}) - \delta_i), \tag{6.3}$$

where $\mathbf{x}$ is the set of variables by which the efficiencies are parameterised and the delta function is 1 if the event's lepton is Tight and 0 if it is Loose. Therefore, the fake contamination in a given bin is taken to be the sum of weights in that bin:

$$N_{\mathrm{f}}^{\mathrm{t}} = \sum_{i=1}^{N^{\mathrm{l}}} w_i. \tag{6.4}$$

Thus, to measure the multijet background the real and fake efficiencies must first be measured.

6.2.3 Efficiencies

Real efficiencies for the MJB estimate are measured using MC in the Signal Region, where the real efficiency in a given bin is the ratio of the number of Tight to Loose signal+electroweak background events in that bin:

$$\epsilon_{\mathrm{r}} = \frac{N^{\mathrm{t}}_{\mathrm{MC(sig+ewk)}}}{N^{\mathrm{l}}_{\mathrm{MC(sig+ewk)}}} \tag{6.5}$$

The MC samples used for the ϵ_{r} estimation have identical selections to those discussed in Section 5.3, with the exception that no trigger selection is applied as ϵ_{r} is calculated entirely using MC. Therefore, events in this sample were also not reweighted by the trigger efficiency scale factor.

The fake efficiency is measured in the Control Region using data, with the W+jets signal and electroweak background MC subtracted to isolate the MJB. Using Eq. 6.1, Eq. 6.2 and Eq. 6.5, the expression for fake efficiency can be defined:

$$\epsilon_{\mathrm{f}} = \frac{N_{\mathrm{f}}^{\mathrm{t}}}{N_{\mathrm{f}}^{\mathrm{l}}} = \frac{N^{\mathrm{t}} - N_{\mathrm{r}}^{\mathrm{t}}}{N^{\mathrm{l}} - N_{\mathrm{r}}^{\mathrm{l}}} = \frac{N_{\mathrm{data}}^{\mathrm{t}} - N_{\mathrm{MC(sig+ewk)}}^{\mathrm{t}}}{N_{\mathrm{data}}^{\mathrm{l}} - N_{\mathrm{MC(sig+ewk)}}^{\mathrm{l}}}. \tag{6.6}$$

Care must be taken with the trigger strategy employed for the ϵ_{f} estimation. In particular, is must be ensured that the online triggers applied to Loose data are looser than than the offline selections, so as not to bias the Loose sample. Unfortunately, the lower p_{T} threshold unprescaled single muon triggers used in 2015 and 2016 both have isolation requirements, a tighter online selection than the offline Loose muons which have no isolation requirement. Therefore, an alternate set of triggers with no isolation requirement was selected for the Loose samples in the denominator of Eq. 6.6. The triggers selected, `mu20_L1MU15` and `mu24` for 2015 and 2016 data respectively have their details discussed in Section 5.3.

As the removal of the isolation requirement greatly increases the rate of event acceptance, each of these Loose triggers were prescaled (described in Section 2.2.4) by a constant factor throughout their respective years. The `mu20_L1MU15` trigger employed in

2015 has a prescale factor of 10, while the 2016 `mu24` trigger has a prescale of 50. [1] To reconcile the normalisation difference between the numerator and the denominator of Eq. 6.6, the Tight events in the numerator are reweighted by a factor of 0.1 or 0.05 according to the year to which they correspond.

As discussed in Section 6.2.2, the real and fake efficiencies used to estimate the MJB are parameterised in terms of the, ideally independent, variables over which they vary. For the W+jets muon channel, the following variables have been chosen, based on a previous application of the matrix method to estimate a fake muon background in association with jets in ATLAS top quark events [126]:

- The transverse momentum of the muon, p_{T}^{μ};
- The absolute value of the pseudorapidity of the muon, $|\eta_{\mu}|$;
- The minimum angular distance between the muon and the nearest jet, min($\Delta R_{\mu,\mathrm{jet}}$).

In an ideal scenario, efficiencies would be extracted from a multidimensional histograms of all the parameterisation variables simultaneously. However, with limited statistics this approach would yield large statistical uncertainties. Instead, the real and fake efficiencies are calculated separately as one-dimensional functions of each parameterisation variable $x_i \in \mathbf{x}$, assuming each of the x_i's to be independent. On the basis of this requirement, the fake efficiencies for this MJB estimate must therefore calculated in-

[1] At present, the trigger efficiency correction scale factors are not available for the prescaled triggers applied to the Loose data. As the MJB will be shown to be a small contribution to the Signal Region in Chapter 9, it has been deemed by the analysis team that the $\mathrm{SF}_{\mathrm{L}}^{\mathrm{trig}} \sim 1$ is a valid approximation. For a more precise estimate, a Tag and Probe style evaluation of the correction factors, similarly to [121], could be performed for the prescaled triggers.

dividually for each value of exclusive jet multiplicity, as it will be shown in Figures 6.5-6.7 that the shape and normalisation of the other variables are correlated with jet multiplicity, and therefore the assumption of independence does not hold. However, assuming the remaining x_i's are approximately uncorrelated with each other, these 1D efficiencies can be treated as conditional probabilities. A total real or fake efficiency for a given event of a given exclusive jet multiplicity, as a function of the full set of parameterisation variables, can then be expressed as:

$$\epsilon_{\mathrm{r,f}}(\mathbf{x}; N_{\mathrm{jet}}) = \epsilon_{\mathrm{r,f}}(N_{\mathrm{jet}}) \prod_{i}^{M} \frac{\epsilon_{\mathrm{r,f}}(x_i; N_{\mathrm{jet}})}{\epsilon_{\mathrm{r,f}}(N_{\mathrm{jet}})}, \tag{6.7}$$

where M is the number of parameterisation variables.

Figures 6.5 - 6.7 show the real and fake efficiencies calculated as described in this section in exclusive jet multiplicity bins from N_{jet} =1 to N_{jet} $\geq$5 for each of the variables by which they are parameterised; p_{T}^{μ} (Figure 6.5), $|\eta^{\mu}|$ (Figure 6.6) and min($\Delta R_{\mu,\mathrm{jet}}$) (Figure 6.7). Real efficiencies calculated in both the Signal Region and the Validation Region are shown. Figure 6.8 shows the the real and fake efficiencies in terms of different inclusive jet multiplicities. The vertical error bands shown are the statistical uncertainties propagated through the addition/subtraction and division in the efficiency estimate. Bins for the p_{T}^{μ} and min($\Delta R_{\mu,\mathrm{jet}}$) distributions were optimised by hand, to ensure the shape of the background is captured while statistical uncertainties are minimised. Two bins are chosen for the efficiencies parameterised in terms of $|\eta^{\mu}|$; one for the barrel region, $|\eta|$ <1.5, and one for the endcap region, 1.5< $|\eta|$ <2.5.

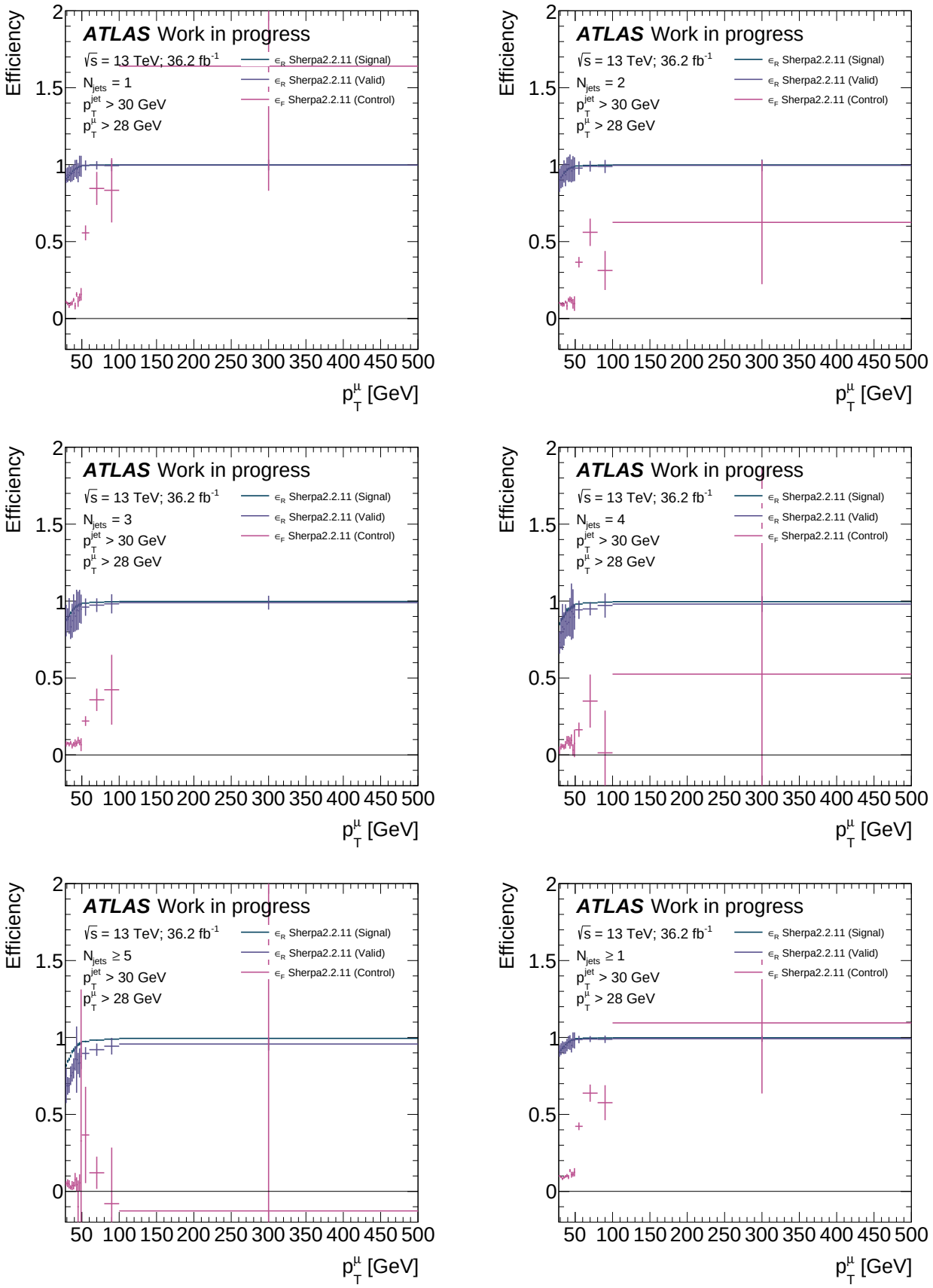


Figure 6.5: 1D conditional ϵ_{R}'s and ϵ_{F}, in terms of p_{T}^{μ} calculated in exclusive jet multiplicity bins from N_{jet} =1 to N_{jet} $\geq$5, and for inclusive jet multiplicity $N_{\mathrm{jet}} \geq 1$ (bottom right).

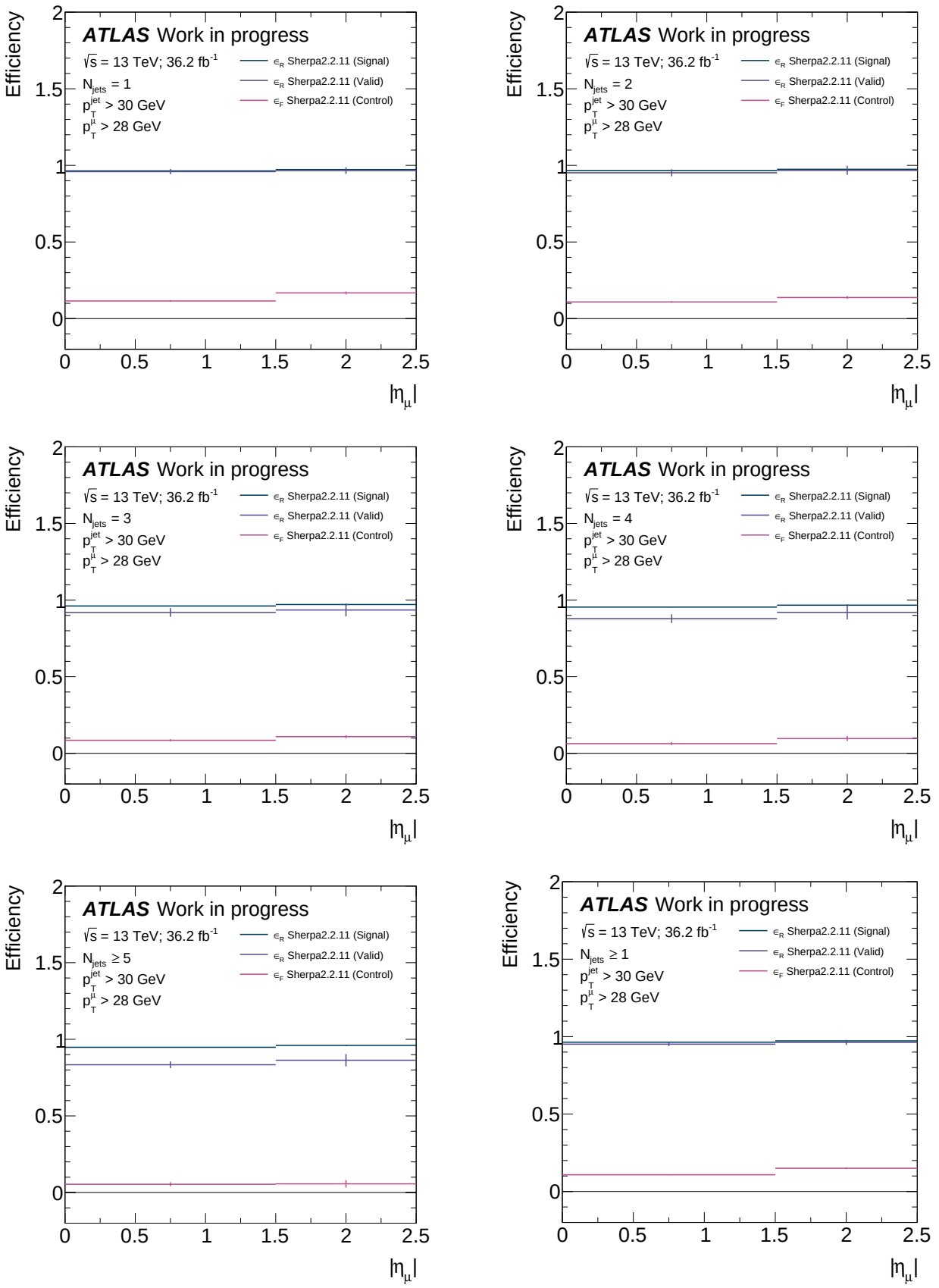


Figure 6.6: 1D conditional ϵ_{R}'s and ϵ_{F}, in terms of $|\eta_{\mu}|$ calculated in exclusive jet multiplicity bins from N_{jet} =1 to N_{jet} $\geq$5, and for inclusive jet multiplicity $N_{\mathrm{jet}} \geq 1$ (bottom right).

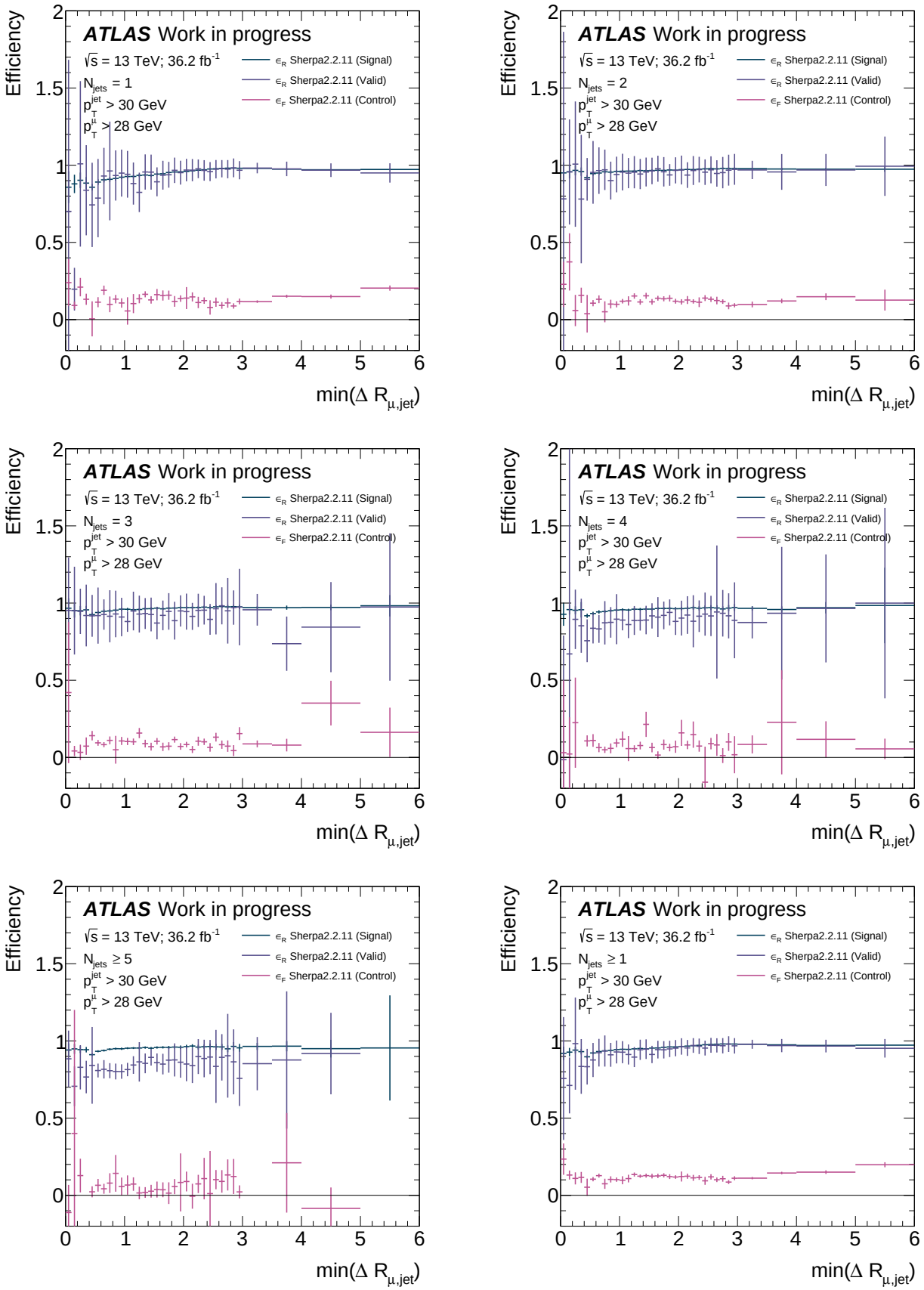


Figure 6.7: 1D conditional ϵ_{R}'s and ϵ_{F}, in terms of min($\Delta R_{\mu,\mathrm{jet}}$) calculated in exclusive jet multiplicity bins from N_{jet} =1 to N_{jet} $\geq$5, and for inclusive jet multiplicity $N_{\mathrm{jet}} \geq 1$ (bottom right).

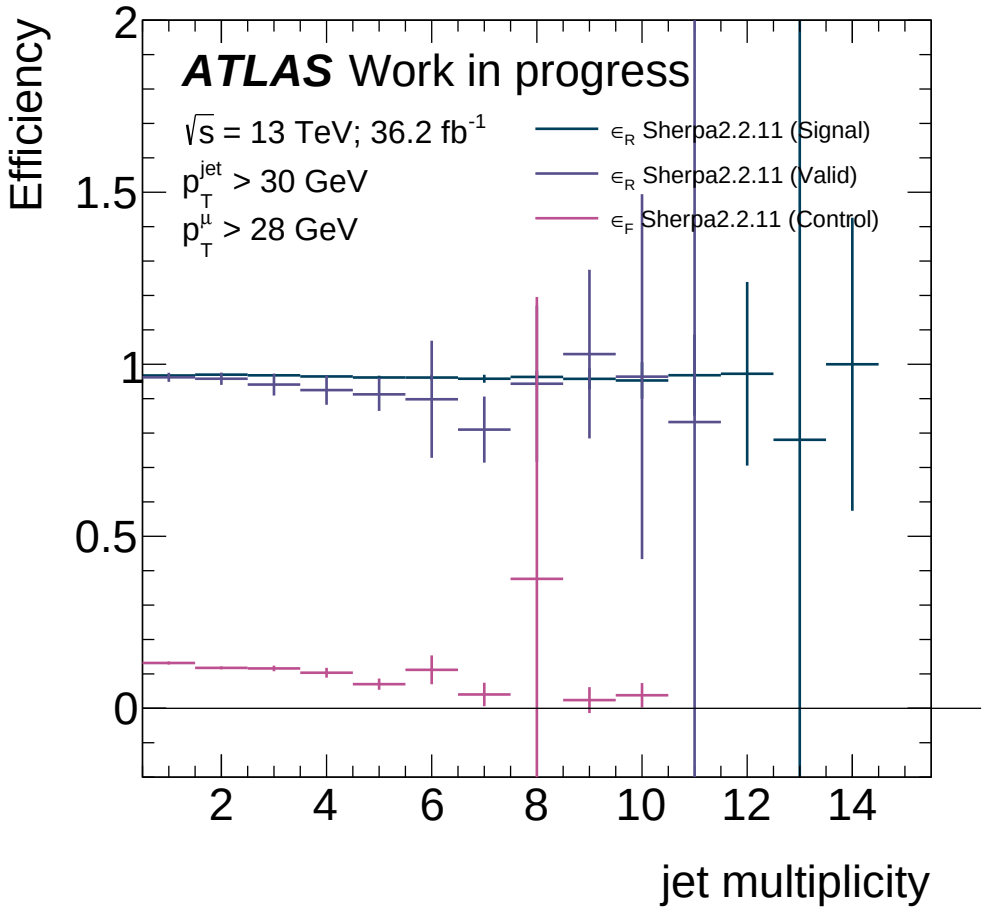


Figure 6.8: Real and fake efficiencies for each inclusive N_{jet} multiplicity.

An important note is that the assumption that the x_i's are completely uncorrelated may not hold exactly true for all variables. This could be accounted for in future studies by extracting those efficiencies from an N-dimensional histogram of the correlated variables. However, the remainder of the Run 2 data must first be added to properly asses where the trade-off with increased uncertainties from low-statistics bins lies, however, this is beyond the scope of the version of the analysis presented in this thesis.

6.2.4 Multi-jet background estimate

The procedure described previously in Section 6.2.2 is then followed to estimate the multijet background contamination in the Signal Region. As the weights calculated in Eq. 6.3 are applied to Loose data in Eq. 6.4, the same sample taken using Loose triggers as used to calculated ϵ_{f} must be utilised. To accurately compare the MJB estimated using this prescaled sample to the data and MC backgrounds in the Signal Region which employ unprescaled triggers, events in the MJB estimate are further reweighted according to the prescale factor applied in their respective year.

In Figure 6.9, the data-driven MJB estimate is shown, and compared to the MC sample of dijet production described in Section 6.1 with the Signal Region selections applied. The vertical error bars represent the statistical uncertainties for each distribution.

While the data-driven background does exhibit fluctuations, particularly in higher momentum bins due to having been derived from a prescaled sample with its normalisation subsequently scaled up, both estimates generally agree to within the same order of mag-

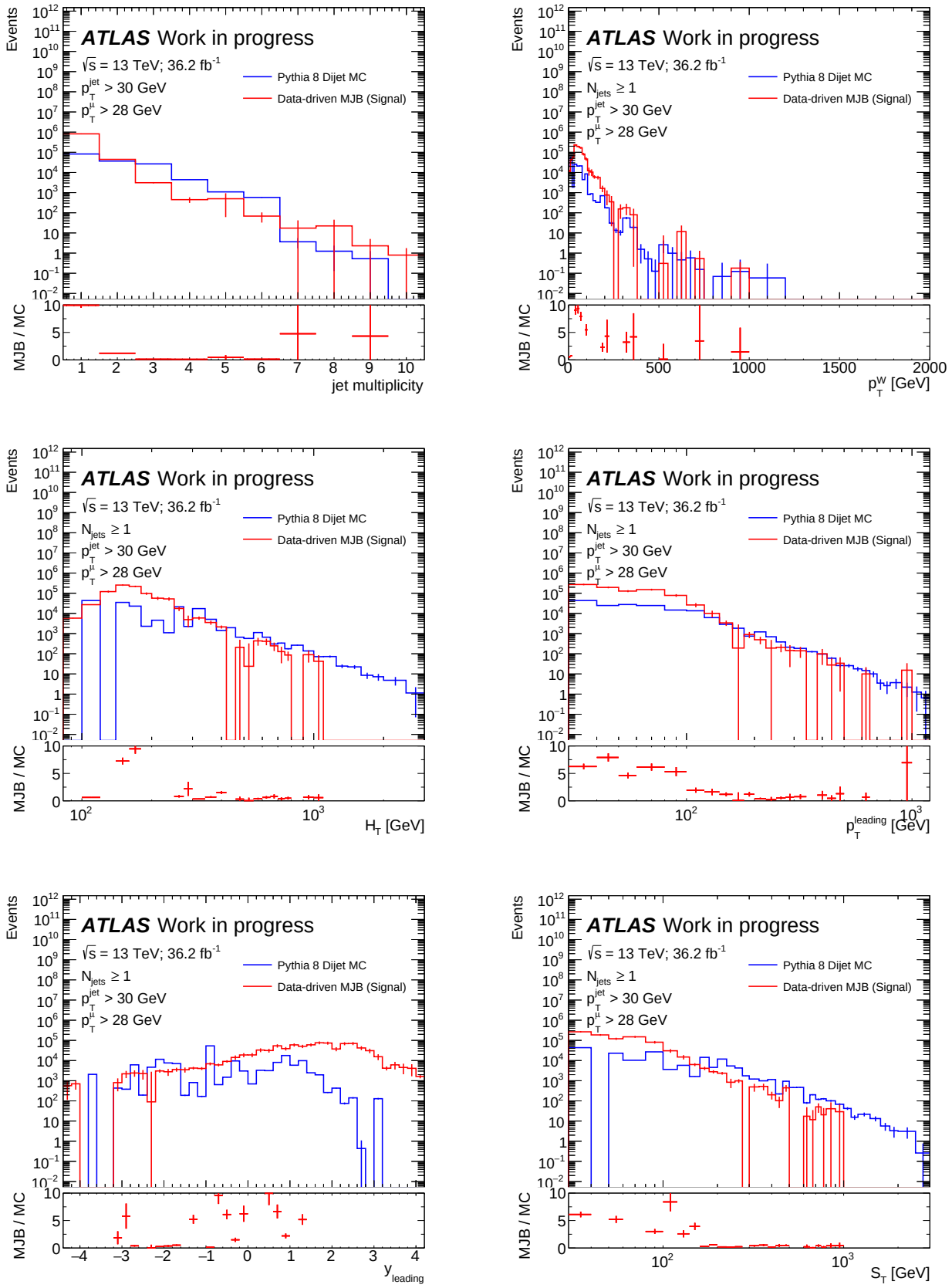


Figure 6.9: the data-driven MJB estimate is shown, and compared to the MC sample of dijet production described in Section 6.1 with the Signal Region selections applied.

nitude, with the exception of the y^{leading} distribution which exhibits strong disagreement with the MC estimate in the forward regions. One strategy for improving the fluctuations in the data-driven estimate would be to smooth the MJB distribution by performing a fit.

6.2.5 The Validation Region

A data-MC comparison in the Validation Region, including the MJB estimated using real efficiencies calculated in the Validation Region, is shown in Figure 6.10, with only statistical uncertainties included. Dedicated real efficiencies calculated in the Validation Region are used, due to the different signal + electroweak background composition with respect to the Signal Region. Similar plots in the Signal Region, with full experimental systematic uncertainties in addition to the statistical uncertainties, will be presented and discussed in Section 9.1.

A consistent underestimation of $\sim$20% is observed in the Validation Region for all observables. One possible explanation for this is the potential mismodelling of the d_0^{sig} distribution, which would result in incorrectly estimated real and fake efficiencies in regions where the mismodelling is present, arising from an incorrect signal + electroweak MC sum or subtraction from data for the re4al and fake efficiencies respectively. Indeed, the difference in the prediction/data ratio in Figure 6.2 between $-6 \leq d_0^{\text{sig}} - 3$ and $3 \leq d_0^{\text{sig}} < 6$, ranges which include both the Validation and Control Regions where the real and fake efficiencies respectively used the in Figure 6.2 MJB estimate are calculated, appears to be approximately the same size as the underestimation in Figure 6.10. This

could be hinted at by Figures 6.5-6.8, where, particularly for larger jet multiplicities, real efficiencies calculated separately in the Signal and Validation Regions do not agree within statistical uncertainties. Such mismodelling may also have an impact on the MJB estimate in the Signal Region, where the overestimation of the prediction/data ratio observed in this region in Figure 6.2 may impact the real efficiency estimates, and fake efficiencies from the affected Control Region are used.

Another plausible explanation arises from an effect which has previously been observed in the electron channel MJB estimate for this analysis, as documented in [127]. This was found to have arisen due to negative MJB event weights generated when two variables used for efficiency parameterisation both included bins where the fake efficiencies were significantly larger than the mean for that jet multiplicity. In the muon channel, this can be observed to be the case for high muon p_{T} bins in Figure 6.5, and low min($\Delta R_{\mu,\mathrm{jet}}$) bins in Figure 6.7. In [127], the negative weights were corrected by substituting the affected 1D efficiencies for 2D efficiencies calculated in terms of both variables, thereby accounting for correlations between them that had previously been neglected. This presents an avenue for further refinement of the MJB estimate in the muon channel; however, any improvement in the background modelling would have to be evaluated with respect to the additional statistical uncertainties 2D efficiencies would introduce.

An additional refinements to the background estimation could include performing an additional data-driven estimate of the top quark background to validate the MC-driven estimate. This is particularly important in high jet multiplicity bins where this background is dominant.

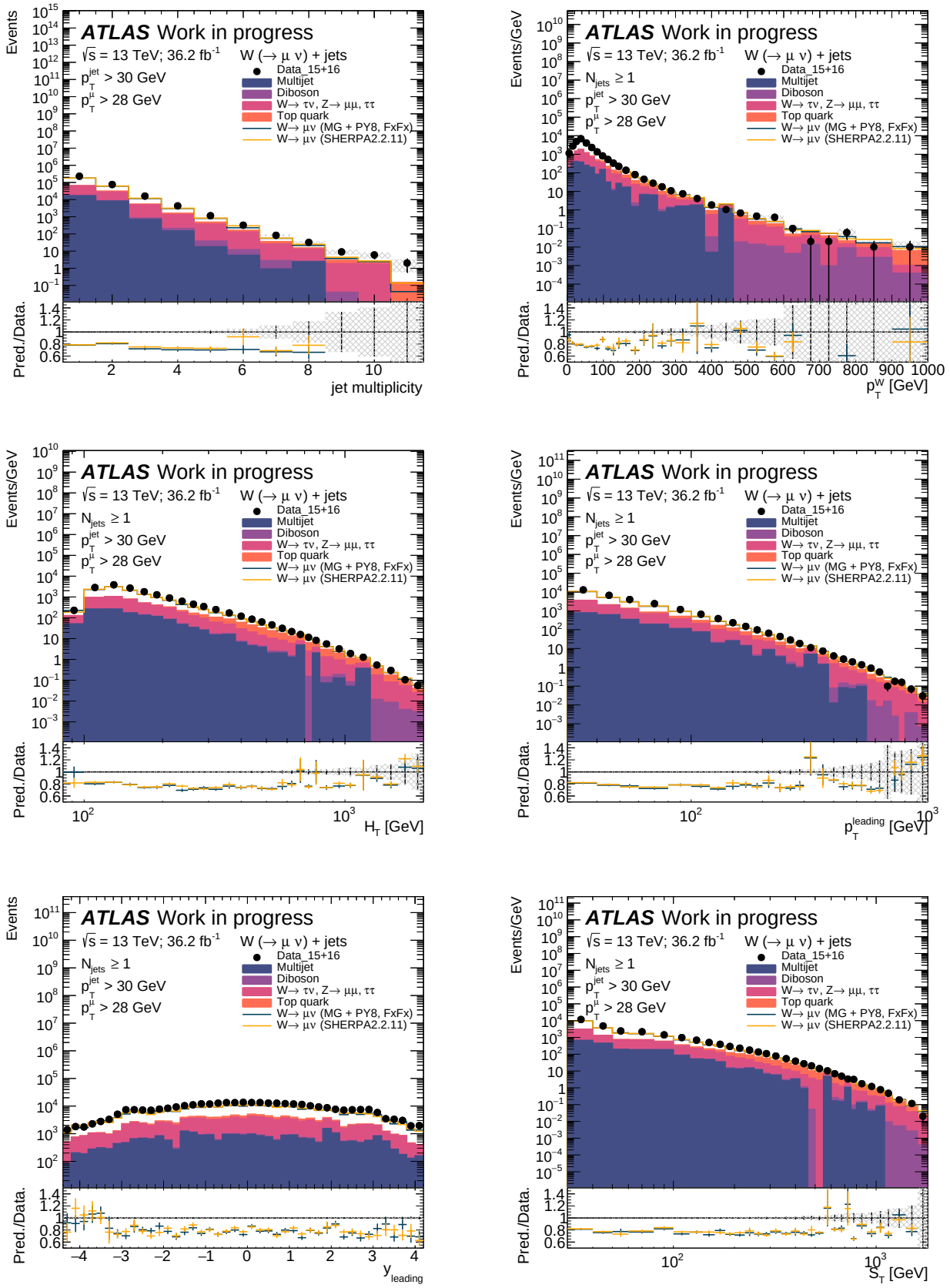


Figure 6.10: Data-MC comparison in the Validation Region, including the MJB estimated using real efficiencies calculated in the Validation Region.

Discussion the uncertainties associated with the data-driven background estimation will be presented in Chapter 8, while the MJB contribution to the Signal Region will be presented in Chapter 9.

Chapter 7

Unfolding

"The truth never set me free so I did it myself"

— Paramore, 'Careful'

As objects pass through and interact with the ATLAS detector, their kinematics become distorted with respect to their original state at the point of production. Due to this, cross-sections measured using objects reconstructed by the detector will be different from the 'true' underlying cross-sections predicted at particle-level. Therefore, to meaningfully interpret a measurement, one must to be able to translate from 'reconstruction-level' to 'particle-level'. This translation, or 'unfolding' problem is non-trivial; the interactions of individual particles with the detector are probabilistic in nature, and may have correlated impacts across many variables. This probabilistic nature or the unfolding problem lends itself naturally, however, to the application of Bayes' Theorem, which is discussed both conceptually and practically in Section 7.1.1 and Section 7.2 respectively. Finally, the first

unfolded $W \to \mu\nu$+jets differential cross-section measurement at 13 TeV is presented in Section 9.2.

7.1 Unfolding: How it works

The unfolding procedure involves the application of several different corrections, each intended to correct for a particular type of inefficiency introduced by the detector:

- **Fake fraction:** This accounts for events which **have not** been produced within the fiducial phase space, but have been reconstructed such that they **pass** the reconstruction-level selections. The fraction of such events can be quantified as $f_{\text{fake}} = \frac{N_{\text{reco}}^{\text{unmatched}}}{N_{\text{reco}}}$, where $N_{\text{reco}}^{\text{unmatched}}$ is the number of reconstructed events which are not matched to a truth-level event and N_{reco} is the number of reconstructed events. Such events should be subtracted from data before unfolding;

- **Bin-to-bin migrations:** This accounts for the migrations between bins of a particular distribution for events matched at both reconstruction and truth-level, due to the observable of interest having been reconstructed at a higher or lower value. These are measured using two-dimensional migration matrices, where a truth-level distribution is compared to the reconstruction-level distribution for events passing both sets of selections;

- **Reconstruction efficiency:** This accounts for events which **have** been produced within the fiducial phase space, but which have been reconstructed such that they

do not pass the reconstruction-level selections. The reconstruction efficiency can be quantified as $\epsilon = \frac{N_{\text{reco}}^{\text{matched}}}{N_{\text{gen}}}$ where $N_{\text{reco}}^{\text{matched}}$ is the number of reconstructed events which are matched to a truth-level event and N_{gen} is the number of generated events.

When performing the unfolding, the order of operations is as follows:

1. Backgrounds are subtracted from reconstruction-level data;

2. The background subtracted reconstruction-level data is corrected to account for the fake fraction;

3. The data is corrected to truth-level by inverting the migration matrices to correct for bin-to-bin migrations. This stage often requires some degree of regularisation - in this analysis, the 'iterative Bayesian' approach is chosen. This is described in detail in Section 7.1.1;

4. The final truth-level distributions are corrected to account for the reconstruction efficiency.

In practice, steps 2-4 above are performed using the `RooUnfold` package [128], using inputs derived from the $W \to \mu\nu$+jets signal MC; distributions of the relevant observables at reconstruction-level, truth-level and the migration matrices.

7.1.1 Iterative Bayesian Unfolding

An estimator $\hat{n}(T_i)$ for the number of events, n in truth bin i, T_i can be expressed [129] [130]:

$$\hat{n}(T_i) = \sum_j M_{ij} n(R_j), \tag{7.1}$$

where $n(R_j)$ is the number of events in the reconstructed bin j, R_j, and the 'unfolding matrix' M_{ij} is

$$M_{ij} = \frac{P(T_i|R_j)}{\sum_j P(R_j|T_i)},$$

where $P(T_i|R_j)$ is the probability an event from T_i is reconstructed in R_j. $P(T_i|R_j)$ is sometimes referred to as the 'response matrix', and is the probability that an event reconstructed in R_j originated from T_i. By application of Bayes' Theorem, $P(T_i|R_j)$ can be expressed as:

$$P(T_i|R_j) = \frac{P(R_j|T_i)n_0(T_i)}{\sum_i P(R_j|T_i)n_0(T_i)},$$

where $n_0(T_i)$ is known as the 'prior'; an initial estimate for the number of events in T_i. Therefore, Eq. 7.1 becomes:

$$\hat{n}(T_i) = \sum_j \frac{P(R_j|T_i)n_0(T_i)n(R_j)}{\sum_i P(R_j|T_i)n_0(T_i) \sum_k P(R_k|T_i)}. \tag{7.2}$$

In practice, the prior $n_0(T_i)$ is estimated using MC simulation of the signal, and the

response matrix $P(T_i|R_j)$ is calculated from an MC signal sample with both particle-level and reconstructed distributions.

As Eq. 7.2 exhibits dependence on $n_0(T_i)$, estimates of $\hat{n}(T_i)$ will be biased by the choice of prior. However, if Eq. 7.2 is iteratively applied with the choice of $n_0(T_i)$ updated with each iteration to be the estimate of $\hat{n}(T_i)$ from the previous, the dependence on the choice of prior is reduced with each iteration. Hence, the method is known as '**iterative Bayesian unfolding**'.

7.2 Unfolding the W+jets muon channel

Reconstructed distributions are unfolded to a reduced particle-level 'fiducial phase space'. This phase space is chosen such that it is kinematically similar to the reconstruction-level phase space described in Section 5.3. However, as these selections, summarised in Table 7.1 are applied at particle-level, any selections which apply only at reconstruction level are removed. This include the veto on b-jets in N_{jet} bins ≥ 3, meaning extrapolation from light-flavour and c quark jets at reconstruction-level to inclusive jets at particle level is being performed; however, this precedent has previously been set by ATLAS in the previous W+jets analysis at 8 TeV [21]. At particle level, leptons are 'dressed' meaning photons produced via Initial-State Radiation and Final-State Radiation are included with the lepton.

The reconstructed $W \to \mu\nu$+jets distributions presented in Section 9.1 are unfolded to

Lepton criteria	
Lepton p_{T}	$p_{\mathrm{T}} >$ 28 GeV
Lepton pseudorapidity	$\lvert\eta\rvert <$ 2.5
W **boson criteria**	
Lepton multiplicity	Exactly one, dressed
Missing transverse momentum	$E_{\mathrm{T}}^{\mathrm{miss}} >$25 GeV
Transverse mass	$\mathrm{m}_{\mathrm{T}} >$ 60 GeV
Jet criteria	
Jet p_{T}	$p_{\mathrm{T}} >$ 30 GeV
Jet rapidity	$\lvert y\rvert <$ 4.4

Table 7.1: Fiducial phase space selections for the W+jets analysis.

particle-level using iterative Bayesian unfolding as implemented in the `RooUnfold` package [128]. Before the final cross-sections can be unfolded, the backgrounds must be subtracted from the reconstruction-level data, and the particle-level prior and response matrices required by Eq. 7.2 must be produced. The truth-level distributions are calculated by applying the fiducial phase space selections in Table 7.1 to the Sherpa 2.2.11 signal MC sample. The production of the response matrices is discussed in further detail in Section 7.2.1. The correct implementation of the unfolding procedure is verified via a technical closure test; this is discussed and results presented in Section 7.2.2.

7.2.1 Response Matrices

The response matrices illustrate the bin-to-bin migrations which occur as particles interact with the detector before being reconstructed. They are calculated by filling a set of two-dimensional histograms with events that pass both the reconstruction-level and particle-level selections, with the particle-level distribution on the y-axis and the reconstruction-

level distribution on the x-axis. An additional matching procedure is applied to jets prior to an event entering the response matrix, to ensure the truth and reconstruction-level distributions contain the same set of objects:

- For each particle-level jet, t, $\Delta R_{r,t}$ is calculated with respect to each reconstruction-level jet r;
- Beginning with higher p_{T} jets, t is matched to the reconstruction-level jet closest in ΔR. Both t and r are subsequently removed from the list of candidate jets;
  - If there are not reconstruction-level jets within $\Delta R < 0.4$ of t, t is removed;
- This procedure is repeated until all particle-level jets have been matched. Any unmatched reconstruction-level jets are removed;
- For the jet-level observables $p_{\mathrm{T}}^{\mathrm{leading}}$ and y^{leading} , the reconstruction-level distributions are filled using the jet matched to the leading particle-level jet.

The jet matching procedure is not applied to the N_{jet} distribution, as this would result in the loss of information regarding just multiplicity migrations.

Response matrices calculated using the Sherpa 2.2.11 MC signal sample are shown in Figure 7.1, normalised to the number of reconstruction-level events in each bin. The bins are defined as described in Section 5.1; the choice of bins should retain as much granularity as possible while producing response matrices which are close to diagonal, as significant migrations will result in large uncertainties on the unfolded distribution.

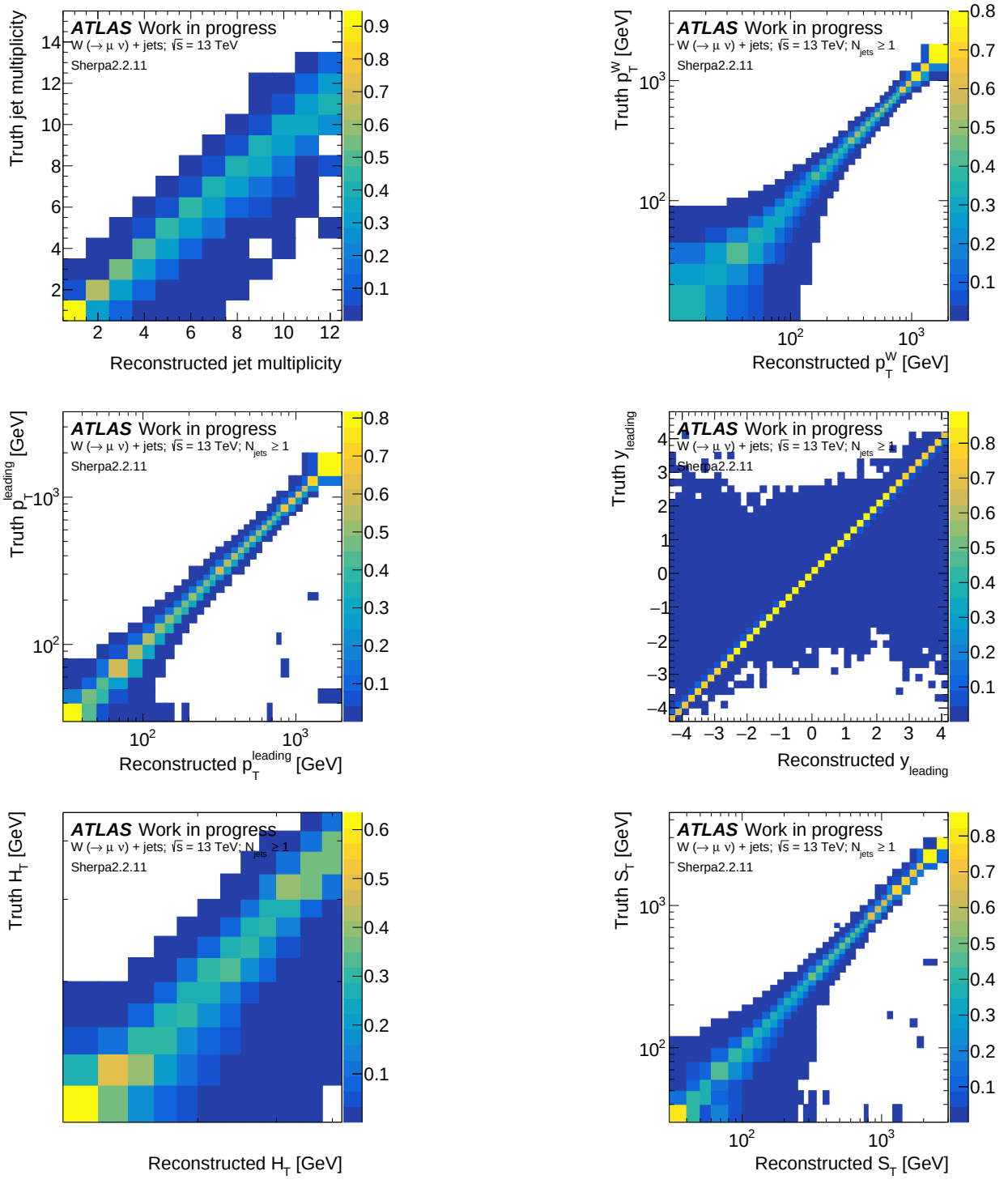


Figure 7.1: Response matrices produced using the Sherpa 2.2.11 signal MC sample. The z-axis is normalised to the number of events in each reconstruction-level (x-axis) bin.

In general, the matrices in Figure 7.1 are reasonably diagonal, with some small migrations observed at low values of p_{T} and throughout y^{leading} spectrum. More significant migrations are observed in the p_{T}^{W} spectrum , particularly at low p_{T} where finer granularity is required to correctly model the shape of the turnover. Migrations in the N_{jet} spectrum appear to bias the reconstruction of additional jets not present at particle-level; this may be caused by pile-up.

The systematic uncertainties described in Chapter 8 are also propagated to the final result via the response matrices. A varied response matrix is produced for each source of systematic uncertainty, which are each used to unfold the background-subtracted data to produce varied unfolded cross-sections, the difference of which from the nominal result is taken to be the propagated uncertainty.

7.2.2 Technical Closure

One simple, yet important test of the correct implementation of the unfolding procedure is the so-called 'technical closure test'. This test is performed by unfolding a reconstruction-level signal MC distribution using a truth distribution and transfer matrix constructed using the same sample. The unfolded result is expected to exhibit perfect agreement with that sample's particle-level distribution if the unfolding has been implemented correctly. Small deviations from perfect closure, due to statistical fluctuations, are permitted.

The results of the technical closure test, performed using the Sherpa 2.2.11 signal MC sample as shown in Figure 7.2 for up to 5 iterations of the Bayesian unfolding. All 5

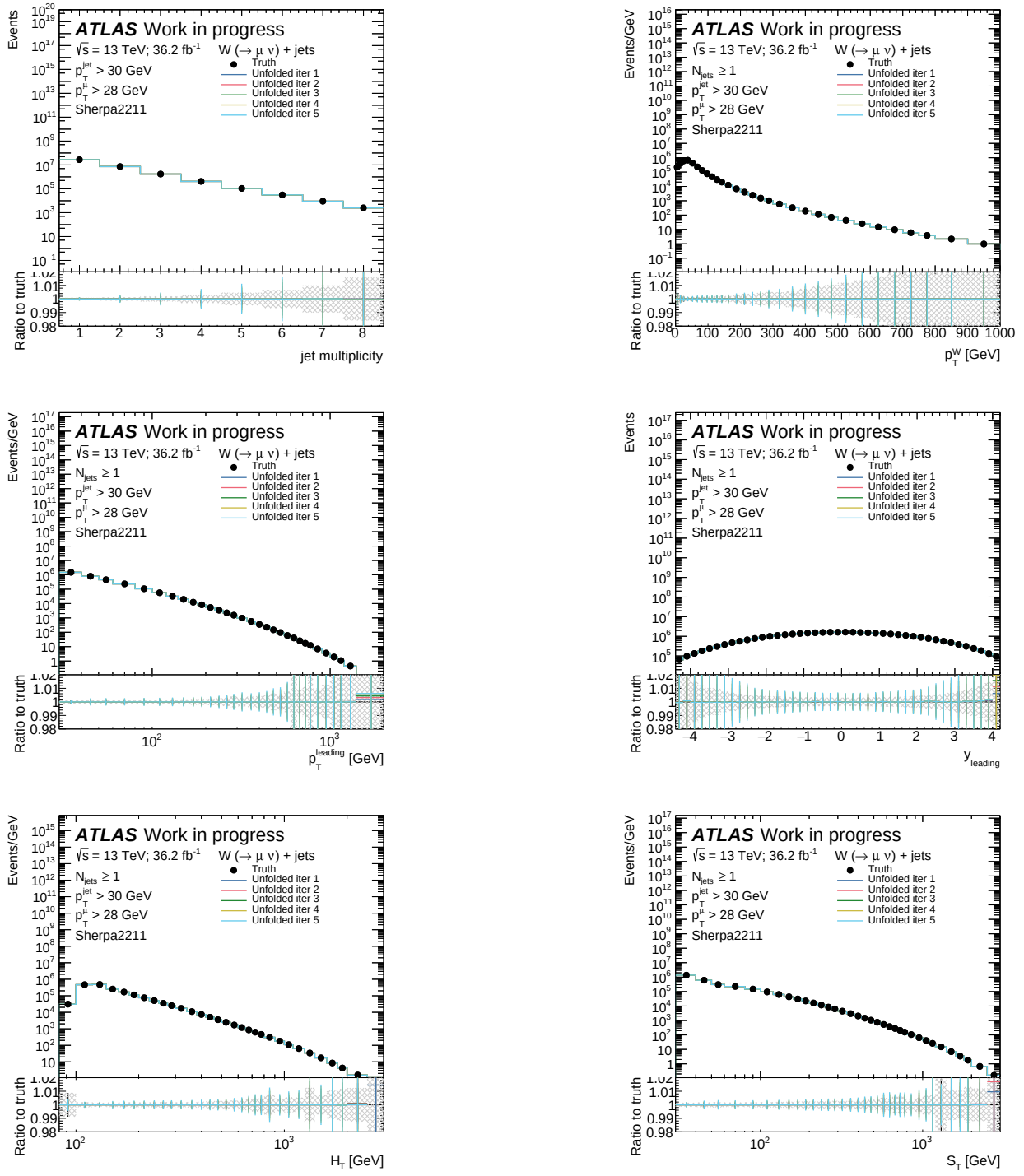


Figure 7.2: The results of the unfolding technical closure test performed using the Sherpa 2.2.11 signal MC sample.

iterations are perfectly overlaid with the Sherpa 2.2.11 particle-level distribution, except for the expected fluctuations in regions of low statistics.

With the response matrices and the technical closure test having been completed, the unfolded cross-sections can be produced. Some aspects of the unfolding remain to be completed before publication, including the optimisation of the number of unfolding iterations and the evaluation of the uncertainty associated with the unfolding.

Chapter 8

Uncertainties

"The truth is rarely pure and never simple"

— Oscar Wilde, 'The Importance of Being Earnest'

When presenting the results of a physical measurement, one must also be able to state their degree of confidence in the findings. All sources of uncertainty from the measurement must, therefore, be identified, quantified and propagated to the final result. Uncertainties in particle physics experiments can be classified into several categories. **Statistical uncertainties** originate from the variance inherent to counting statistics. **Experimental systematic uncertainties** arise from imperfections or limitations in the detector or object reconstruction. Additional systematic uncertainties may arise as a result of the analysis methods employed. Finally, **theoretical uncertainties** are introduced through the simulated MC signal and background samples.

In this chapter, the uncertainties associated with the W+jets muon channel analysis

will be described.

8.1 Statistical uncertainties

In order to make a probabilistic statement regarding the validity of the result of a counting experiment, the differences, or statistical fluctuations between the result obtained for a finite sample and the 'true' underlying value must be quantified. In the context of a particle physics analysis, these fluctuations or 'statistical uncertainties' have several origins:

- The number of data events observed in each bin;
- The number of events in each bin of the distributions used to calculate data-driven backgrounds, propagated through the background estimation;
- The number of events in the distributions used to carry out the unfolding procedure - this was discussed in further detail in Chapter 7.

. The number of events observed can usually be assumed to have an underlying Poisson distribution, which has an equal mean and variance λ, and therefore a standard deviation $\sigma_i = \sqrt{\lambda_i}$ in every bin i of a binned distribution of events. If a sufficient number of events are sampled, this Poisson distribution becomes Gaussian according to the Central Limit Theorem, meaning an approximation for the statistical uncertainty can be made using the number of observed events N in data:

$$\sigma_i^{\text{stat,data}} = \sqrt{N_i}. \tag{8.1}$$

In the case of the present analysis, due to the large production cross-section of W+jets at the LHC, a sufficient number of event are observed that statistical uncertainties do not dominate the measurement in all but the lowest statistics bins.

8.1.1 Statistical uncertainties from unfolding

As described in Chapter 7, the particle-level $W \to \mu\nu$+jets cross-sections are extracted from the measured reconstruction-level distributions via *iterative Bayesian unfolding* [129] [130]. The statistical uncertainty on the unfolded distribution from the finite statistics of the prior and the response matrices is propagated through the iterative Bayesian unfolding procedure using [131]:

$$\frac{\partial \hat{n}(T_i)}{\partial n(R_j)} = M_{ij} + \sum_k M_{ij} n(R_k) \left(\frac{1}{n_0(T_i)} \frac{\partial n_0(T_i)}{\partial \hat{n}(R_j)} - \sum_l \frac{\sum_m P(R_m|T_l)}{n_0(T_l)} \frac{\partial n_0(T_l)}{\partial \hat{n}(R_j)} M_{lk} \right), \quad (8.2)$$

where $\frac{\partial n_0(T_i)}{\partial \hat{n}(R_j)}$ is the uncertainty calculated from the previous iteration and 0 for the first iteration. From Eq. 8.2, it can be observed that the error from the unfolding is increased with each iteration. Therefore, the need to reduce bias from the prior must be balanced with this increase of statistical uncertainty by optimising the number of iterations.

The covariance matrix $V(\hat{n}(T_k), \hat{n}(T_l))$ of the unfolded distribution can be calculated using Eq. 8.2, and is found to be

$$V(\hat{n}(T_k), \hat{n}(T_l)) = \sum_{i,j=1^{n_{\mathrm{T}}}} \frac{\partial \hat{n}(T_k)}{\partial n(R_i)} V(n(R_i), n(R_j)) \frac{\partial \hat{n}(T_l)}{\partial n(R_j)}, \quad (8.3)$$

where n_{T} is the number of truth bins and $V(n(R_i), n(R_j))$ is the covariance matrix of the reconstructed distribution. The statistical uncertainty from unfolding in bin k of the unfolded distribution is therefore extracted from the diagonal element V_{kk}. For detailed discussion of the treatment of uncertainties associated with iterative Bayesian unfolding, see [130].

8.2 Systematic uncertainties

Experimental systematic uncertainties arise due to imperfections or limitations in the setup or operation of the experiment. Examples pertinent to experimental particle physics include uncertainties in object identification, calibration, or resolution, or statistical uncertainties propagated from the samples used in calibration procedures. Additional systematic uncertainties may also be introduced into a measurement based on the data analysis methods employed.

The individual sources of experimental systematic uncertainty in the W+jets muon channel, grouped based on their physical origin, are presented in Table 8.1, and are summarised below.

Muon reconstruction and calibration [63] [66] [132], in addition to the single muon trigger efficiency [121], contribute 14 individual sources of systematic uncertainty to the muon channel analysis:

- 2 associated with the unprescaled single muon trigger efficiencies;

Muon systematics	
MUON_EFF_TRIG_STAT	Statistical error on the muon trigger efficiency SF
MUON_EFF_TRIG_SYST	Systematic error on the muon trigger efficiency SF
MUON_EFF_RECO_STAT	Statistical error on the pT>15 GeV component of the muon reconstruction efficiency SF
MUON_EFF_RECO_SYST	Systematic error on the pT>15 GeV component of the muon reconstruction efficiency SF
MUON_EFF_RECO_STAT_LOWPT	Statistical error on the pT<15 GeV component of the muon reconstruction efficiency SF
MUON_EFF_RECO_SYST_LOWPT	Systematic error on the pT<15 GeV component of the muon reconstruction efficiency SF
MUON_EFF_ISO_STAT	Statistical error on the muon isolation SF
MUON_EFF_ISO_SYST	Systematic error on the muon isolation SF
MUON_EFF_TTVA_STAT	Statistical error on the muon track-to-vertex association SF
MUON_EFF_TTVA_SYST	Systematic error on the muon track-to-vertex association SF
MUON_CB	Variations in the muon track resolution for combined muons
MUON_SAGITTA_DATASTAT	Statistical error on the charge-dependant momentum corrections
MUON_SAGITTA_RESBIAS	Variations in the charge-dependent momentum scale from residual charge-dependent bias and data/MC non-closure
MUON_SCALE	Charge-independent variations in the muon momentum scale
$E_{\mathrm{T}}^{\mathrm{miss}}$ **systematics**	
MET_SoftTrk_ResoPara	Uncertainty from smearing the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term parallel to the direction of $p_{\mathrm{T}}^{\mathrm{hard}}$
MET_SoftTrk_ResoPerp	Uncertainty from smearing the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term perpendicular to the direction of $p_{\mathrm{T}}^{\mathrm{hard}}$
MET_SoftTrk_Scale	Uncertainty from scaling the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term scale up or down in the direction of $p_{\mathrm{T}}^{\mathrm{hard}}$
Jet systematics	
JET_EffectiveNP_Detector (1-2)	JES *in-situ* nuisance parameters due to detector modelling
JET_EffectiveNP_Mixed (1-3)	JES *in-situ* nuisance parameters with mixed origins
JET_EffectiveNP_Modelling (1-4)	JES *in-situ* nuisance parameters due to jet modelling
JET_EffectiveNP_Statistical (1-6)	JES *in-situ* nuisance parameters due to statistical uncertainties
JET_EtaIntercalibration_Modelling	JES η-intercalibration uncertainty due to jet modelling
JET_EtaIntercalibration_NonClosure_2018data	JES η-intercalibration uncertainty due to non-closure with 2018 data
JET_EtaIntercalibration_NonClosure_highE	JES η-intercalibration uncertainty due to non-closure with data for high-energy jets
JET_EtaIntercalibration_NonClosure_negEta	JES η-intercalibration uncertainty due to non-closure with data in the region $\eta <$ -2.4
JET_EtaIntercalibration_NonClosure_posEta	JES η-intercalibration uncertainty due to non-closure with data in the region $\eta >$ 2.4
JET_EtaIntercalibration_TotalStat	JES η-intercalibration statistical uncertainty
JET_Flavor_Composition	JES uncertainty in the proportion of quarks to gluons in calibration samples
JET_Flavor_Response	JES uncertainty in the response of gluon-initiated jets
JET_Pileup_OffsetMu	JES uncertainty of the modelling of μ in simulation
JET_Pileup_OffsetNPV	JES number of primary vertex-dependent pile-up subtraction uncertainty
JET_Pileup_RhoTopology	JES uncertainty on the simulated average pile-up density
JET_Pileup_PtTerm	Remaining JES uncertainty on pile-up p_{T} dependence
JET_PunchThrough_MC16	JES uncertainty on the punch-through correction in the Global Sequential Calibration
JET_SingleParticle_HighPt	JES uncertainty from single-particle and test-beam measurements for high-p_{T} jets
JET_JER_EffectiveNP (1-12restTerm)	JER nuisance parameters
JET_JER_DataVsMC_MC16	JER uncertainty accounting for differences in resolution between data and MC
jvt_SF	Uncertainty on the Jet-Vertex Tagger central pile-up tagging SF
fjvt_SF	Uncertainty on the Forward Jet-Vertex Tagger forward pile-up tagging SF
Pile-up systematics	
pileup	Uncertainty on the pile-up reweighting SF
b-Tagging systematics	
bTag_SF_DL1r_85_eigenvars_B (1-9)	Uncertainty on b-jet b-tagging efficiency
bTag_SF_DL1r_85_eigenvars_C (1-4)	Uncertainty on c-jet b-tagging efficiency
bTag_SF_DL1r_85_eigenvars_Light (1-4)	Uncertainty on light-flavour jet b-tagging efficiency
bTag_SF_DL1r_85_extrapolation	b-tagging efficiency for high-p_{T} jets
bTag_SF_DL1r_85_extrapolation_from_charm	b-tagging efficiency for high-p_{T} jets derived from a charm-enriched sample

Table 8.1: A summary of the experimental systematic uncertainties in the W+jets muon channel.

- 4 associated with muon reconstruction efficiency. This is derived in J/ψ and Z boson events [63] using the Tag and Probe method as discussed in Section 3.2.1, with the former contributing when muon $p_{\mathrm{T}} < 15$ GeV, and the latter when muon $p_{\mathrm{T}} > 15$ GeV. As a result, they each provide independent systematic components;

- 2 associated with muon isolation efficiency;

- 2 associated with muon track-to-vertex association;

- 1 associated with track resolution for 'combined' muons;

- 2 associated with muon charge-dependent momentum corrections;

- 1 associated with charge-independent uncertainties on the reconstructed muon momentum scale.

Most sources listed above have both a statistical component, reflecting the number of events in the distributions used to derive the scale factors, and a systematic component.

$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction [81] is associated with 3 sources of systematic uncertainty from the soft term only, as the uncertainties associated with the objects contributing to the hard term are included in the uncertainties associated with the same objects used for analysis. Two components originate from the resolution of the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term, one of which is smeared with a Gaussian profile parallel to the direction of the hard term, and the other in the perpendicular direction. The final uncertainty component originates from the response of the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term, as measured in a topology with no genuine $E_{\mathrm{T}}^{\mathrm{miss}}$.

Jet systematics primarily originate from the Jet Energy Scale (JES) and Jet Energy Resolution (JER) calibration [50] [76], in addition to uncertainties related to the calibration of pile-up tagging algorithms in the central [83] and forward [82] regions. In theory, all 115 individual nuisance parameters related to the JES calibration could be combined into one single uncertainty; however, many of these nuisance parameters are correlated and such a combination would result in the loss of valuable correlation information. Instead, several reduction schemes are employed by ATLAS to balance the number of JES nuisance parameters with the degree of correlation information retained. For the 'category reduction' scheme used the W+jets analysis, uncertainties related to the *in-situ* JES calibration which are one-dimensional functions of jet p_{T} , an eigenvector decomposition on the covariance matrix of these uncertainties is performed. Only the top 30 (including both up and down variations) are retained, with the final terms being a residual term of the remaining nuisance parameters. The category reduction scheme groups these reduced nuisance parameters according to their origin; detector, jet modelling, statistical and mixed. In total, there are 22 sources of jet-related systematic uncertainty:

- 4 associated with the JES *in-situ* calibration, comprised of 15 sets of up/down variations;
- 6 associated with the JES η-intercalibration, whereby the well-measured jet energy response derived in the central region is extrapolated to the forward region;
- 2 associated with the relative quark-gluon proportion in the JES calibration samples;
- 4 related to the JES pile-up subtraction;

- 2 miscellaneous JES systematics;
- 1 associated with the JER smearing, comprised of 12 sets of up/down variations;
- 1 associated with jet resolution differences between data and MC;
- 2 associated with the data-MC efficiency correction scale factors from pile-up tagging.

One source of systematic uncertainty is associated with '**pile-up reweighting**' [133]. This is the process by which pile-up is simulated in ATLAS MC samples. A variable number of additional inelastic pp collisions are simulated at the interaction point, and MC events are reweighted so that the mean number of interactions per bunch crossing agrees with observations in data from the year to which the MC sample corresponds.

As a veto is placed on events with $\geq$3 jets which have at least one jet from a b-quark which enter the N_{jet} distribution, the uncertainties associated with the b-tagging algorithm calibration [79] [80] must be accounted for in these bins. Similar to the JES uncertainties, a Principal Component Analysis is used to reduce the total number of flavour tagging uncertainties to a smaller number of nuisance parameters. Different sets of nuisance parameters are provided for each jet flavour:

- 9 associated with the b-tagging efficiency of b jets;
- 4 associated with the b-tagging efficiency of c jets;
- 4 associated with the b-tagging efficiency of light flavour jets;

- 2 associated with the extrapolation of data driven calibrations to larger values of jet p_{T} using MC simulation.

In practice, systematic uncertainties are evaluated using MC before being transferred to data distributions. The impact of each systematic on the final measurement is evaluated by varying either the event weight or kinematics, depending on the type of uncertainty, and repeating the analysis in full to produce a varied distribution. Typically, uncertainties on event reweighting scale factors only affect the normalisation of distributions, whereas others which affect event kinematics may impact several quantities at once, in addition to which objects and events pass selections. The uncertainty from an individual systematic source on the measured distributions is evaluated by taking the difference between the varied and nominal distributions. The individual uncertainty components in each bin are then added in quadrature, separately depending on whether they have an up or down effect on the measured cross-section, to produced the total systematic uncertainty bands.

In addition to the experimental systematic uncertainties discussed above, further sources of systematic uncertainty should be considered, originating from the **multijet background** estimation. These include:

- The propagation of the systematic uncertainties discussed above through the matrix method. In principle, this should be performed by repeating the background estimate using varied distributions for each source of systematic uncertainty. However, as Table 9.1 will show, the MJB is a sufficiently small background that shape differences are not expected to strongly contribute to the overall systematic uncertainty in the

Signal Region. Therefore, it has been agreed within the W+jets analysis team that a 20% normalisation envelope is a sufficiently conservative approximation for the full propagation[1];

- A two-point systematic for the choice of $W \to \mu\nu$+jets signal generator used in the efficiency calculation;

- The propagation of the statistical uncertainties of the distributions used to calculate the real and fake efficiencies.

The contribution from each group of systematics described above, including the uncertainties associated with the MJB estimate, to the total W+jets experimental systematic uncertainty is illustrated in Figure 8.1. The binning scheme chosen is the optimised set of bins for the unfolded distributions, as discussed in Section 5.1. The total systematic uncertainty across all distributions is of order 10-20%, with jet-related systematics dominating in most bins of all distributions; this is with the exception of the low p_{T}^{W}, high y^{leading} and mid $p_{\mathrm{T}}^{\mathrm{leading}}$, H_{T} and S_{T} regimes, where uncertainties originating from the MJB estimate dominate.

The total experimental systematic uncertainty is also compared to the statistical uncertainty and, in all but the highest jet p_{T} bins, the measurement is dominated by the systematic uncertainties. The addition of the remainder of the Run 2 data is expected to improve the measurement precision in these bins and possibly make higher p_{T} bins accessible, but will have a limited impact on the total uncertainty in regions where sys-

[1] However; this approach remains under discussion, and an alternative strategy may ultimately be employed in the published paper.

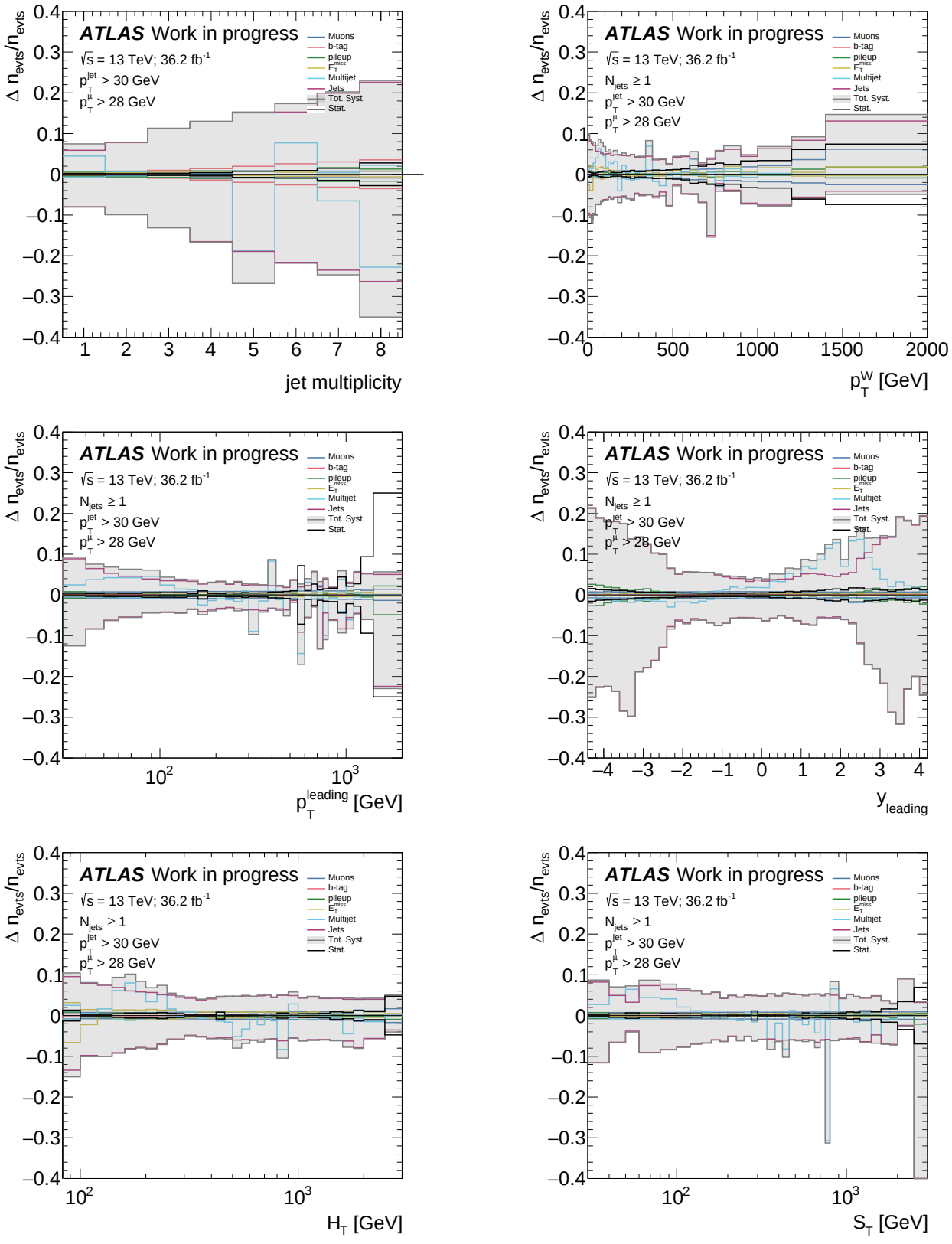


Figure 8.1: The contribution to the total experimental systematic uncertainty (grey) from each group of systematic uncertainty, in each of the distributions measured. The statistical uncertainty (black) is also shown.

tematic uncertainties dominate. The extra data is also expected to reduce the statistical uncertainties from the MJB estimation.

Additionally, the uncertainty on the **luminosity** measurement in ATLAS [30] for Run 2 must be considered. The absolute scale of the luminosity is measured using low-luminosity Van der Meer scans. This is then extrapolated into the high-luminosity regime of the high intensity pp collisions used for physics analyses using the using various methods, including the LUCID detector, the Beam Conditions Monitor, track counting methods and calorimeter algorithms. The largest contribution to the uncertainty on the luminosity measurement is from this extrapolation, followed by components from magnetic nonlinearities and beam-beam interactions. A study of the full Run 2 luminosity delivered a relative uncertainty of 0.83%, compared to 1.8% for the 7 TeV luminosity measurement and 1.9% for 8 TeV. While the full Run 2 dataset has not yet been implemented in the W+jets analysis, the 0.83% luminosity uncertainty value is used as an approximation of the luminosity uncertainty for the 2015+2016 data.

8.3 Theoretical uncertainties

Uncertainties are also introduced into a particle physics analysis through the MC generators used to model the signal, backgrounds and perform the unfolding. These uncertainties originate from a variety of sources, including the values of fundamental constants, the errors on tuning parameters, modelling assumptions, and uncertainties from the phenomenological models underlying the simulation of the hadronisation and the underlying

event. Additionally, uncertainties are introduced due to the choice of factorisation and renormalisation scales, μ_{F} and μ_{R} respectively. These are quantified by varying the scale choice, usually by factors of 0.5 and 2, according to the standard seven-point variation. Furthermore, the PDFs convolved with parton-level predictions to generate particle-level predictions have their own associated uncertainties which must be propagated to an analysis.

In practice, simulations store the impact of variation by each source of uncertainty as a set of varied generator weights. Similarly to the procedure for systematic uncertainties, to evaluate impact of each source of theoretical uncertainty on the final measured distributions, the full analysis is repeated using the varied Monte-Carlo weights. Additional care must also be taken to use the varied sum of generator weights when using Eq. 5.1 to perform the luminosity reweighting.

The full treatment of the theoretical uncertainties is still to be implemented in the W+jets muon channel. When this is ultimately performed, the theoretical uncertainty will be reported as an additional uncertainty alongside the unfolded cross-section results.

Chapter 9

Results

"I was taught that the way of progress is neither swift nor easy."

— Marie Curie

In this chapter, the measurement of W+jets production at $\sqrt{s}$=13 TeV in the $W \to \mu\nu$ decay channel are presented. Reconstruction-level results are shown in Section 9.1, while the unfolded differential cross-sections are presented and discussed in Section 9.2.

9.1 Reconstruction-level distributions

With the MC-driven backgrounds and data-driven backgrounds having been calculated in Chapter 6, and the majority of statistical and systematic uncertainties having been evaluated in Chapter 8, a comparison of each measured data distribution in the Signal

Region and the total Standard Model prediction can be produced, and are presented in Figure 9.1. The proportion of each background stated as a percentage of the number of data events for each value of inclusive jet multiplicity is presented in Table 9.1.

In Figure 9.1, two Standard Model predictions are presented: One with the $W \rightarrow \mu\nu$+jets signal modelled using Sherpa 2.2.11, and the other using Madgraph+Pythia8 with FxFx merging. In general there is good agreement within the total uncertainties of the predictions with data. This agreement breaks down, however, at large values of $p_{\mathrm{T}}^{\mathrm{leading}}$, H_{T} and S_{T} where both predictions overestimate the data outside the uncertainty bands. These same three distributions also deviate from the data at low p_{T} to the border of the uncertainty bands. In the N_{jet} distribution, both predictions agree well up to 4 jets, at which point Madgraph+Pythia8 prediction begins to underestimate the data while the Sherpa 2.2.11 prediction overestimates the data. In the p_{T}^{W} distribution, while the Sherpa prediction agrees with data within uncertainties over the full spectrum, several bins in the Madgraph prediction deviate outside uncertainties from 400 GeV. In all distributions, fluctuations in the MJB due to estimation using prescaled data, as described in Section 6.2.4, are apparent and emphasise the need for smoothing in the final result presented in the paper.

The uncertainties discussed in Section Chapter 8 are also included on the plots; the vertical black lines represent the statistical uncertainties, while the grey bands represent the statistical uncertainties, full experimental systematic uncertainties including those associated with the MJB estimate, and luminosity uncertainties approximated using the full Run 2 uncertainty of 0.83%. Theoretical uncertainties are not included; as discussed in

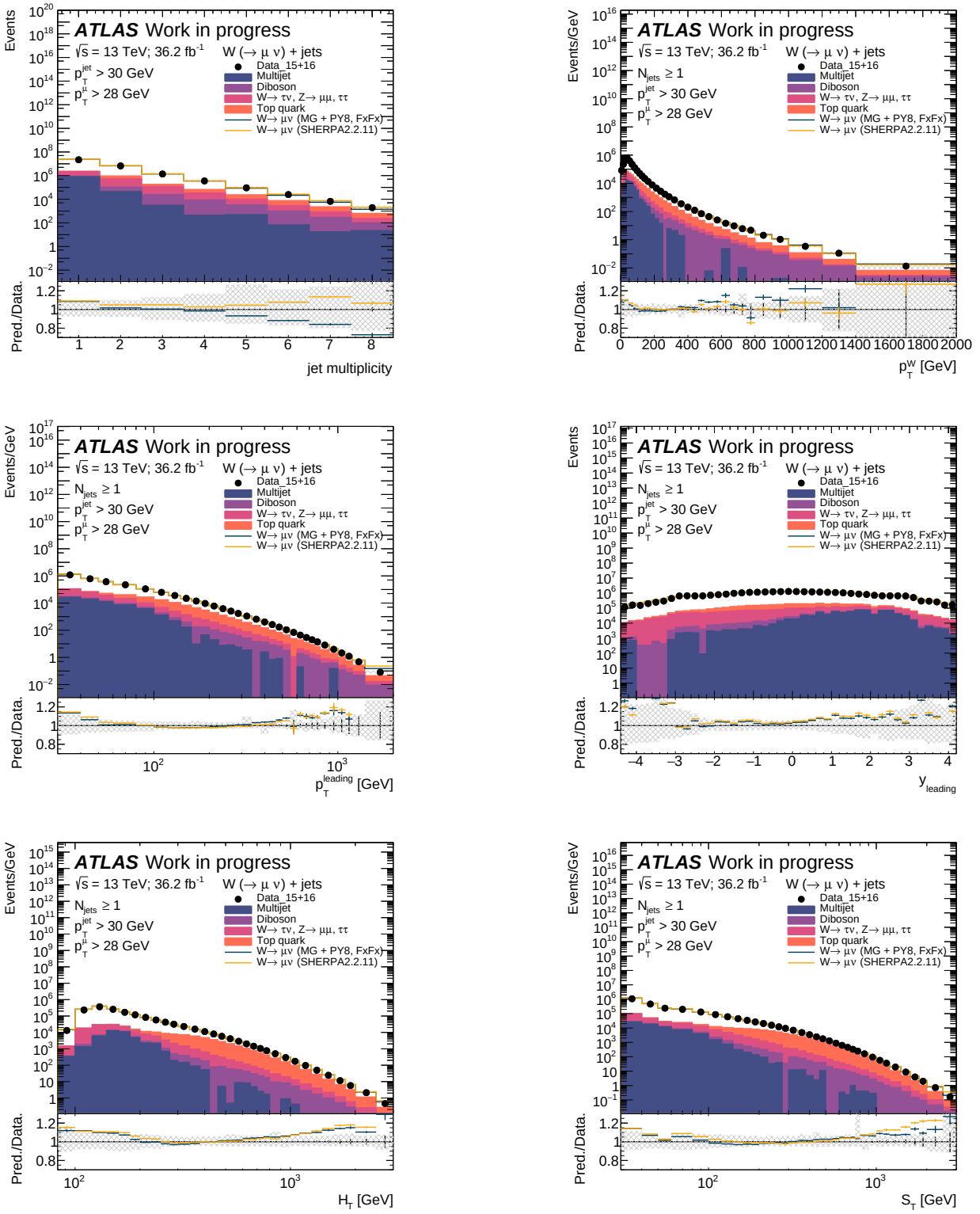


Figure 9.1: A comparison between data and two total Standard Model predictions in the Signal Region; one with the $W \to \mu\nu$+jets signal modelled using Madgraph+Pythia8 and FxFx merging (blue) and one using Sherpa 2.2.11 (yellow). Statistical uncertainties are represented by black vertical bands, while full statistical + systematic uncertainties in addition to the luminosity uncertainty, are represented by the grey bands.

Section 8.3, these will ultimately be expressed as an additional uncertainty alongside the unfolded cross-sections when they are available.

N_{jet}	$W \to \mu\nu$ (MG+PY8)	$W \to \mu\nu$ (SHERPA)	Top quark	$Z \to \mu\mu$ +jets	$W \to \tau\nu$ +jets	Multijet	Diboson	$Z \to \tau\tau$
	%	%	%	%	%	%	%	%
$\geq$1	97.8	98.4	0.7	4.2	1.7	3.7	0.4	0.1
$\geq$2	87.4	90.6	6.2	4.3	2.0	0.7	0.9	0.2
$\geq$3	87.1	91.6	4.5	4.7	2.2	0.2	1.6	0.2
$\geq$4	79.5	83.8	9.4	4.7	2.1	0.1	2.3	0.2
$\geq$5	67.4	78.8	14.8	4.8	2.3	0.5	3.1	0.3
$\geq$6	57.2	76.8	19.1	5.1	2.4	0.3	3.8	0.3
$\geq$7	49.4	78.9	21.8	5.4	2.7	0.3	4.1	0.3
$\geq$8	40.3	74.2	20.4	5.1	1.7	1.2	4.0	0.3

Table 9.1: The Signal Region contribution from each signal generator and background, expressed as a percentage of the number of data events, for different jet multiplicities up to 8 jets.

Additionally, the proportion of $W \to \mu\nu$+jets signal and each background in the Signal Region, expressed as a percentage of data events, is presented in Table 9.1 for different inclusive jet multiplicities up to 8 jets. In lower jet multiplicity bins, the Signal region consists of $>$ 85 % W+jets signal, declining sharply with jet multiplicity for the Madgraph+Pythia8 prediction and more gradually for Sherpa 2.2.11. The top quark background, while a relatively small contribution at lower jet multiplicities, becomes a significant $\sim$ 20% background at higher jet multiplicities, despite the veto placed on b-jets for $N_{\text{jet}} \geq 3$. The $Z \to \mu\mu$+jets background contributes approximately 5% across all jet multiplicities, which is slightly larger than the $\sim$ 1-4% contribution observed in the most recent ATLAS W+jets muon channel measurement at 7 TeV during Run 1 [94]. However, the present analysis has a slightly looser p_{T} threshold for the second muon veto; 10 GeV for this measurement vs 7 GeV for the 7 TeV analysis. Finally, the MJB contribution to the

Signal Region is somewhat smaller than the $\sim$ 2-10% background observed in the muon channel of the 7 TeV analysis; however, these may not be directly comparable as the 7 TeV measurement had a lower m_{T}^{W} >40 GeV cut compared to the m_{T}^{W} >60 GeV selection implemented in this 13 TeV analysis, and utilised jets using calorimeter topoclusters only compared to the Particle Flow jets made available in Run 2.

9.2 Unfolded cross-sections

In this section, the reconstruction-level data distributions presented in Section9.1, with the backgrounds subtracted, are unfolded using the Sherpa 2.2.11 signal MC sample. A choice of 3 unfolding iterations is made, based on studies performed in the electron channel. The unfolded distributions are weighted by a factor of 1/36.2 fb^{-1} to extract the differential cross-sections from the number of unfolded events (see Eq. 2.1). Additionally distributions that are a function of p_{T} are scaled to the bin width.

The systematic uncertainties described in Section 8.2, including the 0.83% luminosity uncertainty are propagated to the unfolded differential cross-sections, as described in Section 7.2.1. As discussed in Section 8.3, the theoretical uncertainties, which are yet to be calculated for the muon channel due to time constraints, will be presented alongside the statistical and systematic uncertainties in Table 9.2 and Table 9.3. These tables present the measured charge-independent production cross-section for W bosons in association with jets from pp collisions at $\sqrt{s}$=13 TeV, in the $W \rightarrow \mu\nu$ decay channel, for inclusive and exclusive jet multiplicities respectively, up to 8 jets. As discussed in Section

9.1, the measurement is dominated by systematic uncertainties for all jet multiplicities.

$N_{\text{jets}}^{\text{incl.}}$	σ [pb]	Stat. unc. (%)	Syst. unc. up (%)	Syst. unc. down (%)
$\geq$1	9461.57	0.2	18.8	17.0
$\geq$2	2573.15	0.1	28.9	27.3
$\geq$3	615.42	0.2	41.7	40.6
$\geq$4	153.64	0.3	52.7	51.5
$\geq$5	39.30	0.8	64.2	61.6
$\geq$6	10.75	1.1	72.0	70.8
$\geq$7	3.24	2.0	103.5	77.8
$\geq$8	1.23	3.3	211.6	85.5

Table 9.2: The charge-independent production cross-section for W bosons in association with jets from pp collisions at $\sqrt{s}$=13 TeV, in the $W \to \mu\nu$ decay channel, for inclusive jet multiplicities up to N_{jet} $\geq$8. The associated statistical and systematic uncertainties are also presented as a percentage of the measured cross-sections.

$N_{\text{jets}}^{\text{excl.}}$	σ [pb]	Stat. unc. (%)	Syst. unc. (%) up	Syst. unc. (%) down
=1	688.84	0.2	15.6	13.6
=2	195.77	0.1	24.9	23.2
=3	46.18	0.2	38.1	37.0
=4	11.43	0.3	49.2	48.0
=5	2.85	1.0	63.7	58.3
=6	0.75	1.3	66.6	68.0
=7	0.20	2.6	71.4	74.1
=8	. 0.06	4.6	118.0	79.5

Table 9.3: The charge-independent production cross-section for W bosons in association with jets from pp collisions at $\sqrt{s}$=13 TeV, in the $W \to \mu\nu$ decay channel, for exclusive jet multiplicities up to N_{jet}=8. The associated statistical and systematic uncertainties are also presented as a percentage of the measured cross-sections.

Figures 9.2-9.7 present, for each observable, the W boson charge-independent differential cross-section, the W^+ differential cross-section, the W^- differential cross-section and the ratio of the W^+/W^- differential-cross sections. The latter are important measurements, as many systematic uncertainties cancel in the ratio, and have historically been

powerful inputs to ATLAS PDF fits - see Chapter 10. The measured differential cross-sections are compared to the particle-level predictions from both the Sherpa 2.2.11 and Madgraph + Pythia 8 FxFx generators.

The results for the charge-independent measurements are similar to those presented at reconstruction-level in Section 9.1. Generally, agreement between the measured cross-sections and the two predictions is observed within the uncertainty bands, with the exception of the very high-p_{T} regime, or for very forward jets. While still technically within the uncertainty bands, it should also be noted that both generators systematically over-predict the data at low-p_{T} , and that, from N_{jet} $\geq$6, the two predictions diverge; Sherpa 2.2.11 begins to over-predict the data while MGPy8+FxFx underestimates it.

The ratios of the charge-dependent differential cross-sections also provide some insights. As expected, the ratio causes many systematic uncertainties cancel, leading to significant reductions in the total uncertainty, except in bins with low statistics where the ratio compounds the larger statistical errors; a prominent example is the p_{T}^{W} spectrum, where no meaningful comparison can be made above 800 GeV. For the N_{jet} and y^{leading} distributions, the W^{+}/W^{-} cross-section ratio agrees very with the data. However, for distributions which are sensitive to jet p_{T} , H_{T} ,S_{T} and $p_{\mathrm{T}}^{\mathrm{leading}}$, both predictions diverge strongly from the data, underestimating the measured ratio strongly at high-p_{T} .

Furthermore, the measured differential cross-sections and cross-section ratios are available for inclusive jet multiplicities larger than N_{jet} $\geq$1. An example is presented in Figure 9.8, where the charge-independent differential cross-section as a function of H_{T} is

presented for inclusive jet multiplicities a) $N_{\mathrm{jet}} \geq$2, (b) $N_{\mathrm{jet}} \geq$3, (c) $N_{\mathrm{jet}} \geq$4 (d) $N_{\mathrm{jet}} \geq$5.

In summary, this chapter has presented an interim result for W+jets production at $\sqrt{s}$=13 TeV in the muon channel. A limited number of further steps, which will be performed by the author, remain before the result is considered 'complete' for publication. These include:

- Performing a fit to smooth the MJB estimate, and understanding the underestimation in the Validation Region;
- Evaluating the theoretical uncertainties, and the unfolding non-closure uncertainty. The latter will simultaneously optimise number of unfolding iterations;
- Evaluating the statistical correlations between observables using the bootstrap method;
- The combination of the muon channel with the electron channel, which has been prepared by analysts other than the author;
- Using the measured cross-sections to constrain proton structure via a QCD analysis.

Additionally, the addition of the remaining $\sim$100 fb^{-1} of Run 2 data would improve the precision of the the high-p_{T} regime measurement, and of the cross-section ratios, where statistical uncertainties dominate. However, at the time of writing, this is intended to be performed in a follow-up paper, which will also include an expanded list of observables and updated unfolding techniques.

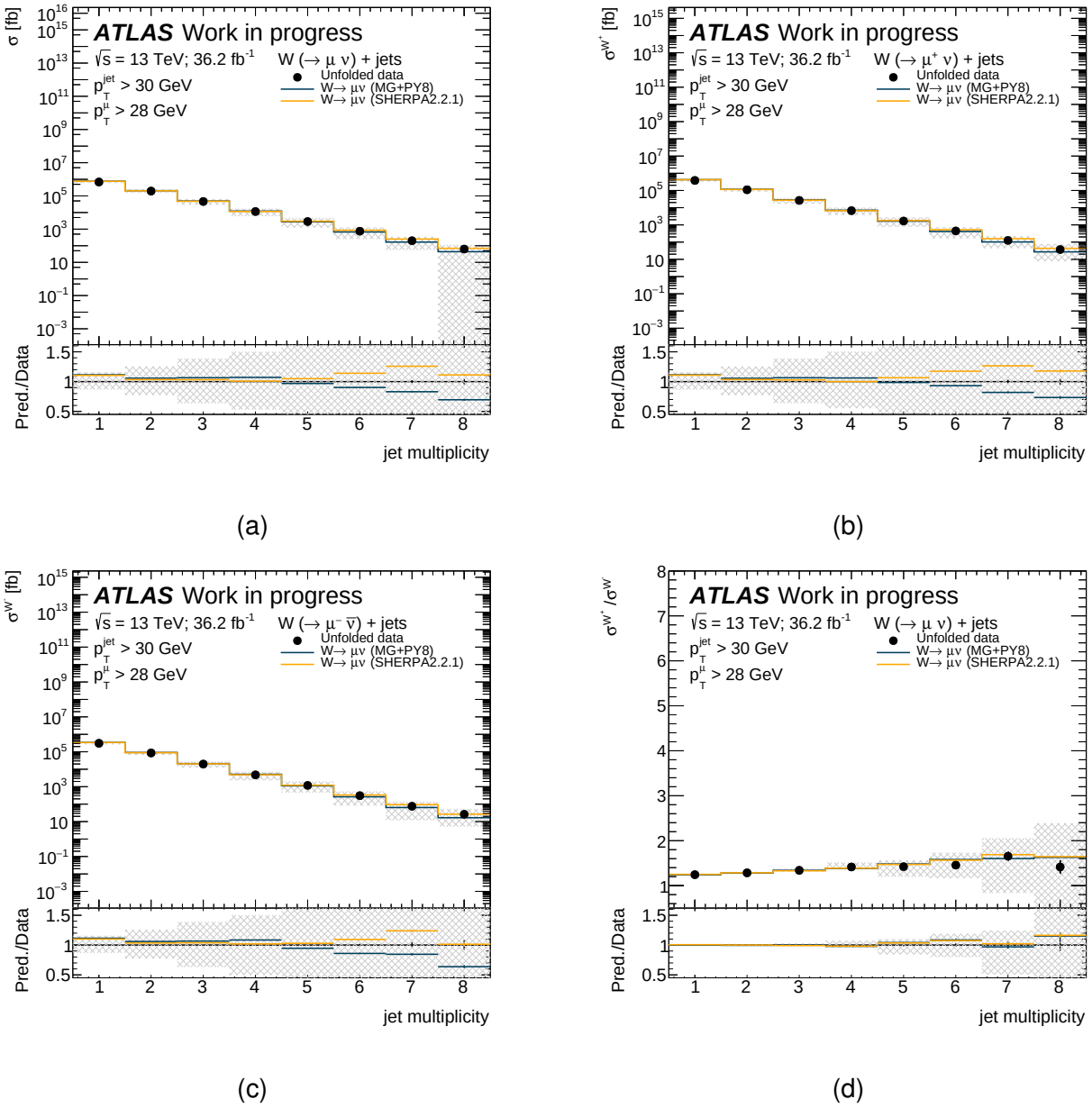


Figure 9.2: The unfolded cross-sections as a function of N_{jet} for (a) charge-independent W+jets production in the $W \to \mu\nu$ channel, (b) $W^+ \to \mu\nu$+jets (c) $W^- \to \mu\nu$+jets (d) the W^+/W^- ratio of differential cross sections. All results are compared to two predictions, generated using Madgraph+Pythia8 with FxFx merging (blue) and one using Sherpa 2.2.11 (yellow). Statistical uncertainties are represented by black vertical bands, while full statistical + systematic uncertainties, in addition to the luminosity uncertainty, are represented by the grey bands.

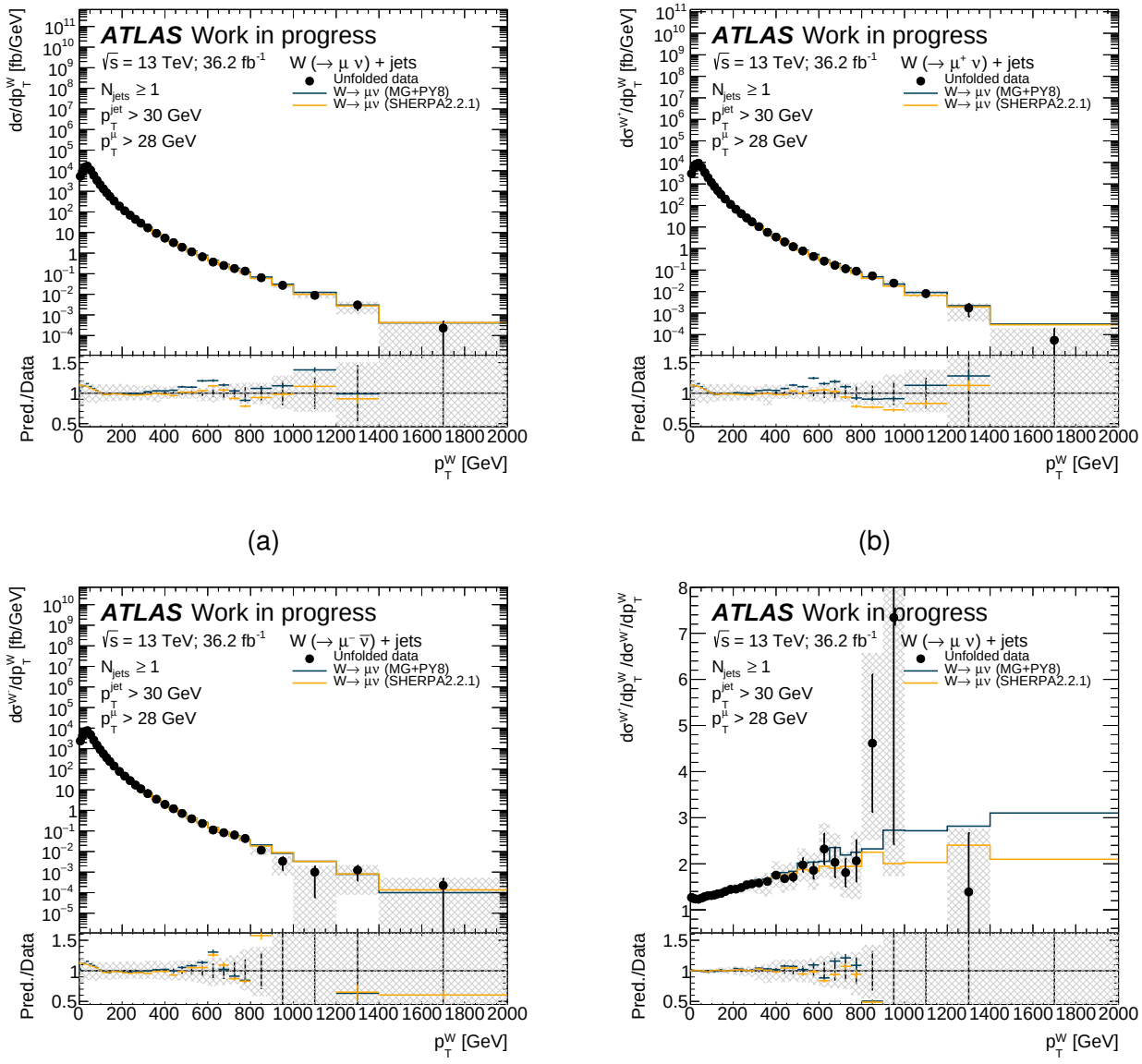


Figure 9.3: The unfolded cross-sections as a function of p_{T}^{W} for (a) charge-independent W+jets production in the $W \to \mu\nu$ channel, (b) $W^{+} \to \mu\nu$+jets (c) $W^{-} \to \mu\nu$+jets (d) the W^{+}/W^{-} ratio of differential cross sections. All results are compared to two predictions, generated using Madgraph+Pythia8 with FxFx merging (blue) and one using Sherpa 2.2.11 (yellow). Statistical uncertainties are represented by black vertical bands, while full statistical + systematic uncertainties, in addition to the luminosity uncertainty, are represented by the grey bands.

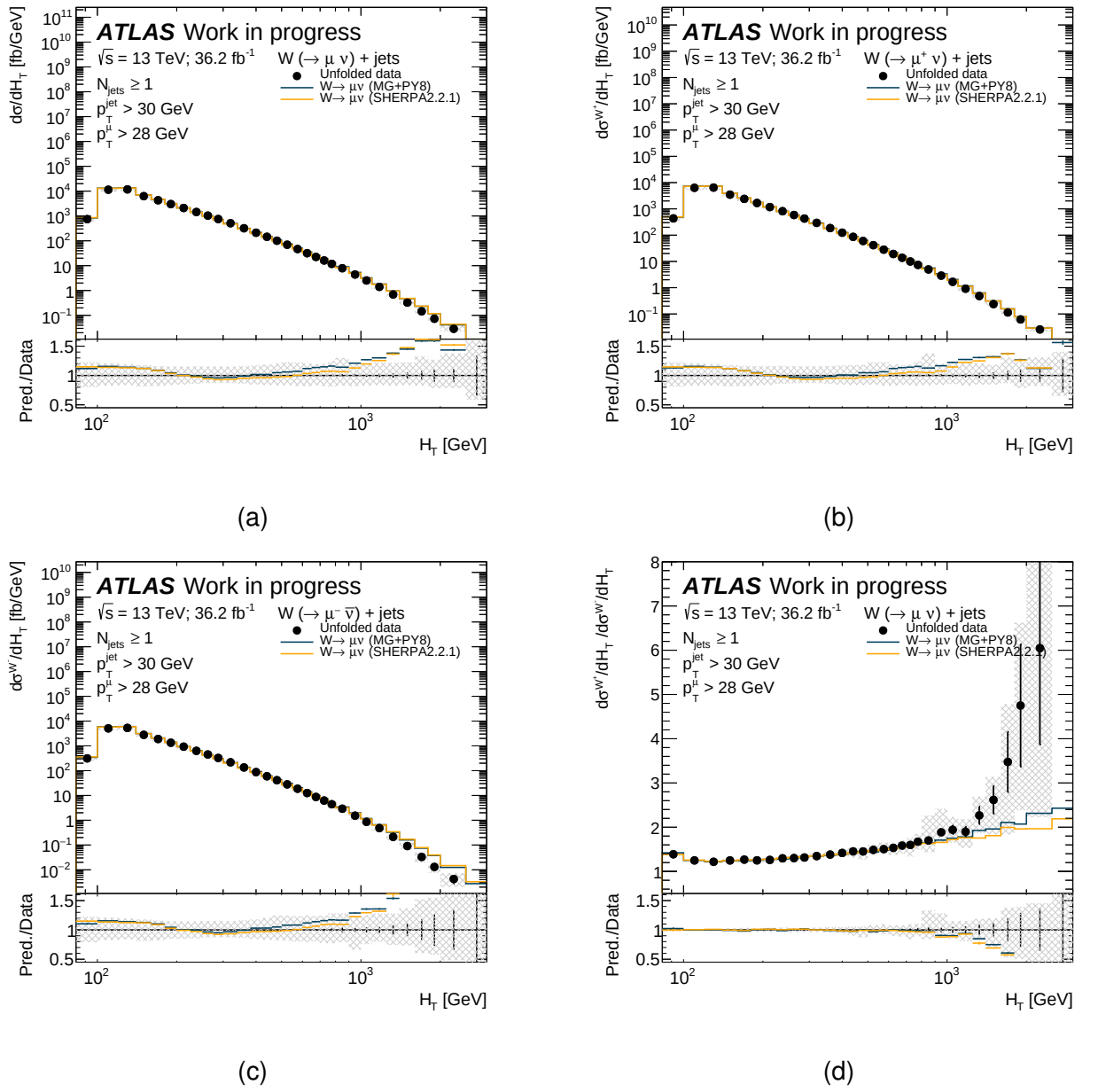


Figure 9.4: The unfolded cross-sections as a function of H_{T} for (a) charge-independent W+jets production in the $W \to \mu\nu$ channel, (b) $W^+ \to \mu\nu$+jets (c) $W^- \to \mu\nu$+jets (d) the W^+/W^- ratio of differential cross sections. All results are compared to two predictions, generated using Madgraph+Pythia8 with FxFx merging (blue) and one using Sherpa 2.2.11 (yellow). Statistical uncertainties are represented by black vertical bands, while full statistical + systematic uncertainties, in addition to the luminosity uncertainty, are represented by the grey bands.

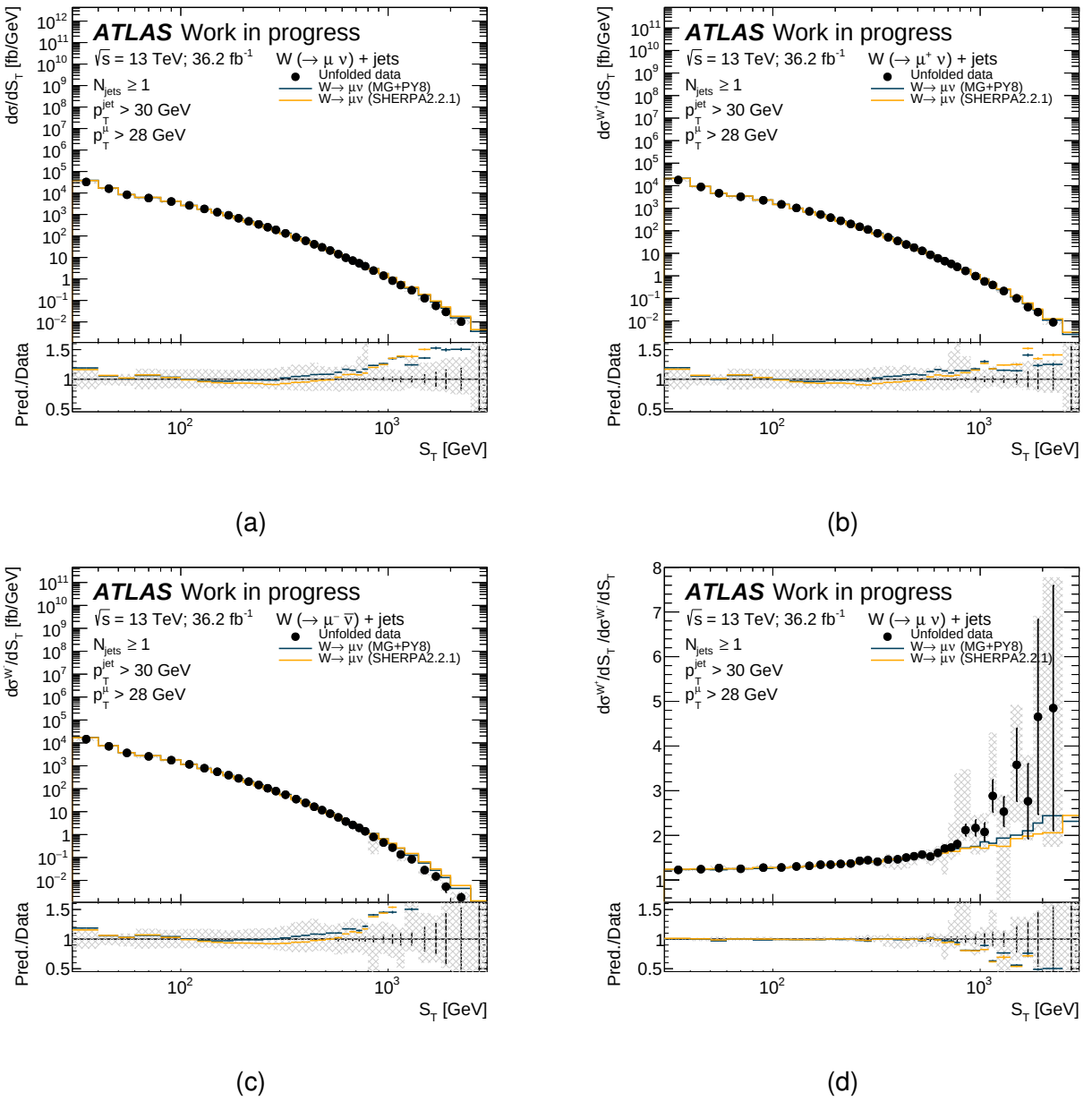


Figure 9.5: The unfolded cross-sections as a function of S_{T} for (a) charge-independent W+jets production in the $W \to \mu\nu$ channel, (b) $W^{+} \to \mu\nu$+jets (c) $W^{-} \to \mu\nu$+jets (d) the W^{+}/W^{-} ratio of differential cross sections. All results are compared to two predictions, generated using Madgraph+Pythia8 with FxFx merging (blue) and one using Sherpa 2.2.11 (yellow). Statistical uncertainties are represented by black vertical bands, while full statistical + systematic uncertainties, in addition to the luminosity uncertainty, are represented by the grey bands.

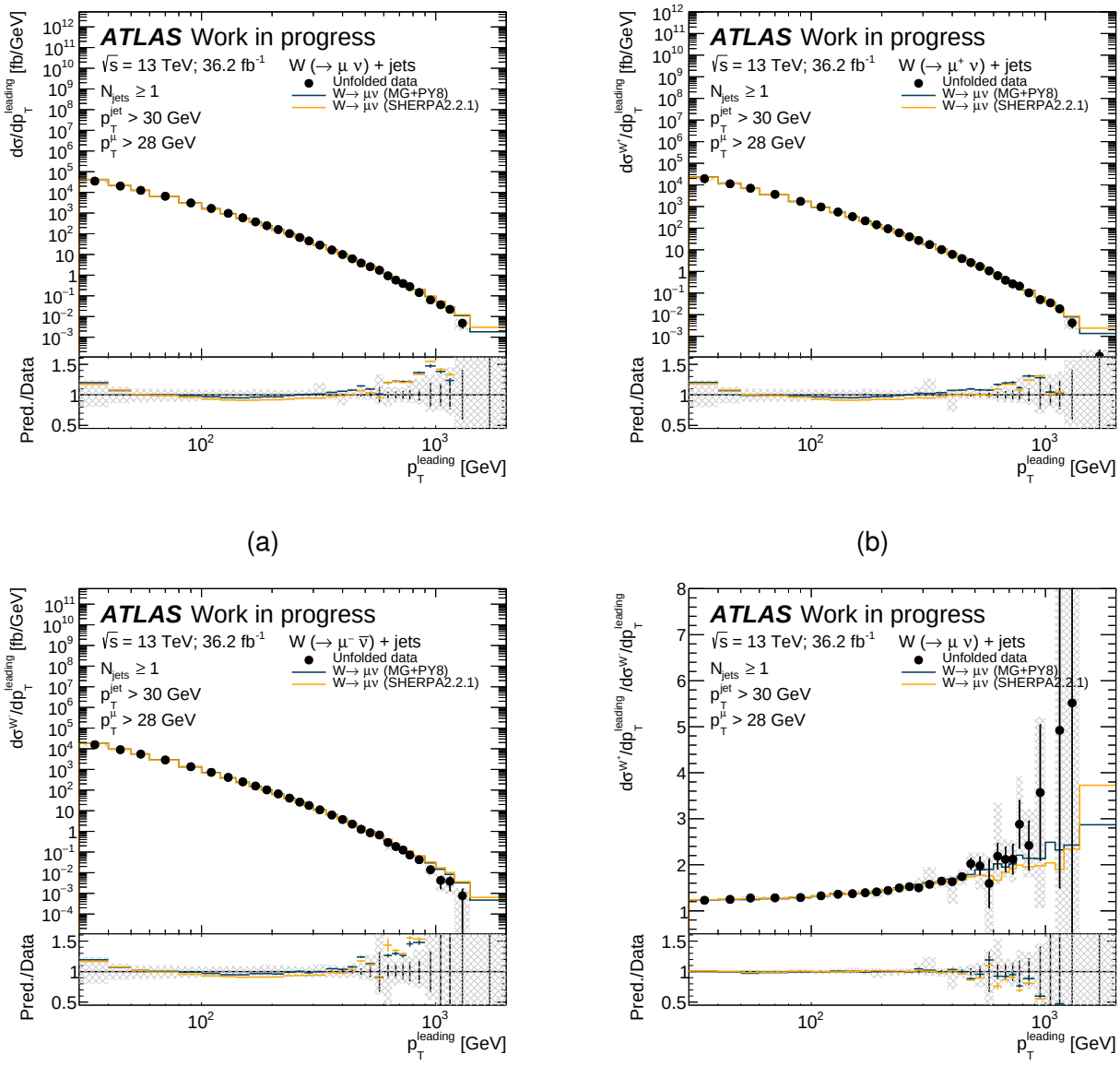


Figure 9.6: The unfolded cross-sections as a function of $p_{\mathrm{T}}^{\mathrm{leading}}$ for (a) charge-independent W+jets production in the $W \to \mu\nu$ channel, (b) $W^+ \to \mu\nu$+jets (c) $W^- \to \mu\nu$+jets (d) the W^+/W^- ratio of differential cross sections. All results are compared to two predictions, generated using Madgraph+Pythia8 with FxFx merging (blue) and one using Sherpa 2.2.11 (yellow). Statistical uncertainties are represented by black vertical bands, while full statistical + systematic uncertainties, in addition to the luminosity uncertainty, are represented by the grey bands.

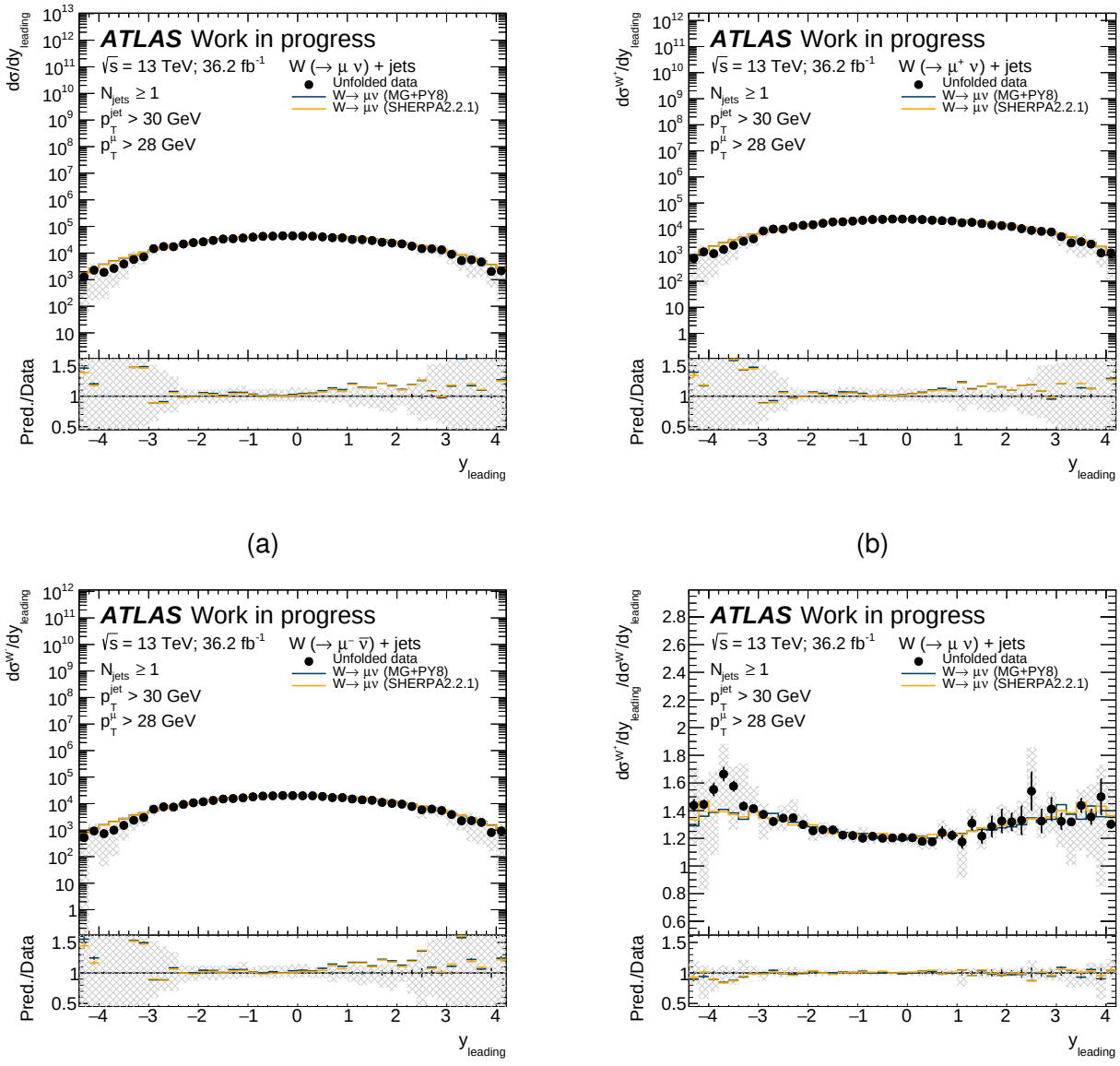


Figure 9.7: The unfolded cross-sections as a function of y^{leading} for (a) charge-independent W+jets production in the $W \to \mu\nu$ channel, (b) $W^+ \to \mu\nu$+jets (c) $W^- \to \mu\nu$+jets (d) the W^+/W^- ratio of differential cross sections. All results are compared to two predictions, generated using Madgraph+Pythia8 with FxFx merging (blue) and one using Sherpa 2.2.11 (yellow). Statistical uncertainties are represented by black vertical bands, while full statistical + systematic uncertainties, in addition to the luminosity uncertainty, are represented by the grey bands.

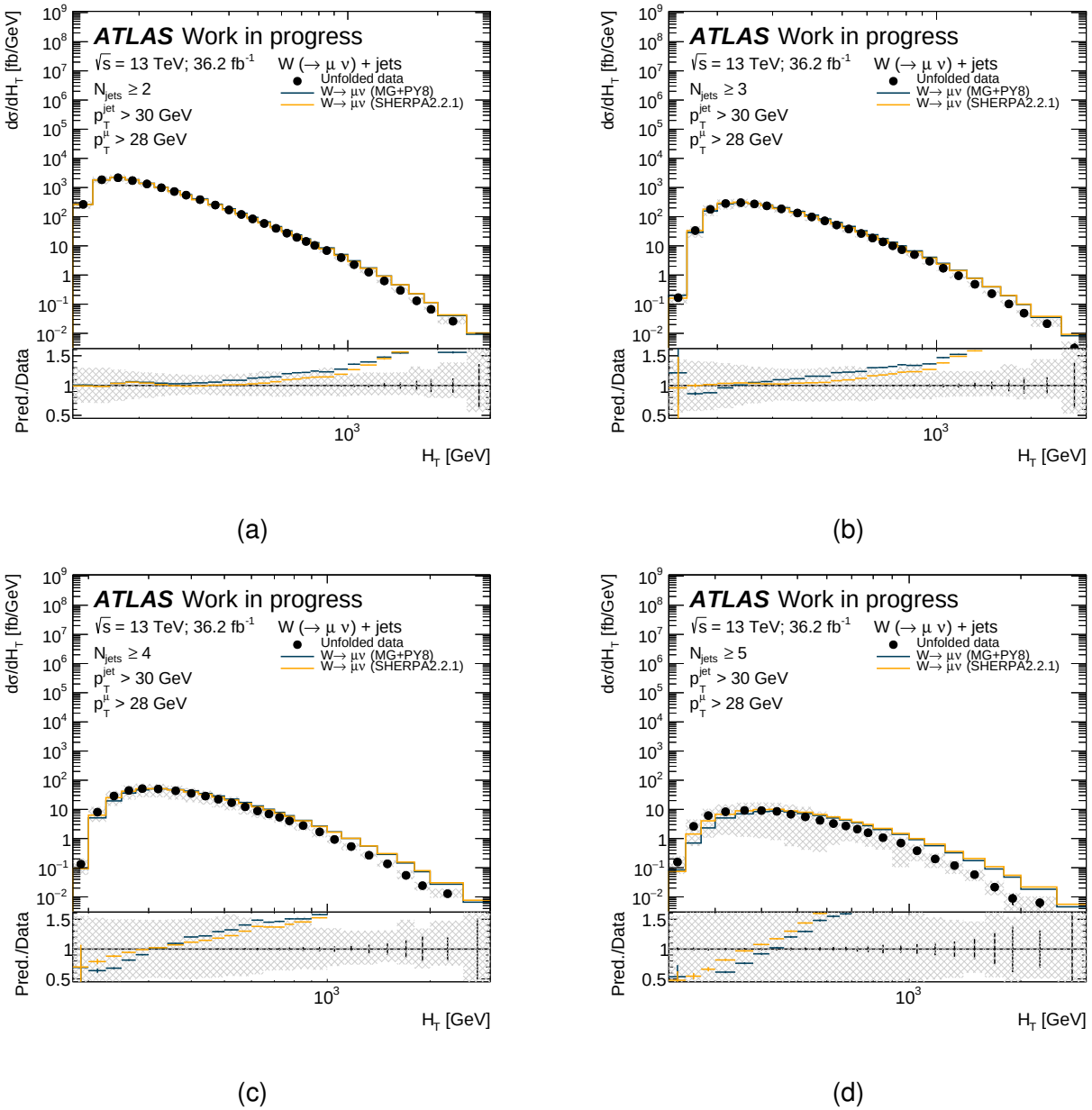


Figure 9.8: The charge-independent unfolded cross-sections as a function of H_{T} for inclusive jet multiplicities greater than 1; (a) $N_{\mathrm{jet}} \geq$2, (b) $N_{\mathrm{jet}} \geq$3, (c) $N_{\mathrm{jet}} \geq$4 (d) $N_{\mathrm{jet}} \geq$5. All results are compared to two predictions, generated using Madgraph+Pythia8 with FxFx merging (blue) and one using Sherpa 2.2.11 (yellow). Statistical uncertainties are represented by black vertical bands, while full statistical + systematic uncertainties, in addition to the luminosity uncertainty, are represented by the grey bands.

Chapter 10

Parton Distribution Functions with ATLAS data

"How do we emerge victorious from the quagmire?"

— Lin Manuel Miranda, 'Hamilton'

Precise knowledge of the parton content of colliding particles is crucial for both the Standard Model and Beyond Standard Model physics programmes of hadron colliders such as the LHC, meaning the PDFs must be thoroughly understood. In order to obtain the cross-section predictions necessary for many of the measurements and searches at such experiments, parton-level predictions must be convolved with the PDFs according to the factorisation theorem in Eq. 1.9. Precision determination of the PDFs is a key area of research, as PDF uncertainties are a limiting theoretical systematic for a wide range of important measurements, including precision SM measurements such as the W

boson mass, e.g. [134] [135], and $\sin^2\theta_{\mathrm{W}}$, e.g. [136] [137], BSM searches at high-mass and, together with the strong coupling constant α_{S} , are a dominant theory uncertainty for Higgs measurements [138].

The DGLAP equations, as specified in Eq. 1.3.2, can predict the Q^2 evolution of the PDFs from a particular starting value Q_0^2 . However, as an exact function of Bjorken-x, the PDFs are not known a priori at this starting value and thus must be extracted from fits to experimental data, usually using a χ^2 fit [139]. In order to achieve good PDF precision, broad Bjorken-x and Q^2 coverage from the input data is desirable, and can be achieved by fitting either a single measurement with broad sensitivity in x, Q^2 , in the case of ep charged-current (CC) and neutral current (NC) inclusive deep inelastic scattering (DIS), or multiple datasets encompassing a range of different processes from different experiments.

Currently, the best constraints and broadest x, Q^2 coverage are provided by the combination into a single coherent dataset of all inclusive ep deep inelastic DIS neutral NC and CC data collected by the H1 and ZEUS experiments during HERA I and HERA II [140]. This dataset has x, Q^2 coverage of 0.045 $\leq Q^2 \leq$ 50000 GeV^2 and 6$\times$10^{-7} $\leq x \leq$ 0.65 for the NC and 200 $\leq Q^2 \leq$ 50000 GeV^2 and 1.3$\times$10^{-2} $\leq x \leq$ 0.40 for the CC, and alone has sufficient sensitivity to form the basis of the HERAPDF2.0 PDF set [140]. However, these data do have certain limitations. For example, inclusive NC and CC DIS cannot discriminate different flavours of up-type quarks or down-type quarks. Global PDF fitters boost the power of the HERA dataset by simultaneously fitting it with measurements from other experiments, for example by using data from fixed-target DIS experiments [141],

[142], [143] [144], fixed target Drell-Yan experiments [145] [146] or Tevatron experiments [147] [148] [149] to further constrain the PDFs at mid to high-x. In recent years, data from LHC experiments such as ATLAS, CMS and LHCb have also proved to be powerful inputs to PDF fits; for complete summary of the LHC data used, please refer to papers by the global PDF fitting groups, for example, NNPDF [117], MSHT [150] or CTEQ [151].

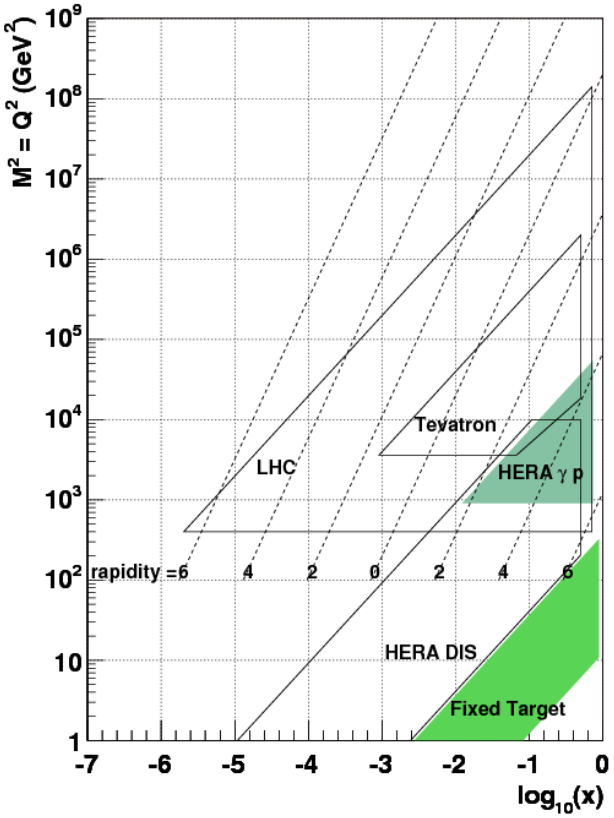
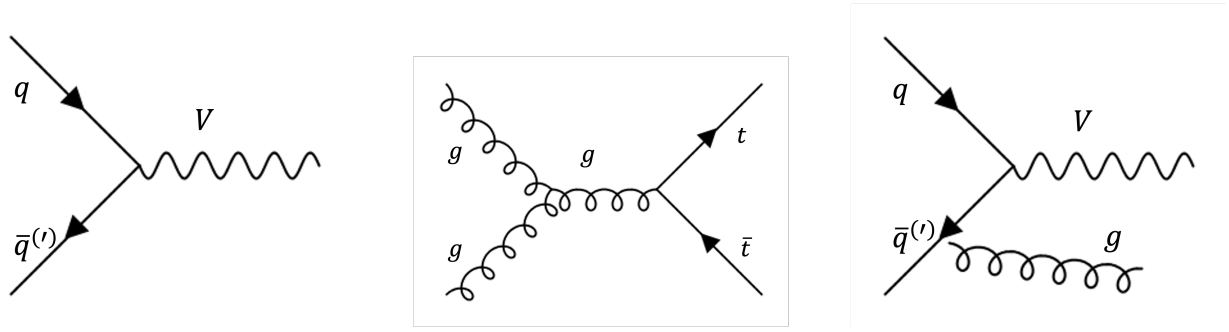


Figure 10.1: The approximate regions of sensitivity to PDFs, in the x, Q^2 kinematic plane, of the LHC, HERA, Tevatron and fixed target experiments. Figure from [152].

10.1 An Overview of PDF Fitting in ATLAS

Proton-proton measurements at the LHC, such as those performed by the ATLAS experiment, provide valuable constraints for PDF fitting, including sensitivity to the gluon and quark distributions at mid-to-high x, as well as information on quark flavour separation. ATLAS has a rich history of leveraging its PDF-sensitive measurements from the LHC Run 1 and Run 2, providing complimentary information to the combined HERA DIS dataset to make important contributions to the understanding of PDFs, following the HERAPDF philosophy [140].



(a) (b) (c)

Figure 10.2: Examples of Feynman diagrams contributing to PDF-sensitive processes that have been included in ATLAS PDF fits: (a) Inclusive vector boson production, (b) $t\bar{t}$ production, and (c) Vector boson production in association with jets.

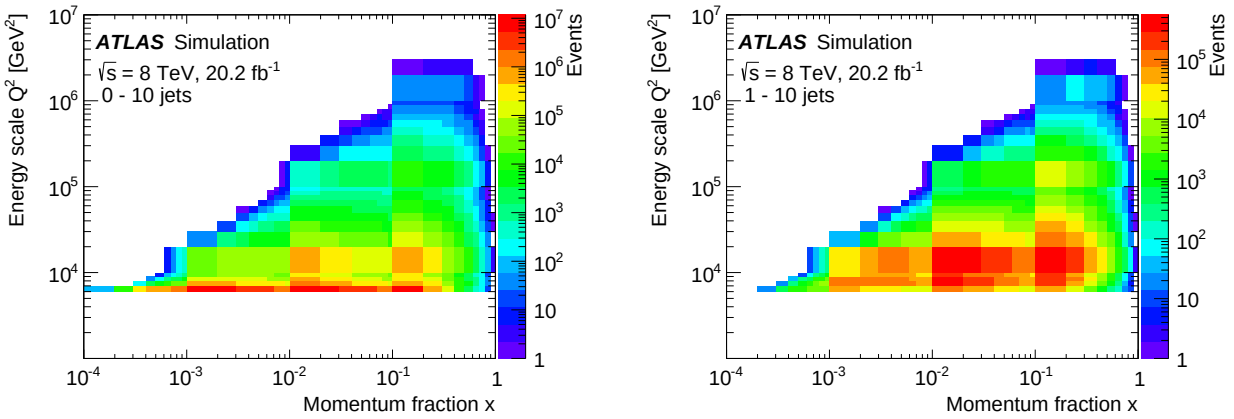
In 2016, the ATLASepWZ16 fit [153] was performed, fitting the combined HERA I+II inclusive DIS data together with the measurement of inclusive W, Z boson production (Figure 10.2a) cross-sections from 4.6 fb^{-1} of pp collision data at 7 TeV. As the inclusive W boson production cross-section is sensitive to the different couplings of the $\overline{d}$ and $\overline{s}$ quarks to the u quark, the inclusion of the ATLAS inclusive W, Z data allowed the strange quark sea to be directly fitted; for ep DIS, the $\overline{s}$ sea must instead be assumed to be a

fixed fraction of the light sea. This measurement found the ratio of the strange to light quark sea distributions, $R_s = \frac{\bar{u}+\bar{d}}{s+\bar{s}}$, to be unsuppressed ($\sim$ 1) at low-x ($x \lesssim 0.05$) This implies the strange and light quark distribution in the sea are of similar strength in this kinematic range, in tension with the measurements of neutrino induced DIS [154] [142] [155] [141] used in global PDF fits, which are more sensitive to higher values of x. A related tension was observed with results from the fixed-target Drell-Yan experiment E866 in the $\overline{d} - \overline{u}$ distribution at high-x [146]. Measurements from E866 indicated that this distribution is positive at high-x, whereas results from the ATLASepWZ16 fit found it to be slightly negative and compatible with zero within uncertainties.

In 2018, data from ATLAS measurements of $t\bar{t}$ production (Figure 10.2b) at 8 TeV in the lepton + jet channel [156] and the dilepton channel [157] were added to the datasets used in the ATLASepWZ16 fit. For the first time in an ATLAS PDF fit, it was possible to include multiple PDF-sensitive spectra from the same measurement in the fit; this was made possible by the provision of both statistical and systematic correlations between the different $t\bar{t}$ spectra by ATLAS. The resulting ATLASepWZtop18 fit [158] fit produced a gluon distribution that was made marginally harder by the inclusion of the $t\bar{t}$ data, and the high-x gluon uncertainties were constrained.

Measurements of the production of vector bosons in association with jets (V+jets), (Figure 10.2c), such as the 13 TeV W+jets measurement described earlier in this document, are powerful inputs to PDF fits. The addition of the jets extends sensitivity in x, Q^2 space with respect to inclusive W, Z measurements, as shown in Figure 10.3. Additionally, at tree-level V+jets production is initiated either by quark-antiquark interaction with

gluon radiation or quark-gluon interaction, meaning these reactions are sensitive to the gluon at their lowest order in QCD. In 2020, the ATLASepWZVjet20 analysis [159] fit the ATLAS 7 TeV inclusive W, Z [153], ATLAS 8 TeV inclusive W [160]and Z [161], ATLAS 8 TeV Z+jets [162] and ATLAS 8 TeV W+jets [21] measurements together with the HERA I+II combination [140].



(a) (b)

Figure 10.3: Coverage in x, Q^2 space for inclusive W boson production (10.3a) and W boson production in association with at least one jet (10.3b). The requirement that at least one jet be present in the final state increases the sensitive range in both x and Q^2 . Figure from [21].

The addition of the V+jets data was found to increase the high-x $\overline{d}$ distribution and decrease the high-x $\overline{s}$ distribution, with the sum of the two constrained by HERA data. The increased high-x $\overline{d}$ also impacted the determination of the $\overline{d} - \overline{u}$ distribution at high-x. The additional control of the down-type sea in this kinematic region supplied by the V+jets data returned a positive high-x distribution in reasonable agreement with the E866 results up to $x \simeq 0.1$, improving the tension between ATLASepWZ16 and E866 data. The ATLASepWZVjet20 fit retained the unsuppressed R_s at low-x observed in previous ATLAS fits. However, the observed tension with results from the global PDF fits was found

to be somewhat reduced, particularly with fits which include the ATLAS 7 TeV inclusive W, Z measurement [153], such as CT18A [151], NNPDF3.1 [117] and NNPDF3.1_strange [163].

10.2 Recipe for an Experimental PDF Fit

The following section describes the statistical framework employed in ATLAS, following the HERAPDF philosophy [140] for performing for a PDF fit to experimental data; the information provided has been summarised from [139]. In general, PDF fits are performed using a χ^2 fit of parton-level cross-section predictions convolved with the PDFs, which have been evolved from some starting scale Q_0^2 up to the scale of the experimental data using the DGLAP equations from Eq. 1.3.2. A variety of functional forms in x may be chosen for the PDFs. However, in general, ATLAS PDF fits employ the HERAPDF scheme [140], with the quark distributions taking the form:

$$xq_i(x) = A_i x^{B_i}(1-x)^{C_i} P_i(x), \tag{10.1}$$

where terms of the polynomial $P_i(x) = (1 + D_i x + E_i x^2 + F_i x^3)$ are added to the central parameterisation only if their inclusion results in a decrease of $\chi^2 > 2$, as described in [164], as the addition of extra parameters beyond this 'χ^2 saturation' point serves only to fit noise in the data. Additionally, the gluon distribution has the more flexible form:

$$xg(x) = A_g x^{B_g}(1-x)^{C_g} P_g(x) - A'_g x^{B'_g}(1-x)^{C'_g}.$$

Physics-motivated constraints can then imposed on these parameters; examples include the quark number sum rules and the momentum sum rule.

A χ^2 formulation must be chosen such that correlated systematic uncertainties are correctly accounted for. One method of accomplishing this is through a definition where the correlated systematic errors and associated nuisance parameters (independent variables of zero mean and unit variance for each source of correlated uncertainty) are included in the theoretical prediction [139]:

$$F_i(p,b) = F_i^{\mathrm{NNLOQCD}}(p) + \sum_j \gamma_{ij} b_j,$$

where the theoretical predictions F for point i are now a function of the theoretical parameters p and the nuisance parameters b associated with the jth correlated source of systematic uncertainty on point i, γ_{ij}.

The χ^2 definition for the fit can then be formulated as:

$$\begin{aligned}\chi^2 = &\sum_{i,k}(D_i - F_i(1-\sum_j \gamma_{ij}b_j))\; C^{-1}_{\mathrm{stat,uncor},ik}(D_i,D_k)\; (D_k - F_k(1-\sum_j \gamma_{kj}b_j)) \\ &+ \sum_i \log\frac{\delta^2_{i,\mathrm{uncor}}F_i^2 + \delta^2_{i,\mathrm{stat}}D_iF_i}{\delta^2_{i,\mathrm{uncor}}D_i^2 + \delta^2_{i,\mathrm{stat}}D_i^2} \\ &+ \sum_j b_j^2,\end{aligned} \tag{10.2}$$

where D_i and F_i are the data and theoretical predictions respectively, $\delta^2_{i,\mathrm{uncor}}$ and $\delta^2_{i,\mathrm{stat}}$ are the uncorrelated systematic and statistical uncertainties on D_i and have the covariance matrix $C^{-1}_{\mathrm{stat,uncor},ik}$. The correlated uncertainties are given by γ_{ij} and accounted for

by the nuisance parameter b_j. The first term in Eq. 10.2 is the 'partial χ^2' of the data set indicated by the summation over i (k). The second term is known as the 'log penalty' and is a small bias correction term. The final term is the 'correlated χ^2' and comes from the j nuisance parameters used to describe the correlations between systematic uncertainties.

One method of performing the fit is the Hessian method [165] [166], where the nuisance parameters are allowed to vary alongside the theoretical parameters in the central fit. In effect, in the Hessian method the theory is not fit to the data, but rather the data points are collectively shifted in accordance with their correlated systematic uncertainties, such that the fit returns the optimal set of correlated shifts that result in the most consistent agreement with the theory across all datasets. Observed shifts much greater than 1σ could indicate inconsistencies between the datasets being fitted, or that the associated uncertainties have not been well-understood.

The errors on the PDF parameters are calculated from:

$$\langle \sigma_F^2 \rangle = T \sum_i \sum_k \frac{\partial F}{\partial p_i} V_{ik} \frac{\partial F}{\partial p_i},$$

where V_{ik} is the covariance matrix expressing how the fit χ^2 varies with respect to both theoretical parameters and the nuisance parameters and T is the fit tolerance. However, as PDF parameters are usually correlated with each other, resulting in Hessian matrices that are generally not diagonal, it can be challenging to separate which individual PDF parameters have the greatest impact. The eigenvector basis can simplify the expression of the results of a PDF analysis. By diagonalising the covariance matrix of the parame-

ters, the resulting eigenvalues correspond to the squares of the errors of the eigenvector combinations of PDF parameters, and should therefore always be positive. PDF sets generally supply their errors in the form of two fit replicas for each eigenvalue, shifted up and down the eigenvector direction by the fit tolerance T. This makes the calculation of PDF uncertainties on physical observables much simpler than if the full covariance matrix must be considered.

In addition to the correlated experimental systematic uncertainties discussed above, ATLAS PDF fits also consider uncertainties due to parameterisation and model choices. Parameterisation uncertainties are evaluated by adding extra D, E, F parameters (controlling the high-x regime) to the expression in Eq. 10.1, or by releasing constraints on the low-x sea. Similarly, model uncertainties are estimated by varying theoretical assumptions, for example Q_0^2 or heavy quark masses The PDF fit is reevaluated for each of these variations, and the difference with respect to the central fit is taken as an uncertainty, each of which are added in quadrature with the experimental uncertainties to return the complete set of PDF uncertainties.

10.3 The ATLASpdf21 Fit

The aim of the ATLASpdf21 fit [1] was to show the PDF constraining power of ATLAS measurements by simultaneously fitting as many PDF sensitive ATLAS datasets as possible, alongside leveraging ATLAS experimental expertise to demonstrate the impact of correctly implementing the correlations of the systematic uncertainties both within and

between the datasets. The latter is non-trivial, as individual measurements are made by different analysis teams within ATLAS, which can result in different naming schemes being employed for the same uncertainties. In addition, there is the need to separate the degree of correlation of systematic uncertainty between different measurements analysing the same data and including some of the same final state objects e.g. 8 TeV $t\bar{t}$ + jets, V+jets , inclusive jets etc.

The HERA I+II combined inclusive NC and CC Deep-Inelastic DIS data [140], providing wide coverage in Bjorken-x and Q^2 were fit simultaneously together with a range of ATLAS cross-section measurements: inclusive W, Z production at 7 TeV [153], inclusive W [160], Z [161] production at 8 TeV, W+jet production at 8 TeV [21], Z+jet production at 8 TeV [162], $t\bar{t}$ production at 8 TeV [156], $t\bar{t}$ production at 13 TeV [167], direct photon production ratios at 8/13 TeV [168] and inclusive jet production at 8 TeV [169]. Two jet radii were provided in the inclusive jet measurement, and the radius R=0.6 was chosen for the fit, as it is considered more theoretically reliable.

A next-to-next-to-leading-order (NNLO) QCD analysis was performed on these data using the open-source xFitter framework [170] with the minimisation performed by the MINUIT programme [171], and was cross-checked using code independently developed previously at Oxford. For the predictions corresponding to the ATLAS data, APPLGRIDs [172] or fastNLO [173] lookup tables were used, either directly at NNLO or at NLO corrected to NNLO using multiplicative bin-by-bin k-factors, derived from the ratio of the NNLO to the NLO QCD calculation. NLO EW corrections were applied to the LHC data[1] in the form of

[1]For the DIS data, electroweak effects are already unfolded to LO, and NLO EW corrections are not well-defined.

k-factors derived from the ratio of NLO to LO calculated cross-sections. Additionally, for LHC processes which included jets in the final state, non-perturbative correction k-factors were also applied to account for hadronisation effects.

Previous ATLAS fits described in Section 10.1 used either 15 or 16 free parameters in their fits. The ATLASpdf21 analysis had 21 free parameters, allowing greater flexibility of the fit; this is possible due to the diversity of datasets included in the fit. Using a starting scale Q_0^2 = 1.9 GeV^2 for DGLAP evolution (chosen to be below the scale of the charm mass), the parton distributions for the valence quarks u_{v} and d_{v} , the light sea anti-quarks $\overline{u}$,$\overline{d}$ and $\overline{s}$ and the gluon g were fit assuming a fixed α_{S} = 0.118 at the scale of the Z mass in line with the PDG global average value [3]. A Q^2_{min} = 10 GeV^2 cut was placed on HERA data to avoid regions requiring additional treatment e.g. small-x resummation [174] [175], motivated by previously observed poorer fit in the excluded region compared to the rest of the HERA data [140]. The heavy quark masses were taken from a recent re-analysis of the optimal values of the heavy-quark masses from HERA heavy quark data [176], and are taken to be m_c = 1.41 GeV and m_b = 4.2 GeV. All theoretical choices follow those employed in the HERAPDF2.0 fit, and are varied while evaluating the fit's model uncertainties. A list of the model and parameter variations studied, and the resulting varied fit χ^2 , can be found in Table 10.4.

Careful considerations were made of the systematic correlations between datasets. These were made public with the PDF set, for other fitters to use as a framework for the correct treatment of ATLAS correlations. Figure 10.5a demonstrates the importance of this - by accounting for systematic correlations between datasets, the central values of

Central χ^2/NDF	2010/1620
Model variations	
Q^2_{min} = 12.5 GeV^2	1947/1571
Q^2_{min} = 7.5 GeV^2	2076/1660
m_c = 1.45 GeV (sym)	2025/1620
Q^2_0 = 1.6 GeV^2 (sym)	2018/1620
m_b = 4.3 GeV	2016/1620
m_b = 4.1 GeV	2014/1620
m_t = 175.0 GeV	2063/1620
m_t = 172.5 GeV	2018/1620
R = 0.4	2080/1620
Parameter variations	
F_{u_v}, $D_{\bar{d}}$	2007/1620

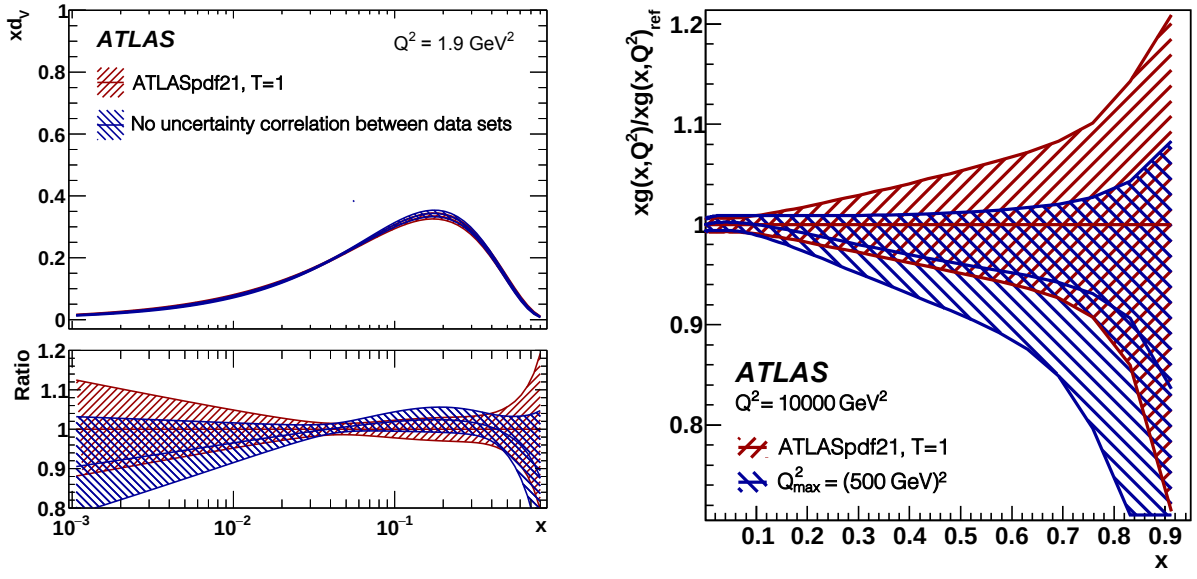
Figure 10.4: The varied theoretical and parameterisation choices used to evaluate the ATLASpdf21 model and parameterisation uncertainties. 'sym' indicates that the variations upwards and downwards have been symmetrised, R=0.4 indicates a varied jet radius for the 8 TeV inclusive jet data, and $F_{u_v}, D_{\bar{d}}$ indicates a variation where the listed parameters have been added. Figure from [1].

the fit shift by $>$ 2% at mid-x at LHC scales. Consideration was also made of the scale uncertainties of the predictions where necessary. For the 7 and 8 TeV inclusive W, Z datasets, the measurements' precision is comparable to the size of NNLO scale uncertainties. To account for this, the scale uncertainties were added as additional systematic. This was found to result in small differences between PDF shapes, but accounting for such differences is important if $\mathcal{O}$(1%) uncertainties are sought on the PDF measurements.

Given the number of datasets and different types of reactions included in the fit, the possibility arises for inconsistencies between the datasets. To account for this, an analysis of the optimum tolerance for the ATLASpdf21 fit was performed using the MSHT dynamic tolerance procedure [177]. The most constraining datasets for the tolerance were found to be the 7, 8 TeV inclusive W, Z measurements, with tolerance values for all eigenvectors falling in the range T=2.4$-$4.2. On this basis, a central value of T=3 was chosen for the

central fit, to ensure no dataset fell too far outside its 68% CL limit. PDF sets with T=1 and T=3 were both made publicly available on the LHAPDF public repository.

A common concern in PDF fitting is in regards to the possibility of 'fitting away' the effects of high-scale BSM physics present in the data, thus distorting the SM background prediction and masking these effects when the PDFs are subsequently used in searches. To ensure that this is not the case for this fit, a Q^2_{max} = 25000 GeV^2 cut was imposed on the input data and the fit was repeated. Figure 10.5b shows that high scale data do not strongly affect the PDF shapes or uncertainties in this measurement within uncertainties. Thus, any BSM effects that may be present in the data are not expected to skew the final PDFs, and the cut was not considered further.



(a) (b)

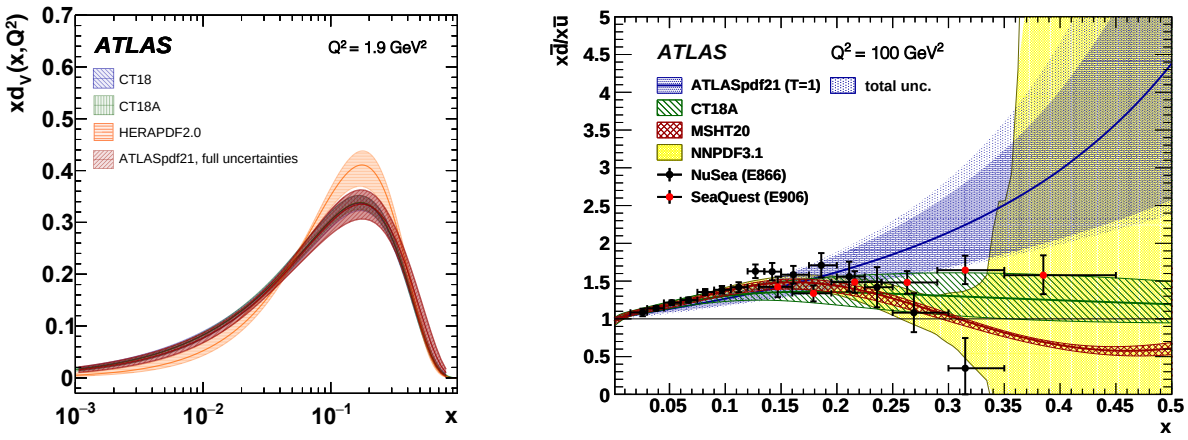
Figure 10.5: (a) The d_{v} distribution for the ATLASpdf21 fit with the correct correlations between systematic uncertainties applied (red) and no uncertainty correlation applied between datasets (blue). (b) The ratio to the gluon distribution of the central ATLASpdf21 T=1 fit (red) of the fit repeated with a $Q^2 < (500\,\text{GeV})^2$ cut applied (blue). The linear x scale emphasises that, even at very large x values, high scale data do not impact the fit within uncertainties. Figures from [1]

The final fit found R_s to be unsuppressed at low-x, in accordance with previous ATLAS results and a 2021 QCD analysis of W+charm production by CMS [178]. However, this fit showed less tension with the global fitters than exhibited by the previous ATLAS fits, particularly with CT18A [151], MSHT20 [150] and NNPDF3.1_strange [163], all of which included the ATLAS 7 TeV W, Z inclusive measurement in their fits. The PDFs produced by this fit were also compared to HERAPDF2.0 and various global fits. As exemplified in Figure 10.6a, the d_{v} and $\overline{d}$ distributions with enhanced tolerance more closely resembled the global fitters than HERAPDF2.0, indicating the ATLAS data used in ATLASpdf21 are replicating features of the Tevatron and fixed-target DIS. data used in the global fits.

Finally, the ATLASpdf21 result was considered in the context of recent fixed-target Drell-Yan data. At high-x, the E906 (SeaQuest) [179] experiment reports a $\bar{d}/\bar{u}$ ratio greater than 1, whereas its predecessor E866 (NuSea) [146] observed a ratio less than 1, in tension with the updated result, When the $T = 1$ ATLAspdf PDF set was used to predict this ratio at high-x, Figure 10.6b indicates a result much more in agreement with the new observation than the old. Additionally, predictions using CT18A [151], which includes the ATLAS 7 TeV W, Z data, were found to replicate the new data well.

10.3.1 Decorrelation Studies of Jet Systematic Uncertainties

For the ATLASpdf21 paper, the author studied the impact of the treatment of systematic correlations in the ATLAS 8 TeV inclusive jet data on the fit [169]. The main impact of these data, as shown in Figure 10.7, is the constraint on the gluon, specifically a mild



(a) (b)

Figure 10.6: (a) The T=3 ATLASpdf21 d_{v} distribution compared to HERAPDF2.0, CT18 and CT18A. The ATLAS fit more closely remembers the global fits than HERAPDF. (b) A comparison of the T=1 ATLASpdf21 prediction for the $\bar{d}/\bar{u}$ distribution to the E866 and E906 results, alongside the prediction from various global PDF fits. Figures from [1].

hardening of the gluon PDF at high-x and a decrease in the high-x gluon uncertainties.

Historically, ATLAS jet data at 7 TeV have proven challenging to fit to QCD predictions at both NLO and NNLO in QCD, resulting in very poor χ^2 values; this was first observed in ATLAS but was subsequently noticed by global fitters [180] [151]. This motivated ATLAS to study options for the decorrelation of 'two-point' jet-related systematic uncertainties in the 8 TeV inclusive jet publication [169].

Two-point systematics are estimated by taking the difference between two MC generators. Their correlations are not known a priori and these systematics are therefore viable candidates for decorrelation treatment. Instead of assuming one internal source of disagreement between the MC generators, which would cause the systematic to be fully correlated across all p_{T} and rapidity bins comprising the jet spectrum, the decorrelation scenarios make the equally reasonable assumption of $\geq$ 2 independent differences that

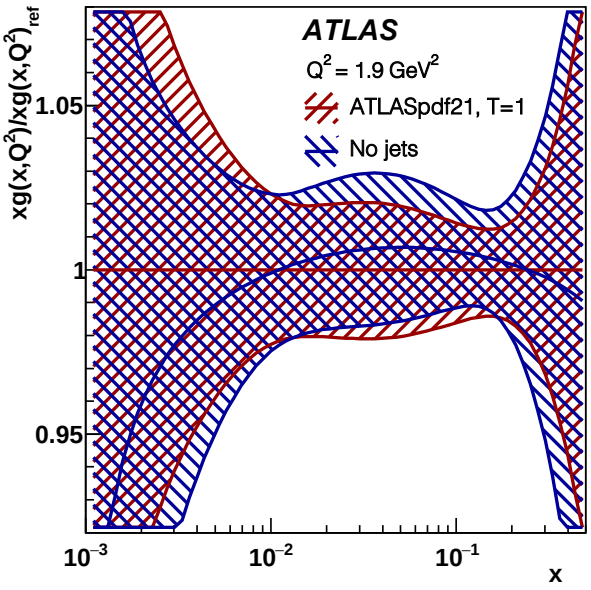


Figure 10.7: The ratio to the central T=1 gluon distribution of the ATLASpdf21 fit (red) of the fit repeated with the 8 TeV inclusive jet data removed (blue), demonstrating the effect of these data in the fit. Figure from [1]

are functions of p_{T} and rapidity. In this case the original uncertainty can be decomposed into uncorrelated components.

Three two-point systematic uncertainties were identified as candidates for decorrelation treatment in the 8 TeV inclusive jet paper:

- 'Jet Flavour Response', the systematic uncertainty introduced by the difference in detector response between jets initiated by quarks and jets initiated by gluons;
- 'Multijet Balance Fragmentation', the systematic uncertainty due to jet fragmentation in the multijet p_{T} balance used for Jet Energy Scale calibration;
- 'Pile-up Rho Topology' is the systematic uncertainty on the average pile-up density ρ used at the pile-up subtraction stage of jet reconstruction.

Two different decorrelation schemes, found in the 8 TeV inclusive jet paper to reduce the data/MC χ^2 , were studied in the context of the ATLASpdf1 fit;

- **Decor 1:** Jet Flavour Response Splitting Option 14, Multijet Balance Fragmentation Splitting Option 17 and Pile-up Rho Topology Splitting Option 16;
- **Decor 2:** Jet Flavour Response Splitting Option 7, Multijet Balance Fragmentation Splitting Option 17 and Pile-up Rho Topology Splitting Option 16,

where the relevant splitting options, extracted from the 8 TeV inclusive jet paper, are given in Table 10.1. Each of the listed sets of one, or two, sub-components are accompanied by one additional complementary sub-component, defined such that the sum in quadrature of all sub-components in a given splitting option returns the magnitude of the original uncertainty. The function $L(x, min, max)$ is defined by:

$$L(x, min, max) = (x - min)/(max - min),$$

when $min < x < max$, and $L =$0 when $x < min$ or $x > max$.

Both of the above decorrelation scenarios were found to result in the significant reduction of the data/MC χ^2 in the ATLAS 8 TeV inclusive jet analysis [169], and differential cross-sections produced using these two splitting schemes were made available by ATLAS.

The author investigated the impact of using the Decor 1 and Decor 2 decorrelation schemes in the ATLASpdf21 analysis, having implemented the published results in an

Splitting option	Definition, completed by complementary
7	L($\|y\|$, 0, 3)· uncertainty
14	$L(\ln(p_{\mathrm{T}}[\mathrm{TeV}]), \ln(0.1), \ln(2.5)) \cdot \sqrt{1 - L(\|y\|, 0, 1)^2}$·uncertainty $L(\ln(p_{\mathrm{T}}[\mathrm{TeV}]), \ln(0.1), \ln(2.5)) \cdot L(\|y\|, 1, 3)$· uncertainty
16	$\sqrt{1 - L(\ln(p_{\mathrm{T}}[\mathrm{TeV}]), \ln(0.1), \ln(2.5))^2} \cdot \sqrt{1 - L(\|y\|, 0, 1.5)^2}$· uncertainty $\sqrt{1 - L(\ln(p_{\mathrm{T}}[\mathrm{TeV}]), \ln(0.1), \ln(2.5))^2} \cdot L(\|y\|, 1.5, 3)$·uncertainty
17	$\sqrt{1 - L(\ln(p_{\mathrm{T}}[\mathrm{TeV}]), \ln(0.1), \ln(2.5))^2} \cdot \sqrt{1 - L(\|y\|, 0, 1)^2}$·uncertainty $\sqrt{1 - L(\ln(p_{\mathrm{T}}[\mathrm{TeV}]), \ln(0.1), \ln(2.5))^2} \cdot L(\|y\|, 1, 3)$·uncertainty

Table 10.1: A table of the relevant two-point systematic splitting options used in the studied decorrelation schemes. Each set of sub-components is accompanied by one additional complementary sub-component, defined such that all sub-components in a given splitting option sum in quadrature to return the magnitude of the original uncertainty. The function $L(x, min, max)$ is defined by $L(x, min, max) = (x - min)/(max - min)$ when $min < x < max$, and L =0 when $x < min$ or $x > max$. Splittings originally provided in [169].

xFitter compatible format, alongside the nominal scheme where all three two-point jet systematic uncertainties are fully correlated in p_{T} and y. One additional decorrelation scheme was also studied; the **JFR** scheme, where the jet flavour response uncertainty is fully decorrelated across y bins. This is not one of the scenarios studied in the 8 TeV inclusive jet paper, but was included as it is straightforward technically to implement and was found early in the analysis to improve the partial χ^2 associated with the jet data. This study was performed using an early, 16 parameter version of the ATLASpdf21 fit which followed the ATLASepWZ16 parameterisation; however, subsequent follow-up studies by the analysis team showed that the more flexible 21 parameter fit also results in the same conclusions.

As shown in Figure 10.8, the final PDFs are insensitive to the correlation scenario chosen, providing freedom to choose the correlation scenario used in the fit. However, as shown in Table 10.2, the scenario chosen does have an impact on the fit χ^2 . In relation to

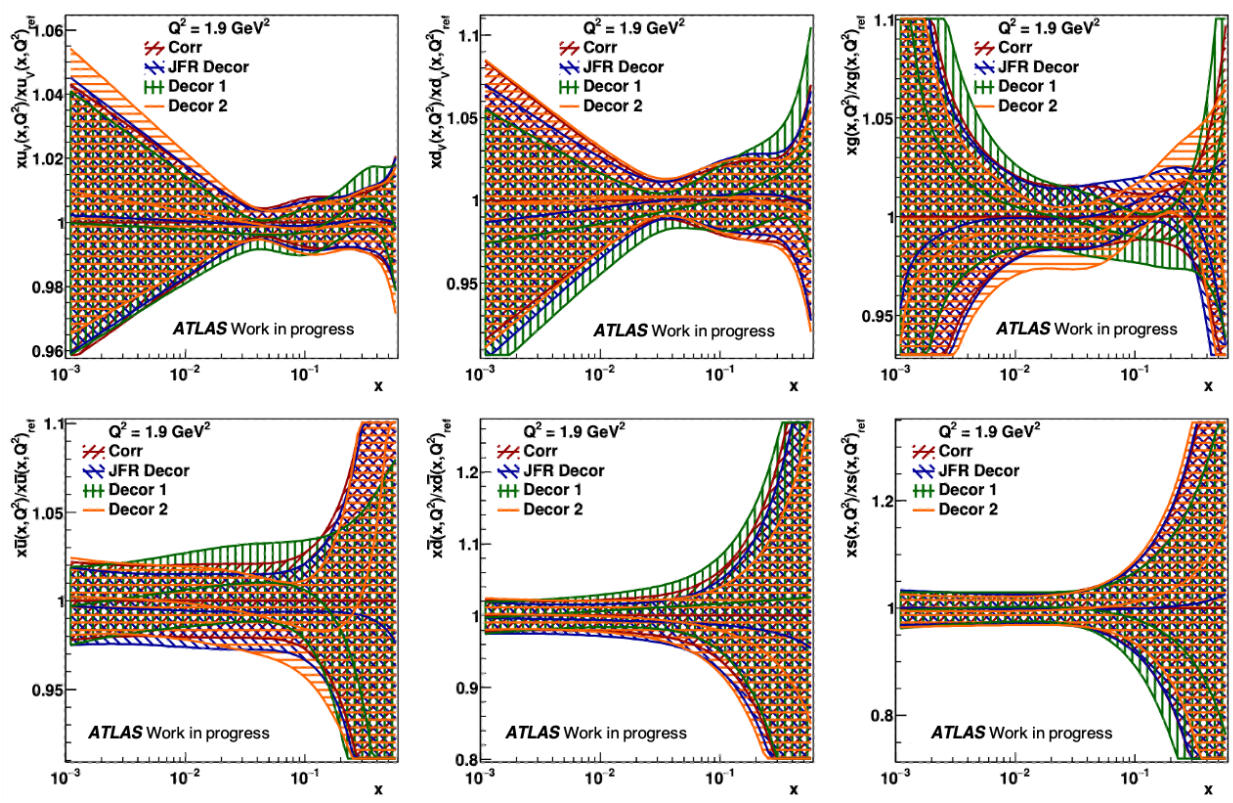


Figure 10.8: Ratios of the fitted parton distributions for each decorrelation treatment to the fully correlated default scenario, using a 16 parameter version of the ATLASpdf21 fit. Top: xu_x, xd_v, xg, Bottom: $x\bar{u}$, $x\bar{d}$, xs. Four correlation treatments for jet two-point systematics are considered: Fully correlated (red), Jet Flavour Response fully decorrelated across rapidity bins (blue), two-point systematic decorrelation scenario 1 (green) and two-point systematic decorrelation scenario 1 (orange). The final PDFs exhibit little sensitivity to the correlation scenario chosen.

Eq. 10.2, the first row of Table 10.2 corresponds to the sum of all three terms i.e. the left-hand side, the second row corresponds to the first term and the third row corresponds to the third term. In the context of the historical difficulties, associated with including ATLAS jet data in PDF fits, the observed χ^2 values associated with the decorrelation of jet-related systematic uncertainties is a significant improvement, proving that ATLAS jet data can be well-fit in a PDF analysis. This opens the door to further exploiting the wealth of ATLAS jet measurements in future PDF studies.

All the decorrelated scenarios performed better than the fully correlated scenario, with the JFR decorrelation performing best. However, as noted above, this was not one of the scenarios recommended in the jet paper and is included only as a cross-check. Ultimately the alternative scenario Decor 2 was selected for the final ATLASpdf21 fit, as this was the scheme which resulted in the strongest data/MC χ^2 reduction in the 8 TeV inclusive jet paper.

	Fully correlated	JFR	Decor 1	Decor 2
ATLASpdf21 global χ^2/NDF	2222/1625 (1.37)	2138/1625 (1.32)	2179/1625 (1.34)	2190/1625 (1.35)
Jet partial χ^2/NDP	273/171 (1.60)	229/171 (1.34)	247/171 (1.44)	247/171 (1.44)
Jet χ^2 correlated term	119	79	106	110

Table 10.2: The χ^2 values for each jet systematic correlation scenario used in an earlier 16 parameter version of the ATLASpdf21 fit. Top: The global χ^2 per degree of freedom for the full fit Middle: The partial χ^2 per data point for the inclusive jet data, Bottom: The correlated term for the inclusive jet data.

Conclusion

"And so we beat on, boats against the current, borne back ceaselessly into the past."

— F. Scott Fitzgerald, 'The Great Gatsby'

The Large Hadron Collider at CERN is the most powerful machine ever built for probing the Standard Model, colliding proton-proton beams accelerated to $\sqrt{s}$=13 TeV to study the high-intensity and high-energy regimes of the theory. While many fascinating high-precision measurements of the Standard Model, and searches for physics beyond it, can be performed at general-purpose LHC experiments such as the ATLAS detector, investigations involving jets provide the opportunity to accomplish the former, by exploring the still in places mysterious theory of Quantum Chromodynamics, while simultaneously assisting the community with the latter, by providing constraints on backgrounds and benchmarks for improvements in modelling.

In order to accomplish this, a necessary evil of high-intensity physics at a hadron collider must first be tamed: the menace of pile-up. To this end, the first application of a topology-based pile-up tagging algorithm in the ATLAS central region has been pre-

sented, along with the first evidence that such a method may complement the default track-based pile-up tagging method in this region. Additionally, it was shown that such an approach can improve the reconstruction of missing transverse energy in events where the default tagger has failed, a finding with lasting implications on missing transverse energy reconstruction in ATLAS and the many searches for new physics which depend on it.

Subsequently, the differential cross-sections for the muon channel of the first ATLAS W+jets measurement at $\sqrt{s}$=13 TeV, and the first measurement of the muon channel of this process since $\sqrt{s}$=7 TeV, were presented. The unfolded results are compared to today's state-of-the-art theoretical predictions and, by noting the regions of fiducial phase space in which their agreement with data break down, provide crucial feedback to modelling groups. This measurement will inform and improve the next generation of W+jets simulations, improving the background estimation for future BSM and Higgs measurements with which this process shares a final state. Additionally, W+jets can be considered a 'stress test' for SM simulation, owing to the need to model both electroweak and QCD processes, its large cross-section extending the possible measurements of several observables up to TeV scales, and the presence of up to many jets in the final state. Therefore, this measurement will provide an important benchmark for the QCD modelling community more broadly for years to come.

Furthermore, W+jets is proven to be a valuable input to PDF fits, due to its sensitivity to down-type sea quark flavour, sensitivity to the gluon at the lowest-order in QCD and extended x, Q^2 sensitivity with respect to inclusive W boson production arising from the

final state jet. Indeed, the ATLASepWZVjets20 [159] PDF fit from 2020 has previously shown that its inclusion of the ATLAS $\sqrt{s}$=8 TeV W+jets measurement [21] allowed for the high-x $\overline{d}$ and $\overline{s}$ distributions to be fit directly for the first time in an ATLAS PDF study, and for tensions to be relaxed between previous ATLAS fits and global PDF fitters. A new QCD analysis including the $\sqrt{s}$ =13 TeV W+jets measurement is planned for the near future, where the results presented in this thesis will continue to have impact, revealing more of the inner workings of the proton.

The W+jets measurements presented in this thesis are interim results; the completed paper, which will include an electron channel and a combined channel, is expected to be published in 2024. This second paper will include the full 139 fb^{-1} integrated luminosity recorded by ATLAS during the LHC Run 2, updated unfolding techniques and the latest software updates and performance recommendations from ATLAS. An expanded list of distributions will be produced, with studies underway within the analysis team to identify additional variables sensitive to MC modelling. As part of this effort, the author has written a 'Rivet' [181] routine for this analysis, which applies the analysis fiducial selections at truth-level to MC.

Efforts such as the aforementioned improvement of our understanding of PDFs using W+jets measurements are fundamental to progressing the physics programme at hadron colliders. Yet, like much of the sometimes-murky field of non-perturbative QCD, there still remains much to be learned about the behaviour of partons inside the proton; for example, even today, several crucial measurements are by limited by PDF uncertainties. ATLAS has a rich history of leveraging its data and experimental expertise to advance our

understanding of of this topic. In this thesis, an overview of, and the author's contributions to the ATLASpdf21 fit have been presented; this is the most comprehensive PDF fit ever produced by ATLAS, and will be the gold-standard for future ATLAS PDF fits and an invaluable framework for the treatment of ATLAS data by the global PDF fitting community in the years ahead.

In conclusion, while the Standard Model has stood up to rigorous testing, there is still much to uncover within it. Moreover, a precise understanding of the Standard Model will be crucial for the discovery of new physics. Looking ahead to future developments such as the High-Luminosity LHC [182], the Electron-Ion Collider at Brookhaven National Laboratory [183] and the next-generation CERN accelerator, there is much on the horizon for experimental high-energy particle physics and the exploration of the fundamental nature of the universe well into the future.

Bibliography

[1] ATLAS Collaboration. "Determination of the parton distribution functions of the proton using diverse ATLAS data from pp collisions at $\sqrt{s}$ = 7, 8 and 13 TeV". In: *Eur. Phys. J. C* 82 (2022), p. 438. DOI: 10.1140/epjc/s10052-022-10217-z. arXiv: 2112.11266 [hep-ex].

[2] Wikimedia Commons. *File: Standard Model of Elementary Particles.svg — Wikimedia Commons, the free media repository*. [Online; accessed 19-September-2023]. 2023. URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_Anti.svg.

[3] Zyla et al. "Review of Particle Physics". In: *PTEP* 2020 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[4] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716.1 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020.

[5] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett. B* 716.1 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021.

[6] E. Noether. "Invariante Variationsprobleme". ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257. URL: http://eudml.org/doc/59024.

[7] Wikimedia Commons. *File:Quark confinement.svg -Wikimedia Commons, the free media repository*. [Online; accessed 25-September-2023]. 2020. URL: https://commons.wikimedia.org/w/index.php?title=File:Quark_confinement.svg.

[8] R. P. Feynman. "The Behavior of Hadron Collisions at Extreme Energies". In: *Special Relativity and Quantum Theory: A Collection of Papers on the Poincaré Group*. Ed. by M. E. Noz and Y. S. Kim. Dordrecht: Springer Netherlands, 1988, p. 289. DOI: 10.1007/978-94-009-3051-3_25.

[9] V. N. Gribov and L. N. Lipatov. "Deep inelastic $e\,p$ scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[10] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language". In: *Nucl. Phys. B* 126.2 (1977), p. 298. DOI: https://doi.org/10.1016/0550-3213(77)90384-4.

[11] Y. L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.” In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[12] C.S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105.4 (1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.

[13] N. Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10.12 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[14] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Phys. Rev. D* 2.7 (1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[15] M. Kobayashi and T. Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[16] S. Glashow. “The renormalizability of vector meson interactions.” In: *Nucl. Phys.* 10 (1959), p. 107. DOI: 10.1016/0029-5582(59)90196-8.

[17] A. Salam and J. C. Ward. “Weak and electromagnetic interactions”. In: *Nuovo Cim* 11 (1959), pp. 568–577. DOI: 10.1007/BF02726525.

[18] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19.21 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[19] Wikimedia Commons. *File:Mecanismo de Higgs PH.png - Wikimedia Commons, the free media repository*. [Online; accessed 25-September-2023]. 2021. URL: https://commons.wikimedia.org/w/index.php?title=File:Mecanismo_de_Higgs_PH.png.

[20] ATLAS Collaboration. *Standard Model Summary Plots March 2021.* Tech. rep. ATL-PHYS-PUB-2021-005. CERN, 2021. URL: http://cdsweb.cern.ch/record/2758261.

[21] ATLAS Collaboration. “Measurement of differential cross sections and W^+/W^- cross-section ratios for W boson production in association with jets at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *JHEP* 05 (2018). [Erratum: JHEP 10, 048 (2020)], p. 77. DOI: 10.1007/JHEP05(2018)077. arXiv: 1711.03296 [hep-ex].

[22] M. Brice. “Aerial View of the CERN taken in 2008.” 2008. URL: https://cds.cern.ch/record/1295244.

[23] L. Evans and P. Bryant. “LHC Machine”. In: *JINST* 3.S08001 (2008). DOI: 10.1088/1748-0221/3/08/S08001.

[24] E. Mobs. “The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019”. In: (2019). General Photo. URL: https://cds.cern.ch/record/2684277.

[25] ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *JINST* 3.S08002 (2008). DOI: 10.1088/1748-0221/3/08/S08002.

[26] LHCb Collaboration. “The LHCb Detector at the LHC”. In: *JINST* 3.S08005 (2008). DOI: 10.1088/1748-0221/3/08/S08005.

[27] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *JINST* 3.S08004 (2008). DOI: 10.1088/1748-0221/3/08/S08004.

[28] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3.S08003 (2008). DOI: 10.1088/1748-0221/3/08/S08003.

[29] W. Herr and B. Muratori. “Concept of luminosity”. In: *CAS - CERN Accelerator School: Intermediate Accelerator Physics* (2006), p. 361. DOI: 10.5170/CERN-2006-002.361. URL: https://cds.cern.ch/record/941318.

[30] ATLAS Collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC”. In: (2022). arXiv: 2212.09379 [hep-ex].

[31] ATLAS Collaboration. Public ATLAS Luminosity Results for Run-2 of the LHC. 2019.

[32] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3*. 2023. arXiv: 2305.16623 [physics.ins-det].

[33] *ATLAS inner detector: Technical Design Report, 1*. Technical design report. ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063.

[34] S. Haywood et al. *ATLAS inner detector: Technical Design Report, 2*. Technical design report. ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331064.

[35] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996. DOI: 10.17181/CERN.FWRW.FOOQ. URL: http://cds.cern.ch/record/331061.

[36] *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996. DOI: 10.17181/CERN.JRBJ.7O28. URL: http://cds.cern.ch/record/331062.

[37] *ATLAS magnet system: Technical Design Report, 1*. Technical design report. ATLAS. Geneva: CERN, 1997. DOI: 10.17181/CERN.9O5C.VDTM. URL: https://cds.cern.ch/record/338080.

[38] *ATLAS central solenoid: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1997. DOI: 10.17181/CERN.ZZVJ.2JYE. URL: https://cds.cern.ch/record/331067.

[39] A. Yamamoto et al. “The ATLAS central solenoid”. In: *Nucl. Instrum. Methods Phys. Res., A* 584.1 (2008), pp. 53–74. DOI: 10.1016/j.nima.2007.09.047.

[40] J. P. Badiouet al. *ATLAS barrel toroid: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1997. DOI: 10.17181/CERN.RF2A.CP5T.

[41] *ATLAS end-cap toroids: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1997. DOI: 10.17181/CERN.P03D.WQLV. URL: https://cds.cern.ch/record/331066.

[42] P. Jenni et. al. *ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity*. Technical design report. ATLAS. Geneva: CERN, 2008. DOI: 10.17181/CERN-LHCC-2008-004. URL: http://cds.cern.ch/record/1095847.

[43] G. Avoni et al. "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS". In: *JINST* 13.07 (2018), P07017. DOI: 10.1088/1748-0221/13/07/P07017.

[44] P. Jenniet al. *Zero Degree Calorimeters for ATLAS*. Tech. rep. Geneva: CERN, 2007. URL: http://cds.cern.ch/record/1009649.

[45] N. Wermes and G. Hallewel. *ATLAS pixel detector: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1998. URL: https://cds.cern.ch/record/381263.

[46] ATLAS Collaboration. *The Silicon Microstrip Sensors of the ATLAS SemiConductor Tracker*. Tech. rep. 1. Geneva: CERN, 2007. DOI: 10.1016/j.nima.2007.04.157. URL: https://cds.cern.ch/record/1019885.

[47] ATLAS Collaboration. "The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance". In: *JINST* 3 (2008), P02013. DOI: 10.1088/1748-0221/3/02/P02013. URL: https://cds.cern.ch/record/1094549.

[48] ATLAS Collaboration. "Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2". In: *Eur. Phys. J. C* 77.10 (2017). DOI: 10.1140/epjc/s10052-017-5225-7. arXiv: 1704.07983 [hep-ex].

[49] M. Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. 2010. URL: https://cds.cern.ch/record/1291633.

[50] ATLAS Collaboration. "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 81 (2021), p. 689. DOI: 10.1140/epjc/s10052-021-09402-3. arXiv: 2007.02645 [hep-ex].

[51] B. Aubert et al. "Construction, assembly and tests of the ATLAS electromagnetic barrel calorimeter". In: *Nucl. Instrum. Methods Phys. Res., A* 558 (2006), pp. 388–418. DOI: 10.1016/j.nima.2005.11.212. URL: http://cds.cern.ch/record/883909.

[52] M. Aleksa et al. "Construction, assembly and tests of the ATLAS electromagnetic end-cap calorimeters". In: *JINST* 3 (2008), P06002. DOI: 10.1088/1748-0221/3/06/P06002. URL: http://cds.cern.ch/record/1114196.

[53] M. L. Andrieux et al. "Construction and test of the first two sectors of the ATLAS barrel liquid argon presampler". In: *Nucl. Instrum. Methods Phys. Res., A* 479.2-3 (2002), pp. 316–33. DOI: 10.1016/S0168-9002(01)00943-3. URL: https://cds.cern.ch/record/567326.

[54] D. M. Gingrich et al. *Construction, assembly and testing of the ATLAS hadronic end-cap calorimeter*. Tech. rep. Geneva: CERN, 2007. DOI: 10.1088/1748-0221/2/05/P05005. URL: http://cds.cern.ch/record/1026994.

[55] R. S. Orr. "The ATLAS Forward Calorimeter". In: (2011). URL: https://cds.cern.ch/record/1386649.

[56] T. Kawamoto et al. *New Small Wheel Technical Design Report*. Tech. rep. ATLAS New Small Wheel Technical Design Report. 2013. URL: https://cds.cern.ch/record/1552862.

[57] ATLAS Collaboration. "Performance of the ATLAS Trigger System in 2010". In: *Eur. Phys. J. C* 72 (2012), p. 1849. DOI: 10.1140/epjc/s10052-011-1849-1. arXiv: 1110.1530 [hep-ex].

[58] ATLAS Collaboration. "Performance of the ATLAS Trigger System in 2015". In: *Eur. Phys. J. C* 77.5 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[59] ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *JINST* 15.10 (2020), P10004. DOI: 10.1088/1748-0221/15/10/P10004. arXiv: 2007.12539 [physics.ins-det].

[60] ATLAS Collaboration. ATLAS Event Displays Repository. 2023.

[61] ATLAS Collaboration (Tracking CP). ATLAS Tracking Software Tutorial. Accessed: 16-02-2024. ATLAS Internal. 2023.

[62] R. Frühwirth. "Application of Kalman Filtering to Track and Vertex Fitting," in: *Nucl. Instrum. Meth.* A.262 (1987). DOI: 10.1016/0168-9002(87)90887-4.

[63] ATLAS Collaboration. "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s}$=13 TeV". In: *Eur. Phys. J. C* 81 (2021), p. 578. DOI: 10.1140/epjc/s10052-021-09233-2. arXiv: 2012.00578 [hep-ex].

[64] W. Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*. Tech. rep. Geneva: CERN, 2008. URL: https://cds.cern.ch/record/1099735.

[65] ATLAS Collaboration. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J.* C.77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].

[66] ATLAS Collaboration. "Studies of the muon momentum calibration and performance of the ATLAS detector with pp collisions at $\sqrt{s}$=13 TeV". In: *Eur. Phys. J. C* 83 (2023), p. 686. DOI: 10.1140/epjc/s10052-023-11584-x. arXiv: 2212.07338 [hep-ex].

[67] ATLAS Collaboration. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data". In: *JINST* 14.12 (2019), P12006–P12006. DOI: 10.1088/1748-0221/14/12/p12006. arXiv: 1908.00005 [hep-ex].

[68] *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*. Tech. rep. Geneva: CERN, 2012. URL: https://cds.cern.ch/record/1449796.

[69] ATLAS Collaboration. "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV". In: *Eur. Phys. J.* C.79 (2019), p. 639. DOI: 10.1140/epjc/s10052-019-7140-6. arXiv: 1902.04655 [hep-ex].

[70] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data”. In: *JINST* 14.03 (2019), P03017–P03017. DOI: 10.1088/1748-0221/14/03/p03017. arXiv: 1812.03848 [hep-ex].

[71] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *Eur. Phys. J. C* 77 (2017), p. 466. DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485 [hep-ex].

[72] G. P. Salam M. Cacciari and G. Soyez. “The anti-k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[73] S. D. Ellis and D. E. Soper. “Successive combination jet algorithm for hadron collisions”. In: *Phys. Rev, D* 48.7 (1993), pp. 3160–3166. DOI: 10.1103/physrevd.48.3160. arXiv: hep-ph/9305266.

[74] M. Wobisch and T. Wengler. *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*. 1999. arXiv: hep-ph/9907280.

[75] G. Corcella et al. *HERWIG 6.5 Release Note*. 2005. arXiv: hep-ph/0210213.

[76] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS detector”. In: *Phys. Rev. D* 96.7 (2017), p. 072002. DOI: 10.1103/physrevd.96.072002. arXiv: 1703.09665 [hep-ex].

[77] T. Sjöstrand et al. “A Brief Introduction to PYTHIA 8.1”. In: *Comput.Phys.Commun* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ex].

[78] ATLAS Collaboration. “Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$ = 7 TeV.” In: *Eur. Phys. J. C* 73 (2013), p. 2304. DOI: 10.1140/epjc/s10052-013-2304-2. arXiv: 1112.6426 [hep-ex].

[79] ATLAS Collaboration. “ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset”. In: *Eur. Phys. J. C* 83 (2023), p. 681. DOI: 10.1140/epjc/s10052-023-11699-1. arXiv: 2211.16345 [hep-ex].

[80] ATLAS Collaboration. “ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$=13 TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex].

[81] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s}$=13TeV”. In: *Eur. Phys. J. C* 78 (2018), p. 903. DOI: 10.1140/epjc/s10052-018-6288-9. arXiv: 1802.08168 [hep-ex].

[82] ATLAS Collaboration. “Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector”. In: *Eur. Phys. J. C* 77 (2017), p. 580. DOI: 10.1140/epjc/s10052-017-5081-5. arXiv: 1705.02211 [hep-ex].

[83] ATLAS Collaboration. “Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”. In: *Eur. Phys. J. C* 76 (2016), p. 581. DOI: 10.1140/epjc/s10052-016-4395-z. arXiv: 1510.03823 [hep-ex].

[84] *Expected pileup values at the HL-LHC*. Tech. rep. Geneva: CERN, 2013. URL: https://cds.cern.ch/record/1604492.

[85] ATLAS Collaboration. Public ATLAS Luminosity Results for Run-3 of the LHC. 2023.

[86] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2009. arXiv: physics/0703039 [physics.data-an].

[87] P. Nason. “A New method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 2004.11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146; P. Nason S. Frixione and C. Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 2007.11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph]; S. Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 2010.6 (2010), p. 043. DOI: 10.1007/jhep06(2010)043. arXiv: 1002.2581 [hep-ph].

[88] T. Sjöstrandet al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[89] ATLAS Collaboration. “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector”. In: *JHEP* 2014.09 (2014), p. 145. DOI: 10.1007/jhep09(2014)145. arXiv: 1406.3660 [hep-ex].

[90] Pumplin et al. “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis”. In: *JHEP* 2002.07 (2002), p. 012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: arXiv:hep-ph/0201195.

[91] *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2206965.

[92] Ball et al. “Parton distributions with LHC data”. In: *Nucl. Phys. B* 867.2 (2013), p. 244. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[93] *Object-based missing transverse momentum significance in the ATLAS detector*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2630948.

[94] ATLAS Collaboration. “Measurements of the W production cross sections in association with jets with the ATLAS detector”. In: *Eur. Phys. J. C* 75.82 (2015). DOI: 10.1140/epjc/s10052-015-3262-7. arXiv: 1409.8639v2 [hep-ex].

[95] CMS Collaboration. “Measurement of the differential cross sections for the associated production of a W boson and jets in proton-proton collisions at $\sqrt{s}$=13 TeV”. In: *Phys. Rev. D* 96.7 (2017), p. 072005. DOI: 10.1103/physrevd.96.072005. arXiv: 1707.05979 [hep-ex].

[96] CMS Collaboration. “Measurement of electroweak production of a W boson in association with two jets in proton–proton collisions at $\sqrt{s}$=13 TeV”. In: *Eur. Phys. J. C* 80.1 (2020), p. 43. DOI: 10.1140/epjc/s10052-019-7585-7. arXiv: 1903.04040 [hep-ex].

[97] E. Bothman et al. “Event Generation with Sherpa 2.2”. In: *SciPost Phys.* 7.034 (2019). DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].

[98] B. Andersson et al. “Parton fragmentation and string dynamics”. In: *Phys. Rept.* 97 (1983), p. 31. DOI: 10.1016/0370-1573(83)90080-7.

[99] J.-C. Winter, F. Krauss, and G. Soff. “A modified cluster-hadronisation model”. In: *Eur. Phys. J. C* 36 (2004), p. 381. DOI: 10.1140/epjc/s2004-01960-8. arXiv: hep-ph/0311085.

[100] ATLAS Collaboration. “Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment”. In: *JHEP* 2022.08 (2022). DOI: 10.1007/jhep08(2022)089. arXiv: 2112.09588 [hep-ex].

[101] T. Gleisberg and S. Höche. “Comix, a new matrix element generator”. In: *JHEP* 2008.12 (2008), p. 39. DOI: 10.1088/1126-6708/2008/12/039. arXiv: 0808.3674 [hep-ph].

[102] NNPDF Collaboration. “Parton distributions for the LHC Run II”. In: *JHEP* 2015.40 (2015). DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[103] S. Carrazza et al. “An unbiased Hessian representation for Monte Carlo PDFs”. In: *Eur. Phys. J. C* 75 (2015), p. 369. DOI: 10.1140/epjc/s10052-015-3590-7. arXiv: 1505.06736 [hep-ph].

[104] F. Buccioni et al. “OpenLoops 2”. In: *Eur. Phys. J. C* 79 (2019), p. 866. DOI: 10.1140/epjc/s10052-019-7306-2. arXiv: 1907.13071 [hep-ph].

[105] F. Cascioli, P. Maierhöfer, and S. Pozzorini. “Scattering Amplitudes with Open Loops”. In: *Phys. Rev. Lett.* 108.11 (2012), p. 111601. DOI: 10.1103/physrevlett.108.111601. arXiv: 1111.5206 [hep-ph].

[106] F. Buccioni, S. Pozzorini, and M. Zoller. “On-the-fly reduction of open loops”. In: *Eur. Phys. J. C* 78 (2018), p. 70. DOI: 10.1140/epjc/s10052-018-5562-1. arXiv: 1710.11452 [hep-ph].

[107] A. Denner, S. Dittmaier, and L. Hofer. “Collier: A fortran-based complex one-loop library in extended regularizations”. In: *Computer Physics Communications* 212 (2017), p. 70. DOI: 10.1016/j.cpc.2016.10.013. eprint: 1604.06792 (hep-ph).

[108] D.R. Yennie, S.C. Frautschi, and H. Suura. “The infrared divergence phenomena and high-energy processes”. In: *Annals of Physics* 13.3 (1961), pp. 379–452. DOI: https://doi.org/10.1016/0003-4916(61)90151-8.

[109] M. Schönherr and F. Krauss. “Soft photon radiation in particle decays in SHERPA”. In: *JHEP* 2008.12 (2008), pp. 018–018. DOI: 10.1088/1126-6708/2008/12/018. arXiv: 0810.5071 [hep-ph].

[110] S. Schumann and F. Krauss. “A parton shower algorithm based on Catani-Seymour dipole factorisation”. In: *JHEP* 2008.03 (2008), pp. 038–038. DOI: 10.1088/1126-6708/2008/03/038. arXiv: 0709.1027 [hep-ph].

[111] S. Höche et al. “A critical appraisal of NLO+PS matching methods”. In: *JHEP* 2012.9 (2012). DOI: 10.1007/jhep09(2012)049. arXiv: 1111.1220 [hep-ph].

[112] S. Catani et al. “QCD Matrix Elements + Parton Showers”. In: *JHEP* 2001.11 (2001), pp. 063–063. DOI: 10.1088/1126-6708/2001/11/063. arXiv: hep-ph/0109231.

[113] S. Höche et al. “QCD matrix elements and truncated showers”. In: *JHEP* 2009.05 (2009), pp. 053–053. DOI: 10.1088/1126-6708/2009/05/053. arXiv: 0903.1219 [hep-ph].

[114] S. Höche et al. “QCD matrix elements + parton showers. The NLO case”. In: *JHEP* 2013.4 (2013). DOI: 10.1007/jhep04(2013)027. arXiv: 1207.5030 [hep-ph].

[115] 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”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ex].

[116] V. Hirschi et al. “Automation of one-loop QCD computations”. In: *JHEP* 2011.5 (2011). DOI: 10.1007/jhep05(2011)044. arXiv: 1103.0621 [hep-ex].

[117] R. D. Ball et al. “Parton distributions from high-precision collider data”. In: *Eur. Phys. J. C* 77 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[118] V. Bertone et al. “Illuminating the photon content of the proton within a global PDF analysis”. In: *SciPost Physics* 5.1 (2018). DOI: 10.21468/scipostphys.5.1.008. arXiv: 1712.07053[hep-ph].

[119] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. Tech. rep. ATL-PHYS-PUB-2014-021. CERN, 2014. URL: https://cds.cern.ch/record/1966419.

[120] R. Frederix and S. Frixione. “Merging meets matching in MC@NLO”. In: *JHEP* 2012.12 (2012). DOI: 10.1007/jhep12(2012)061. arXiv: 1209.6215 [hep-ph].

[121] ATLAS Collaboration. “Performance of the ATLAS muon triggers in Run 2”. In: *JINST* 15 (2020), P09015. DOI: 10.1088/1748-0221/15/09/p09015. arXiv: 2004.13447 [physics.ins-det].

[122] *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*. Tech. rep. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2037702.

[123] S. Frixione, P. Nason, and G. Ridolfi. “A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. In: *JHEP* 2007.09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

[124] S. Alioli et al. “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions”. In: *JHEP* 2009.09 (2009). [Erratum: JHEP 02, 011 (2010)], p. 111. DOI: 10.1088/1126-6708/2009/09/111. arXiv: 0907.4076 [hep-ph].

[125] Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev. D* 82.07 (2010), p. 074024. DOI: 10.1103/physrevd.82.074024. arXiv: 1007.2241 [hep-ph].

[126] ATLAS Collaboration. *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2014-058. CERN, 2014. URL: https://cds.cern.ch/record/1951336.

[127] G. Pownall. “QCD analysis and measurement of W boson production in association with jets at the ATLAS detector”. Presented 15 Jan 2021. 2020. URL: https://cds.cern.ch/record/2774953.

[128] T. Adye. *Unfolding algorithms and tests using RooUnfold*. 2011. arXiv: 1105.1160 [physics.data-an].

[129] G. D’Agostini. *A multidimensional unfolding method based on Bayes’ Theorem*. Tech. rep. Hamburg: DESY, 1994. URL: https://cds.cern.ch/record/265717.

[130] G. D’Agostini. *Improved iterative Bayesian unfolding*. 2010. arXiv: 1010.0632 [physics.data-an]

[131] T. Adye. *Unfolding algorithms and tests using RooUnfold*. 2011. arXiv: 1105.1160 [physics.data-an].

[132] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$=13 TeV.” In: *Eur. Phys. J. C* 76 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

[133] W. Buttinger. *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*. Tech. rep. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2014726.

[134] ATLAS Collaboration. “Measurement of the W-boson mass in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 78.2 (2018). [Erratum: Eur. Phys. J.C78.11, (2018), 898], p. 110. DOI: 10.1140/epjc/s10052-017-5475-4. arXiv: 1701.07240 [hep-ex].

[135] LHCb Collaboration. “Measurement of the W boson mass”. In: *JHEP* 2022.1 (2022). DOI: 10.1007/jhep01(2022)036. arXiv: 2109.01113 [hep-ex].

[136] LHCb Collaboration. “Measurement of the forward-backward asymmetry in $Z/\gamma^* \to \mu^+\mu^-$ decays and determination of the effective weak mixing angle”. In: *JHEP* 2015.11 (2015). DOI: 10.1007/jhep11(2015)190. arXiv: 1509.07645 [hep-ex].

[137] CMS Collaboration. “Measurement of the weak mixing angle using the forward-backward asymmetry of Drell-Yan events in pp collisions at 8 TeV”. In: *Eur. Phys. J. C* 78.9 (2018). DOI: 10.1140/epjc/s10052-018-6148-7. arXiv: 1806.00863 [hep-ex].

[138] D. de Florian et al. *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*. 2017. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[139] R. Devenish and A. Cooper-Sarkar. "Extraction of the parton denisties". In: *Deep Inelastic Scattering*. Oxford University Press, 2004, pp. 165–174.

[140] H. Abramowicz et al. "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data". In: *Eur. Phys. J. C* 75 (2015), p. 580. DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 [hep-ex].

[141] A. Kayis-Topaksu et al. "Measurement of charm production in neutrino charged-current interactions". In: *New J. Phys.* 13.9 (2011), p. 093002. DOI: 10.1088/1367-2630/13/9/093002. arXiv: 1107.0613 [hep-ex].

[142] D. Mason et al. "Measurement of the Nucleon Strange-Antistrange Asymmetry at Next-to-Leading Order in QCD from NuTeV Dimuon Data". In: *Phys. Rev. Lett.* 99 (19 2007), p. 192001. DOI: 10.1103/PhysRevLett.99.192001.

[143] BCDMS Collaboration. "A high statistics measurement of the deuteron structure functions $F2(x, Q^2)$ and R from deep inelastic muon scattering at high Q^2". In: *Phys. Lett. B* 237.3 (1990), p. 592. DOI: https://doi.org/10.1016/0370-2693(90)91231-Y.

[144] The New Muon Collaboration (NMC). "Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T". In: *Nucl. Phys. B* 483.1 (1997), p. 3. DOI: https://doi.org/10.1016/S0550-3213(96)00538-X. arXiv: hep-ph/9610231.

[145] NuSea Collaboration. *Absolute Drell-Yan Dimuon Cross Sections in 800 GeV/c pp and pd Collisions*. 2003. arXiv: hep-ex/0302019.

[146] FNAL E866/NuSea Collaboration. "Improved measurement of the $\bar{d}/\bar{u}$ asymmetry in the nucleon sea". In: *Phys. Rev. D* 64 (2001), p. 052002. DOI: 10.1103/PhysRevD.64.052002. arXiv: hep-ex/0103030.

[147] CDF Collaboration. "Measurement of $d\sigma/dy$ of Drell–Yan e^+e^- pairs in the Z mass region from $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV". In: *Phys. Lett. B* 692.4 (2010), p. 232. DOI: https://doi.org/10.1016/j.physletb.2010.06.043. arXiv: 0908.3914 [hep-ex].

[148] D0 Collaboration. "Measurement of the shape of the boson rapidity distribution for $p\bar{p} \to X/\gamma^* \to ee^- + X$ events produced at $\sqrt{s}$ of 1.96 TeV". In: *Phys. Rev. D* 76 (1 2007), p. 012003. DOI: 10.1103/PhysRevD.76.012003. arXiv: hep-ex/0702025.

[149] CDF Collaboration. "Measurement of the inclusive jet cross section at the Fermilab Tevatron $p\bar{p}$ collider using a cone-based jet algorithm". In: *Phys. Rev. D* 78 (5 2008), p. 052006. DOI: 10.1103/PhysRevD.78.052006. arXiv: 0807.2204 [hep-ex].

[150] S. Bailey et al. "Parton distributions from LHC, HERA,Tevatron and fixed target data: MSHT20 PDFs". In: *Eur. Phys. J. C* 81 (2021), p. 341. DOI: 10.1140/epjc/s10052-021-09057-0. arXiv: 2012.04684 [hep-ph].

[151] T.-J. Hou et al. “New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC”. In: *Phys. Rev. D* 103 (2021), p. 014013. DOI: 10.1103/physrevd.103.014013. arXiv: 1912.10053 [hep-ph].

[152] J.M. Butterworth and T. Carli. “QCD uncertainties at the LHC and the implications of HERA”. In: (2004). Contributed to the Proceedings of DIS04, Strbske Pleso, Slovakia, p. 22. DOI: 10.1103/PhysRevD.71.024010. arXiv: hep-ph/0408061.

[153] ATLAS Collaboration. “Precision measurement and interpretation of inclusive W^+, W^- and Z/γ^* production cross sections with the ATLAS detector”. In: *Eur. Phys. J. C* 77 (2017), p. 367. DOI: 10.1140/epjc/s10052-017-4911-9. arXiv: 1612.03016 [hep-ex].

[154] M. Goncharov et al. “Precise measurement of dimuon production cross sections in muon neurtrino Fe and muon antineutrino Fe deep inelastic scattering at the Fermilab Tevatron”. In: *Phys. Rev. D* 64.11 (2001). DOI: 10.1103/physrevd.64.112006. arXiv: hep-ex/0102049.

[155] NOMAD Collaboration. “A precision measurement of charm dimuon production in neutrino interactions from the NOMAD experiment”. In: *Nucl. Physics B.* 876.2 (2013), p. 339. DOI: https://doi.org/10.1016/j.nuclphysb.2013.08.021. arXiv: 1308.4750 [hep-ex].

[156] ATLAS Collaboration. “Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector”. In: *Eur. Phys. J. C* 76 (2016), p. 538. DOI: 10.1140/epjc/s10052-016-4366-4. arXiv: 1511.04716 [hep-ex].

[157] ATLAS Collaboration. “Measurement of top quark pair differential cross sections in the dilepton channel in pp collisions at $\sqrt{s}$ 7 TeV and 8 TeV with ATLAS”. In: *Phys. Rev. D* 94.9 (2016). DOI: 10.1103/physrevd.94.092003. arXiv: 1607.07281 [hep-ex].

[158] ATLAS Collaboration. *Determination of the parton distribution functions of the proton from ATLAS measurements of differential W and Z/γ^* and $t\bar{t}$ cross sections*. Tech. rep. ATL-PHYS-PUB-2018-017. CERN, 2018. URL: https://cds.cern.ch/record/2633819/.

[159] ATLAS Collaboration. “Determination of the parton distribution functions of the proton from ATLAS measurements of differential $W^{\pm}$ and Z boson production in association with jets”. In: *JHEP* 07 (2021), p. 223. DOI: 10.1007/JHEP07(2021)223. arXiv: 2101.05095 [hep-ex].

[160] ATLAS Collaboration. “Measurement of the cross-section and charge asymmetry of W bosons produced in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 79 (2019), p. 760. DOI: 10.1140/epjc/s10052-019-7199-0. arXiv: 1904.05631 [hep-ex].

[161] ATLAS Collaboration. “Measurement of the Drell-Yan triple-differential cross section in pp collisions at $\sqrt{s}$ = 8 TeV”. In: *JHEP* 12 (2017), p. 59. DOI: 10.1007/jhep12(2017)059. arXiv: 1710.05167 [hep-ex].

[162] ATLAS Collaboration. “Measurement of the inclusive cross-section for the production of jets in association with a Z boson in proton-proton collisions at 8 TeV using the ATLAS detector”. In: *Eur. Phys. J. C* 79 (2019), p. 847. DOI: 10.1140/epjc/s10052-019-7321-3. arXiv: 1907.06728 [hep-ex].

[163] F. Faura et al. “The strangest proton?” In: *Eur. Phys. J. C* 80 (2020), p. 1168. DOI: 10.1140/epjc/s10052-020-08749-3. arXiv: 2009.00014 [hep-ph].

[164] F. D. Aaron et al. “Combined measurement and QCD analysis of the inclusive e $\pm$ p scattering cross sections at HERA”. In: *JHEP* 2010.1 (2010). DOI: 10.1007/jhep01(2010)109. arXiv: 0911.0884 [hep-ex].

[165] A.D. Martin et al. “Uncertainties of predictions from parton distributions I: Experimental errors”. In: *Eur. Phys. J. C* 28 (2003), p. 455. DOI: 10.1140/epjc/s2003-01196-2. arXiv: hep-ph/0211080.

[166] J. Pumplin et al. “Uncertainties of predictions from parton distribution functions. II. The Hessian method”. In: *Phys. Rev. D* 65.1 (2001). DOI: 10.1103/physrevd.65.014013. arXiv: hep-ph/0101032.

[167] ATLAS Collaboration. “Measurements of top-quark pair differential and double-differential cross-sections in the l+jets channel with pp collisions at $\sqrt{s}$= 13 TeV using the ATLAS detector”. In: *Eur. Phys. J. C* 79 (2016), p. 1028. DOI: 10.1140/epjc/s10052-019-7525-6. arXiv: 1908.07305 [hep-ex].

[168] ATLAS Collaboration. “Measurement of the ratio of cross sections for inclusive isolated-photon production in pp collisions at $\sqrt{s}$ = 13 and 8 TeV with the ATLAS detector”. In: *JHEP* 04 (2019), p. 93. DOI: 10.1007/jhep04(2019)093. arXiv: 1901.10075 [hep-ex].

[169] ATLAS Collaboration. “Measurement of the inclusive jet cross-sections in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector”. In: *JHEP* 09 (2017), p. 20. DOI: 10.1007/jhep09(2017)020. arXiv: 1706.03192 [hep-ex].

[170] S. Alekhin et al. “HERAFitter”. In: *Eur. Phys. J. C* 75 (2015), p. 304. DOI: 10.1140/epjc/s10052-015-3480-z. arXiv: 1410.4412 [hep-ex].

[171] “Minuit - a system for function minimization and analysis of the parameter errors and correlation”. In: *Comput. Phys. Commun.* 10 (1975), p. 343. DOI: 10.1016/0010-4655(75)90039-9.

[172] Carli et al. “A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: the APPLGRID project”. In: *Eur. Phys. J. C* 66.3–4 (2010), p. 503. DOI: 10.1140/epjc/s10052-010-1255-0. arXiv: 0911.2985 [hep-ph].

[173] T. Kluge et al. “Fast pQCD Calculations for PDF Fits”. In: *Proceeding of the XIV International Workshop on Deep Inelastic Scattering and Related Subjects*. 2006, p. 483. DOI: 10.1142/9789812706706_0110. arXiv: hep-ph/0609285.

[174] xFitter Developers’ team. *Impact of low-x resummation on QCD analysis of HERA data*. 2018. arXiv: 1802.00064 [hep-ph].

[175] R. D. Ball et al. “Parton distributions with small-x resummation: evidence for BFKL dynamics in HERA data”. In: *Eur. Phys. J. C* 78.4 (2018), p. 321. DOI: 10.1140/epjc/s10052-018-5774-4. arXiv: 1710.05935 [hep-ph].

[176] H. Abramowicz et al. “Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA”. In: *Eur. Phys. J. C* 78 (2018), p. 473. DOI: 10.1140/epjc/s10052-018-5848-3. arXiv: 1804.01019 [hep-ex].

[177] A. Martin et al. “Parton distributions for the LHC”. In: *Eur. Phys. J. C* 63 (2009), p. 189. DOI: 10.1140/epjc/s10052-018-5848-3. arXiv: 0901.0002 [hep-ph].

[178] CMS Collaboration. “Measurements of the associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s}$ = 8 TeV”. In: *Eur. Phys. J. C* 82 (2022), p. 1094. DOI: 10.1140/epjc/s10052-022-10897-7. arXiv: 2112.00895 [hep-ex].

[179] J. Dove et al. “The asymmetry of antimatter in the proton”. In: *Nature* 590 (2021), p. 561. DOI: 10.1038/s41586-021-03282-z. arXiv: 2103.04024 [hep-ph].

[180] L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. “The impact of LHC jet data on the MMHT PDF fit at NNLO”. In: *Eur. Phys. J. C* 78 (2018), p. 248. DOI: 10.1140/epjc/s10052-018-5710-7. arXiv: 1711.05757 [hep-ph].

[181] C. Bierlich et al. “Robust Independent Validation of Experiment and Theory: Rivet version 3”. In: *SciPost Phys.* 8 (2020), p. 026. DOI: 10.21468/scipostphys.8.2.026. arXiv: 1912.05451 [hep-ph].

[182] O. Aberle et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2020. DOI: 10.23731/CYRM-2020-0010. arXiv: 1705.08830 [physics.acc-ph].

[183] R. Abdul Khalek et al. “Science Requirements and Detector Concepts for the Electron-Ion Collider”. In: *Nucl. Phys. A* 1026 (2022), p. 122447. DOI: 10.1016/j.nuclphysa.2022.122447. arXiv: 2103.05419 [physics.ins-det].