Measurement of the production of a vector boson in association with a charmed hadron in pp collisions at $\sqrt{s}$= 13 TeV with the ATLAS detector

by

César González Rentería

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Marjorie Shapiro, Chair
Professor Yury Kolomensky
Professor Karl Van Bibber

Fall 2023

Measurement of the production of a vector boson in association with a charmed hadron in pp collisions at $\sqrt{s}$= 13 TeV with the ATLAS detector

César González Rentería

Abstract

Measurement of the production of a vector boson in association with a charmed hadron in pp collisions at $\sqrt{s}$= 13 TeV with the ATLAS detector

by

César González Rentería

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Marjorie Shapiro, Chair

The production of a Vector Boson in association with a single charm quark is measured and presented here using 140 fb^{-1} of proton–proton collision data collected with the ATLAS detector at the Large Hadron Collider. Measurements are performed using either a Z boson or W boson produced with a charm meson. The charm quark is tagged by the presence of a charmed hadron, reconstructed with a secondary-vertex fit. The W boson is reconstructed from the decay to either an electron or a muon and the missing transverse momentum present in the event. The Z boson is reconstructed from the decay to either a pair of electrons or muons. The reconstruction modes used for charm hadron reconstruction are $D^+ \to K^-\pi^+\pi^+$ and $D^{*+} \to D^0\pi^+ \to (K^-\pi^+)\pi^+$ and the charge conjugate decays. The W+$D^{(*)}$ integrated and normalized differential cross-sections as a function of the pseudorapidity of the lepton from the W boson decay, and of the transverse momentum of the charmed hadron, are extracted from the data using a profile likelihood fit. The fiducial region is defined to include events passing $p_{\mathrm{T}}(e,\mu) > 30\,\mathrm{GeV}$, $|\eta(e,\mu)| < 2.5$, $p_{\mathrm{T}}(D^{(*)}) > 8\,\mathrm{GeV}$, and $|\eta(D^{(*)})| < 2.2$. The measured total fiducial cross-sections are: $\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}(W^-+D^{*+}) = 51.1 \pm 0.4\,(\mathrm{stat.})\,^{+1.9}_{-1.8}\,(\mathrm{syst.})\,\mathrm{pb}$, and $\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}(W^++D^{*-}) = 50.0 \pm 0.4\,(\mathrm{stat.})\,^{+1.9}_{-1.8}\,(\mathrm{syst.})\,\mathrm{pb}$. Results are compared with the predictions of next-to-leading-order quantum chromodynamics calculations performed using state-of-the-art parton distribution functions. Additionally, the ratio of charm to anti-charm production cross-sections is studied to probe the s-$\bar{s}$ quark asymmetry. The ratio is found to be $R_c^{\pm} = 0.971 \pm 0.006\,(\mathrm{stat.}) \pm 0.011\,(\mathrm{syst.})$. The ratio and cross-section measurements are consistent with the predictions obtained with parton distribution function sets that have a symmetric s-$\bar{s}$ sea, indicating that any s-$\bar{s}$ asymmetry in the Bjorken-x region relevant for this measurement is small. The Z+$D^{(*)}$ integrated and normalized differential cross-sections as a function of the transverse momentum of the charmed hadron, are extracted from the data using the same fit strategy. The fiducial region is defined to include events passing $p_{\mathrm{T}}(e,\mu) > 27\,\mathrm{GeV}$, $|\eta(e,\mu)| < 2.5$,

$p_{\mathrm{T}}(D^{(*)}) > 8\,\mathrm{GeV}$, and $|\eta(D^{(*)})| < 2.2$. The measured total fiducial cross-sections in the $Z+D^{(*)}$ analysis are $\sigma_{\mathrm{fid}}(Z^0+D^{\pm}) = 21.9 \pm 0.04\,(\mathrm{stat.})\,^{+1.5}_{-1.4}\,(\mathrm{syst.})\,\mathrm{pb}$, and $\sigma_{\mathrm{fid}}(Z^0+D^{*\pm}) = 31.1 \pm 0.23\,(\mathrm{stat.})\,^{+1.3}_{-1.3}\,(\mathrm{syst.})\,\mathrm{pb}$. These results are also compared with the predictions of next-to-leading-order quantum chromodynamics calculations performed using state-of-the-art parton distribution functions.

Dedicated to my wonderful family

Everything in this world that I accomplish I do for and because of you. Without your unending support I could not get as far in life as I've been blessed and honored to. En su apoyo y amor encontré la paz, humildad y sabiduría a enfrentar lo extraño y lo dificil. Del fondo de mi corazón, gracias totales.

Contents

List of Figures

List of Tables

Acknowledgments

There are too many people to name that I would do a disservice to them to name some, and leave others out but here's my best attempt. To those at UC Berkeley and LBL that have supported me and guided me through this whole process, I thank you for having the patience and energy to hold my hand through the strange world of particle physics. Particularly I'd like to thank my advisor Marjorie Shapiro for being not only an advisor but a friend that I could lean on to talk all things physics and of life. I'd also like to thank Greg Ottino, Miha Muŝkinja and Heather Gray for all of their support in taking on the charming physics we studied throughout these last few years. I'd like to thank those that were there for a large part of my journey at LBL: Maurice Garcia-Sciveres, Simone Pagan Griso, Timon Heim, Rebecca Carney, William Patrick McCormack, Sai Neha Santpur.

I'd like to thank Carlos Figueroa of Cabrillo College, who believed in me before most, and whom without, I would not have made it past the community college level. Thank you for not seeing me as a poor, uneducated cholillo attempting to reach to heights not meant for me and mine. My whole physics career and all of its successes can be attributed to your trust, belief and guidance.

I'd like to thank the loveliest, warmest ferns I met in my time at UCSC. Gabriela Aguilar, Paul Anderson, Astrid Anker, Yevgeniya Shtalenkova, your unending love, laughter and support got me through the dimmest of days and never failed to brighten my life no matter what state I was in.

And lastly, to my family, I love you, you are everything. Pedro González Ríos y Maria Victoria Rentería Félix, de orígenes humildes pudieron sembrar las semillitas que dieron fruto al hermoso jardín que es nuestra familia. Sin mucha educación formal, sin dinero, y sin materia pudieron llegar al colmo de lo más importante en la vida: amar sin límites. Que teniendo tantos hijos, nunca entendí como pudieron amarnos con todos sus corazones y en abundancia. Algo que ni mi doctorado ni años de "educación formal" me servirá a explicar. A mis queridos hermanos y hermanas: Ana, Vero, Gaby, Moni, Angelica, Jesus, Pedro, Ernesto y Eduardo. Juntos enfrentamos los días más oscuros, de lo que es tener poco, de lo dificil que es vivir en este mundo con los límites que nos cargaba. Con su amor y apoyo sé que podemos llegar adelante y confrontar lo que nos de la vida. Familia, al mundo puede ser que parezco ser el "estimado doctor" César González Rentería, pero en sus corazónes suelo ser siempre "Cesítar".

To those who have touched my life, no matter how big or small, I thank you deeply.

Chapter 1

Introduction

This dissertation includes two similar yet distinct measurements of vector bosons produced in association with a charm hadron. In particular these measurements are of a W boson or a Z boson plus a $D^{(*)}$ meson. These measurements are important contributions to understanding the proton structure, the modeling of vector bosons plus heavy flavor physics (V+HF) and a test of perturbative Quantum Chromodynamics (pQCD). For the $W+D^{(*)}$ analysis, only the parts of the analysis in which I was heavily involved are presented. This documents performing the full measurement of the $W+D^{*+}$ mode. The $W+D^{+}$ mode is included insofar as it informs the case where both production modes were used to construct an inclusive $D^{(*)}$ meson quantity. The full $W+D^{(*)}$ analysis can be found in Ref.[1]. The $Z+D^{(*)}$ analysis was performed by myself and is presented in full.

Establishing the Standard Model and theoretical context of these analyses is given in this chapter. Chapter 1.1 describes the Standard Model of particle physics. This includes an overview of the particle content and the Lagrangian which includes the information that describes particle interactions. Chapter 1.2 describes the physics of proton–proton collisions, including how cross-sections are measured and how parton distribution functions of the proton can be probed. Chapter 1.3 gives the context of the $V+D^{(*)}$ measurements and what measurements have been done by other collaborations.

The remainder of this dissertation is structured as follows. Chapter 2 introduces the ATLAS detector. The data and Monte Carlo simulation samples used in both analyses are discussed in Chapter 3. Chapter 4 describes the physics objects used and their selection criteria. The reconstruction and selection of charmed mesons are discussed in Chapter 5. The event selection is summarized in Chapter 6. Signal and background modeling are described in Chapter 7. Chapter 8 presents the method used to extract the $V+D^{(*)}$ differential cross-sections and Chapter 9 summarizes the relevant systematic uncertainties. The cross-section measurements and their comparison with theoretical predictions are presented in Chapter 10. Conclusions are provided in Chapter 11.

1.1 The Standard Model

The Standard Model of Particle Physics is the theory of the properties of particles and their interactions with one another. It is the most established and robust description of how particles behave in nature to date. It is, however, an incomplete picture of the universe: it cannot predict the Dark Matter abundance, Dark Energy, whether supersymmetric particles exist (SUSY theory), unification of forces, neutrino masses, and other unanswered questions. These concepts are placed under the umbrella of Beyond SM (BSM) theories.

1.1.1 The Particle Content of the Standard Model

The particle content of the SM can be broken down into three major types of particles: quarks, leptons, and gauge bosons. Figure 1.1 shows the full list of elementary particles of the SM. Quarks and leptons in the SM are elementary particles that have no further sub-structure. Quarks are fermions (1/2 spin) with fractional charge ($\pm$2/3e or $\pm$1/3e), varying masses, and carry a property known as color. There are three colors in this construction: red/anti-red, blue/anti-blue, green/anti-green. They are not found individually in nature and very quickly hadronize into a baryon (3 quark state) or a meson (quark–anti-quark state). Leptons are also fermions (1/2 spin) with an integer charge (0, $\pm$1), varying masses, and can exists independently in nature. Gauge bosons are the "interaction" particles in the sense that quarks and leptons do not directly interact with each other but indirectly through the exchange of these gauge bosons.

Each gauge boson interacts through one of the force forces found in nature: electric, weak, strong, gravitational. Note that at the energies discussed in this dissertation the electric and weak forces are treated as a singular electroweak force. The photon γ, $W^{\pm}$ and Z^0 bosons are spin-1 particles and are the electroweak mediators. They interact with any particle that has either charge or isospin (a combination or angular momentum and spin). Functionally, this means these bosons directly interact with all leptons and quarks. The strong force is mediated by bosons with spin-1 known as gluons. Gluons interact with particles which contain the property known as a color charge (they themselves carrying a color–anti-color charge). They therefore directly interact with themselves and with quarks only. All of these possible interactions between particles and gauge bosons are summarized in the SM Lagrangian Density.

The Higgs boson is a scalar (spin-0) neutral charge boson that is necessary to maintain the consistency of the SM. Without it, unitarity is not preserved, and the interaction probabilities of certain particles would blow up to infinity. The presence of the Higgs boson and the Higgs mechanism generates the masses of the electroweak gauge bosons which preserved the local gauge invariance of the SM. The Higgs boson directly interacts with any particle that contains mass.

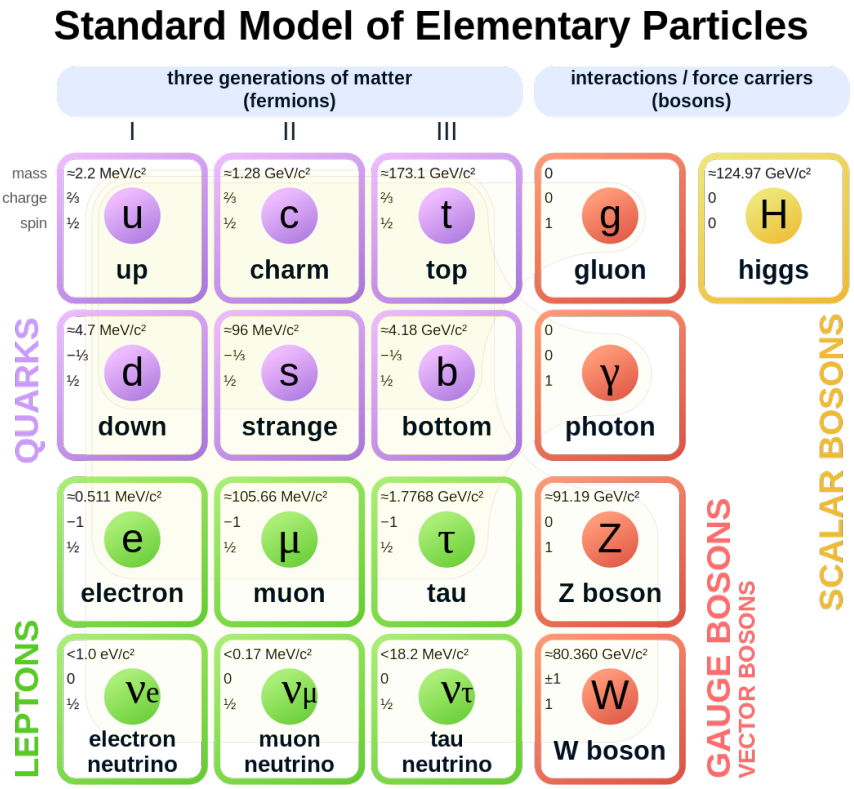


Figure 1.1: The particle content of the Standard Model.[2]

1.1.2 The Standard Model Lagrangian

The SM is established as a Quantum Field Theory (QFT) that encompasses the strong and electroweak forces but is missing the gravitational force. In QFT, particles are described as excitations of a quantum field that give rise to states physically manifesting into existence from vacuum. The full dynamics of the QFT of the SM can be summarized in its Lagrangian density ($\mathcal{L}$). The Lagrangian Density (Lagrangian for short) contains information of the fields and their kinematic properties. A complete Lagrangian contains only terms that conserve local gauge and global symmetries of the model they describe.

In the case of the SM, these symmetries can be written as: $SU(3)_c \times SU(2)_L \times U(1)_Y$. These terms are group theory representations of a non-Abelian Lie algebra. This Lie Algebra defines the number of generators (gauge bosons) in the group and its interactions through structure functions f. $SU(3)_c$ is the color symmetry group representing the quarks and gluons which contains 3 quarks, 3 anti-quarks and 8 gluons. $SU(2)_L \times U(1)_Y$ collectively represents the Electroweak group which is a left-handed theory (conserves left-handed interactions) and

hypercharge. In this formulation particles are either written as a doublet (for left-handed symmetry) or a singlet state (for hypercharge symmetry). There are 3+1 generators in this group that correspond to the W/Z bosons and the photon.

Once the Lagrangian is constructed, the full dynamics of the theory can be obtained from the Euler-Lagrange Equation (Equation 1.1):

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0 \tag{1.1}$$

where $\phi_i = \phi_i(x^\mu)$ represents a continuous quantity with a value at each point in space-time and can be a scalar, vector or tensor field.

The Lagrangian of the SM is written out in Equation 1.2:

$$\mathcal{L}_{SM} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (i\overline{\psi} \not{D} \psi + h.c.) + (\psi_i y_{ij} \psi_j \phi + h.c.) + |D_\mu H|^2 - V(H) \tag{1.2}$$

where "h.c." is the hermitian conjugate of a term. This Lagrangian is the condensed form of the full SM theory which contains all the information on what particles and gauge bosons interact with one another and the strength of their interactions. Every single term in the Lagrangian conserve the symmetries of the SM which is required to be a consistent theory.

The full Lagrangian can be broken down into a set number of terms that describe specific physics phenomenology and interactions in the SM. The first is the gauge bosons kinetic term shown in Equation 1.3:

$$\mathcal{L}_{gauge} = -\frac{1}{4} (G^a_{\mu\nu})^2 - \frac{1}{4} (W^a_{\mu\nu})^2 - \frac{1}{4} B^2_{\mu\nu} \tag{1.3}$$

This term represents the kinetic energy of the gauge bosons and contains an entry for each gauge boson: G for gluons, W for W/Z bosons, and B for the photon. The "a" designation represents a sum over all gluons in the G term (over all color combinations), and all three EW bosons (W^+, W^-, Z^0) in the W term. The field terms $F_{\mu\nu}$ (where F is either G, W, or B) is expanded to be written $F^a_{\mu\nu} = \partial_\mu F^a_\nu - \partial_\nu F^a_\mu + g f^{abc} F^b_\mu F^c_\nu$. This explicitly displays the kinetic term of the field (∂F) and the boson-boson self-interaction term ($gfFF$). g is a coupling constant which gives the strength of an interaction, while f^{abc} is the structure function associated with the group.

The next term is the kinetic energy of the fermions shown in Equation 1.4:

$$\mathcal{L}_{fermions} = \sum_{fermions} i\overline{\psi} \left(\partial_\mu - i g_c q_c G^a_\mu \frac{\lambda^a}{2} - i g_L q_L W^a_\mu \frac{\sigma^a}{2} - i g_Y q_Y B_\mu \right) \psi + h.c. \tag{1.4}$$

This term has the covariant derivative $\not{D}$ fully written out as $\not{D} \propto \partial - k_1 G - k_2 W - k_3 B$. The use of the covariant derivative instead of solely a standard derivative is necessary to

conserve local gauge symmetry. By requiring this conservation, the derivative introduces terms with the gauge bosons. Thus the kinetic term for fermions goes from a simple $\overline{\psi}\partial\psi$ to terms that involve an interaction of these fermions to the gauge bosons like $\overline{\psi}G\psi$. The strength of the interactions are given by the constants present in each gauge boson term specific to each symmetry group (color, left-handedness, hypercharge): g is the coupling constant, q is the charge and λ and σ are the generating matrices. These are the Gell-Mann matrices for $SU(3)_c$ and the Pauli matrices for $SU(2)_L$.

The following term is the Yukawa term which is written in Equation 1.5:

$$\mathcal{L}_{Yukawa} = (-y\overline{L}^i_L H l^i_R - Y^d_{ij}\overline{Q}^i_L H d^j_R - Y^u_{ij}\overline{Q}^i_L \tilde{H} u^j_R) + h.c. \tag{1.5}$$

This displays multiple terms that all have the form of $Y\phi^{particle}_L H \phi^{particle}_R$. These terms are the interactions of the fermion fields to the Higgs field. This gives rise to the fermion masses (no term exists for neutrinos which are massless in the SM). Lepton terms include Yukawa constants which are the coupling constants of the Higgs interaction. Yukawa matrices Y are introduced for the quark terms which allow different generations to mix Y_{ij}.

The last term is the Higgs term written in Equation 1.6:

$$\mathcal{L}_{Higgs} = (D_\mu H)^\dagger (D^\mu H)^\dagger + \mu^2 (H^\dagger H) - \lambda (H^\dagger H)^2 \tag{1.6}$$

This term includes the kinetic term and the potential term which makes the Higgs mechanism and symmetry breaking possible. The kinetic term, when the covariant derivative is fully written out, contains Higgs-Gauge Bosons interaction (W^2H, W^2H^2), and Higgs-Higgs interactions (H^3, H^4) and Gauge Boson mass terms (W^2). In the second term, $\mu^2 < 0$, which leads to a non-zero vacuum expectation value for the lowest energy state of the Higgs field ($H = \pm v = \pm\sqrt{-\mu^2/\lambda}$). The actual vacuum state of the field is equal to either $+v$ or $-v$, the choice of which state it is in is known as the spontaneous symmetry breaking which gives rise to the Higgs and Gauge Boson masses. Here v is known as the vacuum expectation value or *vev*. The Higgs mass can be written $m_H = \sqrt{2\lambda}v$ and the mass of the gauge bosons can be written $m_B \propto gv$. The fermion masses can also be written in terms of this non-zero *vev*: $m_f \propto gv$. In both these expressions g is an interaction coupling constant that is distinct for each particle and gauge boson. These gs can be written in terms of other constants since the SM has more degrees of freedom than constraints leading to degeneracies.

1.2 Proton-Proton Collisions

The SM is a robust theory but must be verified through experiment. It is also an incomplete theory of the universe and must therefore be extended and modified as new data is taken. A potent probe of the Standard Model is through collider experiments. Collider experiments have primarily two categories in which they can probe the Standard Model: Find a new particle as a resonance or final state radiation through converting collision energy into

mass, or by measuring the interaction of particles to find if a new feature exists. The latter could be through indirect detection of a particle through its effect on known particles or particle properties.

Proton-proton collisions at high-energy provide a great vehicle to cover both categories of probing the SM. The proton and its substructure allows for probing various quark and gluon initial state interactions whilst the high-energy allows for coverage of a large range of SM and BSM phenomenology. Functionally, protons have the advantage over electrons of not losing large amounts of energy due to radiation during the acceleration process and can therefore be used in circular colliders that have much higher CM energies further extending the reach of its measurements. The high instantaneous luminosities that these accelerators can achieve leads to a large output of data that when combined can reach a level of precision necessary to test the SM.

In order to produce massive particles, the energy available in the CM frame must be greater than the sum of masses of the particles being produced. The CM energy of a collision is calculated through a Lorentz invariant quantity "s" from the total energy and momentum of the two initial state particles. This quantity in its most general form is written out in Equation 1.7:

$$s = \left(\sum_{i=1}^{2} E_i\right)^2 - \left(\sum_{i=1}^{2} \mathbf{p}_i\right)^2 \tag{1.7}$$

This Lorentz invariant has units of [Energy]^2, thus the CM Energy is calculated through taking the square-root of this quantity $\sqrt{s}$.

1.2.1 Cross-Sections and Luminosity

The cross-section of a particular physics process dictates the rates at which two particles interact and depends on the properties of the particles, properties of their interaction with one another, and the interaction kinematics. Qualitatively a cross-section can be described as the rate of the number of interactions per unit time per target particle all divided by the incident flux. The rate of the number of interactions is described by the SM and calculated through Quantum Field Theory calculations (QFT). In the context of particle-particle collisions the cross-section can further be simplified to the following expression in Equation 1.8:

$$\sigma = \frac{N}{\int L dt} \tag{1.8}$$

where N is the number of events that match the physics process being measured and $\int L dt$ is the total integrated luminosity. L, the instantaneous luminosity is the rate of an event is defined by Equation 1.9:

$$L = f_{coll} \frac{n_1 n_2}{4\pi \sigma_x \sigma_y} \tag{1.9}$$

f_{coll} is the frequency of collisions, n_1 and n_2 are the number of particles in the colliding bunches, and σ_x and σ_y are the root-mean-squared (rms) horizontal and vertical beam sizes.

Protons are not point-like particles and contain a complex sub-structure of partons that make up the proton (quarks and gluons) all interacting with one another. When two protons collide at high center-of-mass energy, it is not the protons that interact but the partons inside the protons that interact. This makes the modeling of cross-sections more complicated due to the distinct way in which one particle interacts with another. The cross-section determination of a gluon in one proton colliding with a gluon in the other proton is different from the cross-section determination of a gluon interacting with a quark (or any combination of gluons, quarks and anti-quarks). Thus any cross-section calculated from proton-proton collisions must utilize a complete picture of which partons took place in the interaction. Taking this into account, the cross-section of two partons from the colliding protons interacting with one another can be written out as in Equation 1.10:

$$\sigma_{pp\to X} = \sum_{a,b} \int dx_1 dx_2 f_a^{(p_1)}(x_1,\mu) f_b^{(p_2)}(x_2,\mu) \hat{\sigma}_{(q/g)_a(q/g)_b\to X}(x_1 x_2 s,\mu) \tag{1.10}$$

where the following variables correspond to:

- $\sigma_{pp\to X}$ is the total cross-section of the process of two protons colliding generating a final state "X"
- $f_i^{(H_j)}(x_i,\mu)$ is the probability of finding parton i inside hadron j with x_i fraction of the momentum (also known as a parton distribution function)
- $\hat{\sigma}$ is the partonic cross-section calculated for the interaction the two partons that actually interacted with one another in the collision
- μ is the energy scale of the interaction and is often chosen to be $\sqrt{s}$

The summary of several Standard Model cross-section measurements performed by the ATLAS collaboration for $\sqrt{s}$ = 7, 8, 13 TeV is shown in Figure 1.2.

1.2.2 Parton Distribution Functions

A full description of proton-proton collisions must include the knowledge of which partons inside each proton actually collided to create the final state. This was explicitly included in the calculation through the inclusion of a factor for each proton known as the parton distribution function. Parton distribution functions (PDFs) describe the momentum distributions of quarks and gluons inside nucleons. At low interaction energies, a proton can be treated as a point-like particle. At high enough interaction energies, the sub-structure of the proton is exposed and the quarks and gluons that make up a proton are now the actual particles that interact in a collision. Moreover, since gluons can be generated inside the proton that

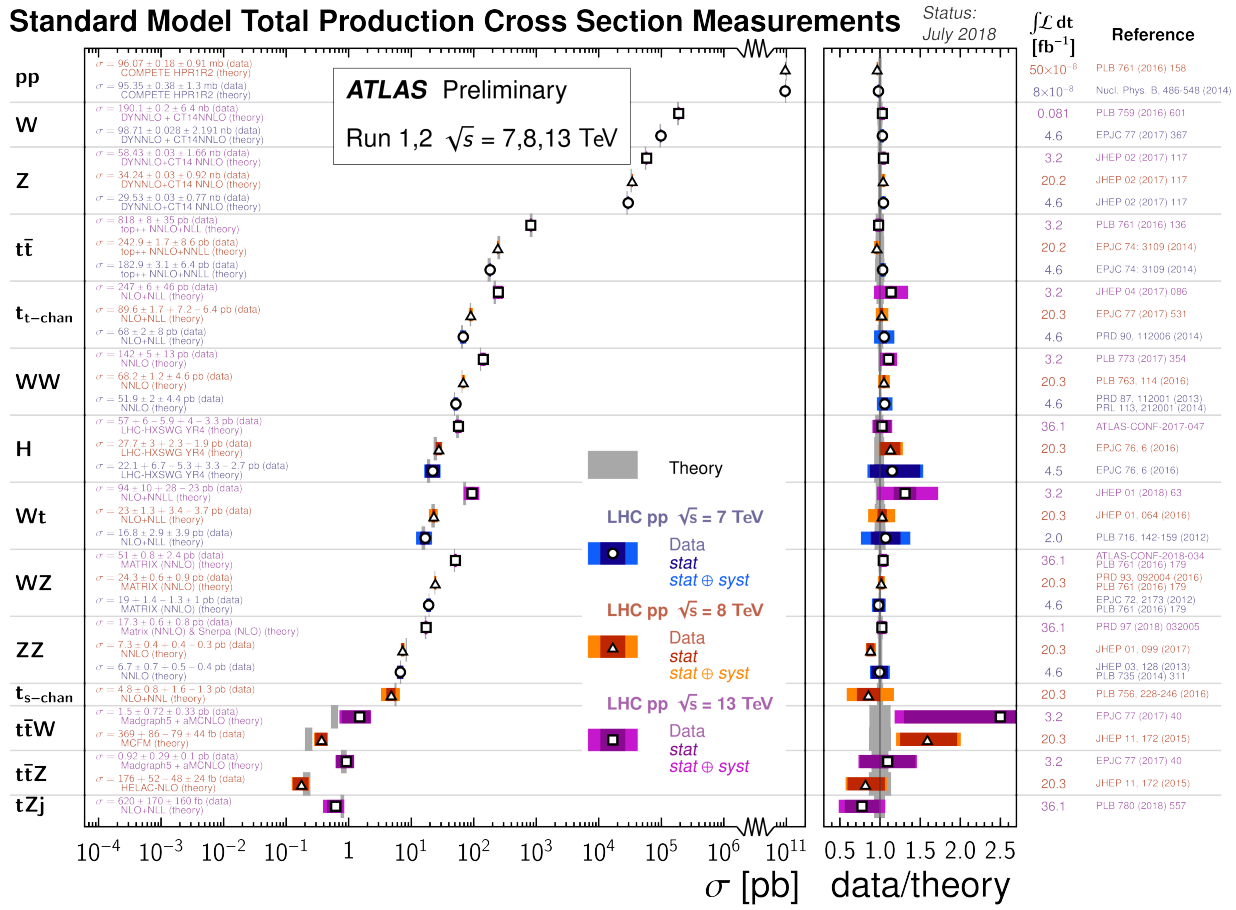


Figure 1.2: Summary of several Standard Model total production cross-section measurements, corrected for leptonic branching fractions, compared to the corresponding theoretical expectations. All theoretical expectations were calculated at NLO (Next-to-leading-Order) or higher. The dark-color error bar represents the statistical uncertainty. The lighter-color error bar represents the full uncertainty, including systematic uncertainties and luminosity uncertainties. The data/theory ratio, luminosity used and reference for each measurement are also shown. Uncertainties for the theoretical predictions are quoted from the original ATLAS papers.[3]

convert to quark–anti-quark pairs and back to a gluon, the complete make-up of the proton is dynamic.

Through a wide range of experimental data from Deep Inelastic Scattering experiments (DIS) it has been revealed that the particle contribution of a proton is three valence quarks (uud), a sea of virtual quarks and anti-quarks (from gluon splitting), and a sea of virtual gluons (radiated by valence quarks). The PDFs of various particles inside the proton are shown in Figure 1.3.

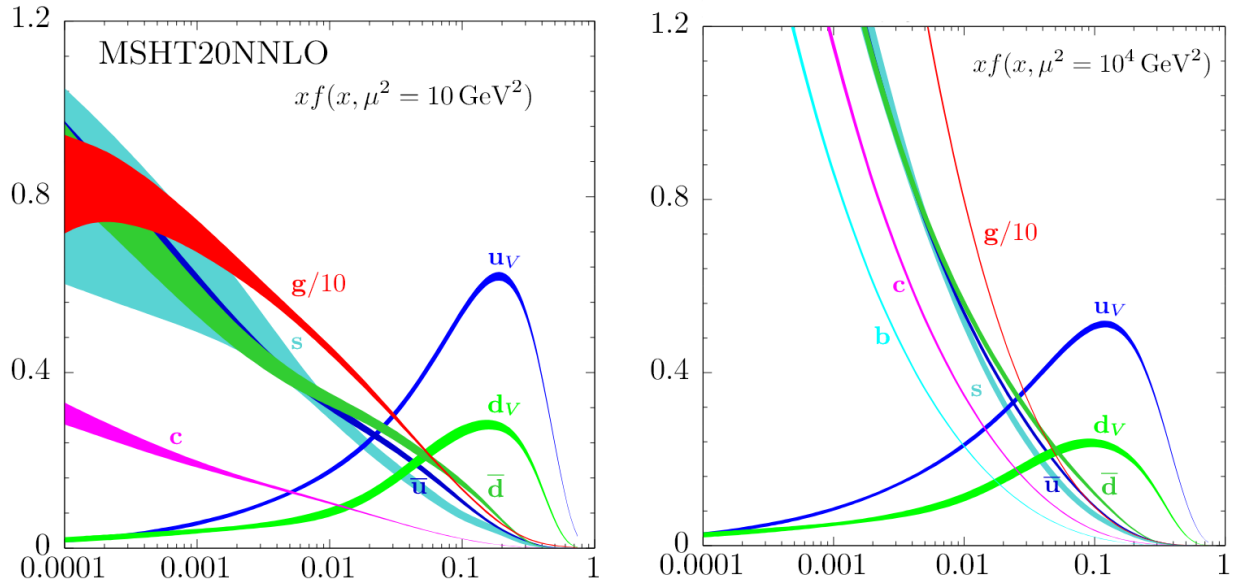


Figure 1.3: Parton distribution functions of the proton at $\mu^2 = 100\,\mathrm{GeV}$ and $\mu^2 = 10\,000\,\mathrm{GeV}$. The bands are the parton's momentum fraction times the unpolarized parton distributions f(x) obtained by the NNLO MSHT20 global analysis.[4]

These experiments also measured the gluon contribution to the proton's momentum to be about 50%. PDFs must obey momentum and energy conservation like all processes in physics as described by Equation 1.11:

$$\int_0^1 dx\; x \left(\sum_{i=1}^{n_f} \left[q_i(x,Q^2) + \overline{q}_i(x,Q^2) \right] + g(x,Q^2) \right) = 1 \tag{1.11}$$

where the sum over all quark and gluon PDFs must equal the full momentum of the proton, in this case normalized to 1.

In the high energy pp collisions at the Large Hadron Collider (LHC), protons collide at a high enough energy and rate that the constituents of the protons colliding can be considered static during the time of interaction. Measuring physics observables in the final state provide

a probe into properties of the initial state particles. It is due to this that measuring a physics process which contains either a charm or strange quark in the initial state can provide a probe into their properties and how they contribute to the set of proton PDFs.

1.3 Vector Boson + Charm Meson Measurements

The following Chapters describe the two V boson plus charm meson measurements carried out using $140\,\text{fb}^{-1}$ of $\sqrt{s} = 13\,\text{TeV}$ proton–proton (pp) collision data recorded by the ATLAS detector at the LHC. These measurements contribute to the fitting of proton PDFs, modeling of V+HF modeling and provides a great test and validation of pQCD.

1.3.1 W Boson + $D^{(*)}$ Meson Measurement

Currently, only limited information is available about the PDF of strange quarks in the proton. The sea distributions for the three light quarks, up, down and strange, might be equal due to flavor SU(3) symmetry; alternatively the strange quark distribution might be suppressed due to its larger mass. Current knowledge of the strange PDF comes largely from measurements of deep-inelastic lepton–proton scattering [5, 6] and charged-current neutrino scattering [7, 8, 9, 10, 11], and from vector-boson measurements at the LHC [12, 13, 14, 15]. However, constraints on the strange quark and antiquark PDFs are much weaker than those on the up and down sea quarks and antiquarks [16].

In pQCD, the production of a W boson in association with a single charm quark occurs through the scattering of a gluon and a down-type quark, i.e. down, strange or bottom, at leading order (LO), as shown in Figure 1.4. The relative contributions to the cross-section of $W+c$ production from each of the three different quarks depends on their PDFs and on the values of the three relevant terms from the Cabbibo–Kobayashi–Maskawa (CKM) mixing matrix [17, 18]: V_{cd}, V_{cs}, and V_{cb}. At the LHC, the process $gs \to W^-c$ and its charge conjugate are dominant, while the process $gd \to W^-c$ ($g\overline{d} \to W^+\overline{c}$) contributes only ∼10% (∼5%) to the W^-c ($W^+\overline{c}$) rate. The difference between the d and $\overline{d}$ contributions can be attributed to the presence of valence d-quarks [19]. The contribution from b-quark-initiated processes is negligible. The largest next-to-leading-order (NLO) contributions are the one-gluon-loop corrections to $gs \to W^-c$ ($gs \to W^+\overline{c}$); however, various other partonic initial states such as qq', gg and sq or $\overline{s}q$ are also present.

The idea of using $W+c$ events to measure the strange PDF was first proposed in Refs. [20, 21] and their production was first observed at the Tevatron [22]. At the LHC, it has been measured both by ATLAS and CMS using data taken at $\sqrt{s} = 7\,\text{TeV}$ [23, 24, 25] and by CMS using data taken at $\sqrt{s} = 8\,\text{TeV}$ and $13\,\text{TeV}$ [26, 27]. Measurements of $W+c$ production in the forward region at $\sqrt{s} = 7\,\text{TeV}$ and $8\,\text{TeV}$ have also been performed by the LHCb Collaboration [28]. In these measurements, the charm quark or antiquark is tagged either by the presence of a jet of particles containing a secondary vertex or a semileptonic decay

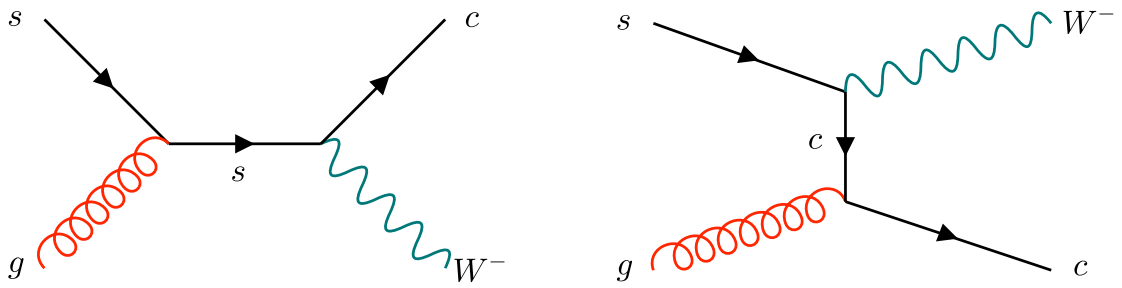


Figure 1.4: The leading-order diagrams for $W^- + c$ production

to a muon, or by explicit reconstruction of a D^+ or D^{*+} meson or its charge conjugate, collectively written as $D^{(*)}$.

This analysis is a measurement of W boson production in association with a $D^{(*)}$ meson using $140\,\text{fb}^{-1}$ of $\sqrt{s} = 13\,\text{TeV}$ proton–proton (pp) collision data recorded by the ATLAS detector at the LHC. Events in which the W boson decays to an electron or a muon (and the associated neutrino) are studied and the presence of the charm quark is detected through explicit charmed hadron reconstruction. The measurement does not require the presence of a reconstructed jet. The production of charmed hadrons is studied using the following decay mode (and its charge conjugate):

- $D^{*+} \to D^0\pi^+ \to (K^-\pi^+)\pi^+$.

The $D^+ \to K^-\pi^+\pi^+$ mode was also measured in this analysis but is not included in this dissertation. The signal W+D^{*+} events are extracted through a likelihood [29] fit to the reconstructed secondary-vertex mass difference $m(D^{*+}-D^0)$ for the D^{*+}. The main backgrounds are single-W-boson events that do not contain the requisite $D^{(*)}$ decays and $t\bar{t}$ events.

In W+c production, at LO the W boson and charm quark always have opposite-sign electric charges, i.e., either $W^+ + \bar{c}$ or $W^- + c$. For those processes where one of the initial-state partons is a strange or anti-strange quark, this charge correlation remains at NLO and next-to-next-to-leading order (NNLO) [19]. [1] However, many of the backgrounds (e.g. heavy-flavor pair production or b-hadron production from $t\bar{t}$ events) have equal rates for the production of leptons and $D^{(*)}$ with opposite-sign (OS) or same-sign (SS) charges. This is exploited in the analysis by extracting the signal as the difference of the numbers of OS and SS candidates, denoted by OS-SS, and extrapolating the background estimate from SS candidates. The $t\bar{t}$ background with events containing $W \to cs$ decays is not charge

[1] If there is a significant asymmetry between the charm and anti-charm PDFs, there would be a contribution from processes with charm quarks in the initial state, i.e. $dc \to W^-uc$ and $d\bar{c} \to W^-u\bar{c}$, but this is expected to be small [19].

symmetric and is measured in situ by categorizing the events according to whether b-tagged jets separated in phase space from the $D^{(*)}$ candidate are present.

The $W+D^{(*)}$ cross-sections, $\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W+D^{(*)})$, are measured in a fiducial region defined by requirements for W boson and $D^{(*)}$ meson selection. The requirements for W boson selection are a charged lepton, ℓ (e or μ), of transverse momentum $p_T^{\ell} > 30\,\text{GeV}$ and pseudorapidity $|\eta^{\ell}| < 2.5$. The requirements for $D^{(*)}$ meson selection are $p_T(D^{(*)}) > 8\,\text{GeV}$ and $|\eta(D^{(*)})| < 2.2$. The total fiducial cross-section is presented along with two differential cross-sections, in $p_T(D^{(*)})$ and $|\eta(\ell)|$. The measurements are performed separately for events with positively and negatively charged W bosons and the ratio $R_c^{\pm} \equiv \sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^+ + D^-)/\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^- + D^+)$ is also presented. These measurements are compared with QCD predictions obtained using state-of-the-art PDF sets [30, 13, 31, 4, 32, 33, 16, 34].

An important development in the theoretical study of $W+c$ production is the recent publication of the first NNLO calculation [35] of the process. This calculation includes an off-shell treatment of the W boson and is performed in a five-flavor scheme using the infrared- and collinear-safe flavored k_t algorithm [36] and neglecting c-quark finite-mass effects. Non-diagonal CKM matrix elements and the dominant NLO electroweak (EW) corrections are included. Scale uncertainties obtained using this calculation are below 2%, significantly smaller than PDF uncertainties for most PDF set choices. Such NNLO calculations will ultimately allow the incorporation of $W+c$ measurements into NNLO PDF fits. For $W+c$-jet measurements, comparisons with NNLO predictions require that cross-sections be unfolded to jet observables calculated in the flavored k_t scheme; such unfolded results are not currently available. Alternatively, in the case of $W+D^{(*)}$ measurements, the charm fragmentation function could in the future be incorporated into theory predictions using methods pioneered in Ref. [37].

The measurements presented here are compared with QCD calculations with NLO plus parton shower accuracy. The baseline framework for these calculations and the QCD scale uncertainties associated with them is MadGraph5_aMC@NLO [38]. Theoretical uncertainties associated with the choice of matching scheme are assessed using the difference between predictions obtained with MadGraph5_aMC@NLO and those obtained with recent calculations [39] implemented in the Powhel event generator [40].

1.3.2 Z Boson + $D^{(*)}$ Meson Measurement

Studies of Z boson production in association with a $D^{(*)}$ meson provide an important tests of perturbative quantum chromodynamics (pQCD) calculations. This signature is also a major background for a variety of physics processes, including standard model Higgs boson production in association with a Z boson. It is also helpful for searches of new physics signatures with leptons and heavy flavor final states. Therefore a precise measurement of the modeling of Z bosons plus heavy flavor will be useful for both theory and phenomenological measurements. This measurement is sensitive to the charm PDF where the charm quark in the initial state comes from the evolution of the gluon PDF (eg. $g \to c\bar{c}$) in the initial state.

It can also contribute to the PDF fits that measure how "intrinsic" the charm contribution is to the proton PDFs however this analysis is not sensitive to intrinsic charm. [41, 42]

In pQCD, the production of a Z boson in association with a single charm quark occurs through the scattering of a gluon and a charm quark, at leading order (LO), as shown in Figure 1.5. At the LHC the process $gc \to Z^0 c$ is dominant and the gluon-spitting process $q\bar{q} \to Z^0 c\bar{c}$ is much smaller. [43] The gluon-splitting process and contributions from b-quark-initiated processes are not measured directly in this analysis but will contribute to the signal. The b-quark-initiated processes contribute through $B \to D$ decays. The largest next-to-leading-order (NLO) contributions are the one-gluon-loop corrections to $gq \to Z^0 q$ and $q\bar{q} \to Z^0 g$ and are sub-dominant.

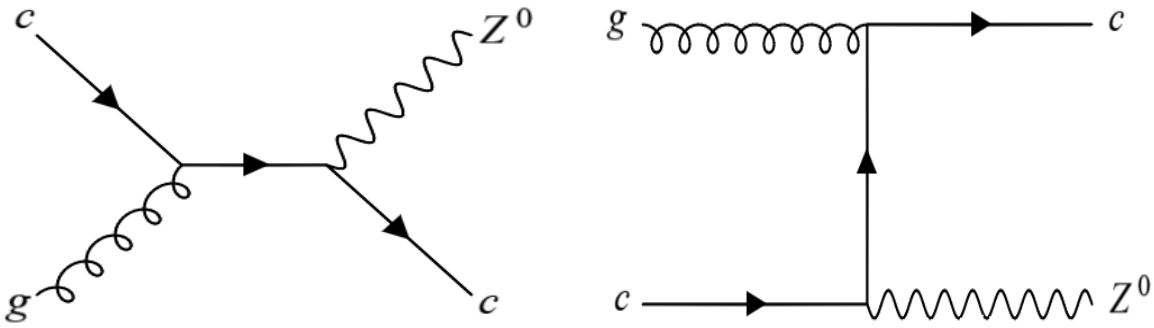


Figure 1.5: The leading-order diagrams for $Z^0 + c$ production

The idea of measuring the production of Z+c events was first proposed in Refs. [44, 45] with intrinsic charm as the main motivator. At the LHC, this measurement has been performed by CMS using data taken at $\sqrt{s} = 7\,\text{TeV}$ [46], at $\sqrt{s} = 8\,\text{TeV}$ [47] and with partial data at 13 TeV [48]. Measurements of Z+c production in the forward region at $\sqrt{s} = 13\,\text{TeV}$ have also been performed by the LHCb Collaboration [49]. In these measurements, the charm quark or antiquark is tagged either by the presence of a jet of particles containing a secondary vertex or a semileptonic decay to a muon, or by explicit reconstruction of a D^+ or D^{*+} meson or its charge conjugate, collectively written as $D^{(*)}$.

This analysis is a measurement of Z boson production in association with a $D^{(*)}$ meson using $140\,\text{fb}^{-1}$ of $\sqrt{s} = 13\,\text{TeV}$ proton–proton (pp) collision data recorded by the ATLAS detector at the LHC. Events in which the Z boson decays to a pair of electrons or muons are studied and the presence of the charm quark is detected through explicit charmed hadron reconstruction. The measurement does not require the presence of a reconstructed jet. The production of charmed hadrons is studied using the following decay modes (and their charge conjugates):

- $D^+ \to K^-\pi^+\pi^+$ and

- $D^{*+} \to D^0\pi^+ \to (K^-\pi^+)\pi^+$.

The signal $Z{+}D^{(*)}$ events are extracted through a profile likelihood [29] fit to the reconstructed secondary-vertex mass distribution for the D^+ and the mass difference $m(D^{*+}{-}D^0)$ for the D^{*+}. The main backgrounds are single-Z-boson events that do not contain the requisite $D^{(*)}$ decays, Diboson and $t\bar{t}$ events.

The $Z{+}D^{(*)}$ cross-sections, $\sigma_{\text{fid}}(Z{+}D^+)$, are measured in a fiducial region defined by requirements for Z boson and $D^{(*)}$ meson selection. The requirements for Z boson selection are the presence of oppositely-charged, same-sign leptons, $\ell\ell$ (ee or $\mu\mu$), of transverse momentum $p_T^\ell > 27\,\text{GeV}$ and pseudorapidity $|\eta^\ell| < 2.5$. The invariant mass of the Z boson must fall within $66 < m(Z) < 116$ GeV. The requirements for $D^{(*)}$ meson selection are $p_T(D^{(*)}) > 8\,\text{GeV}$ and $|\eta(D^{(*)})| < 2.2$. The total fiducial cross-section is presented along with $p_T(D^{(*)})$ differential cross-sections. These measurements are compared with QCD predictions obtained using state-of-the-art PDF sets [30, 13, 31, 4, 32, 33, 16, 34].

Chapter 2

The ATLAS Experiment

This chapter provides a description of the Large Hadron Collider and the ATLAS detector. Chapter 2.1 gives the description of the Large Hadron Collider, how protons are brought up to collision energy and physics motivations for its construction. Chapter 2.2 summarizes the ATLAS detector, including each of the detector elements that go into measuring and reconstructing the particles generated in the collision and the trigger and data acquisition system.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring-superconducting-hadron accelerator and collider housing four main experiments: ATLAS, CMS, ALICE, LHCb. The full layout of the CERN complex is show in Figure 2.1[50][12]. Various accelerator components are also shown that are used to bring the protons up to collision energy. The LHC was constructed in order to probe particle physics at high energies. The high luminosity and increased cross-sections enabled further high precision tests of QCD, electroweak interactions, and flavour physics. The main physics drivers were the search of the Standard Model Higgs boson, validation of the Supersymmetry (SUSY) model, search for extra dimensions, and new heavy gauge bosons W' and Z'.

The LHC was built to be a proton-proton collision machine operating at high energies. Protons are brought up to the correct center-of-mass energy (CM Energy) through a sequence of energy boosts that include multiple smaller accelerator sections. Magnets and radio frequency cavities are utilized to provide these boosts. Hydrogen ions are accelerated to 150 MeV in a linear accelerator (LINAC). The electrons are stripped off just before being injected into the Booster. The Booster accelerates protons to 1.4 GeV and sends them to proton synchrotron (PS) where the protons increase their energy to 25 GeV. The protons are further accelerated by super proton synchrotron (SPS) to 450 GeV, and then finally injected into LHC main ring. Here the proton follows the path of the beam line accelerating until its

energy reaches its desired CM Energy.

The LHC was operated with a 40 MHz proton bunch frequency, and hence the bunch crossing occurs every 25 ns. The peak Luminosity of collisions reached $2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ collecting a total of 140 fb^{-1} through 2015-2018 known as Run 2 of data taking at the LHC.

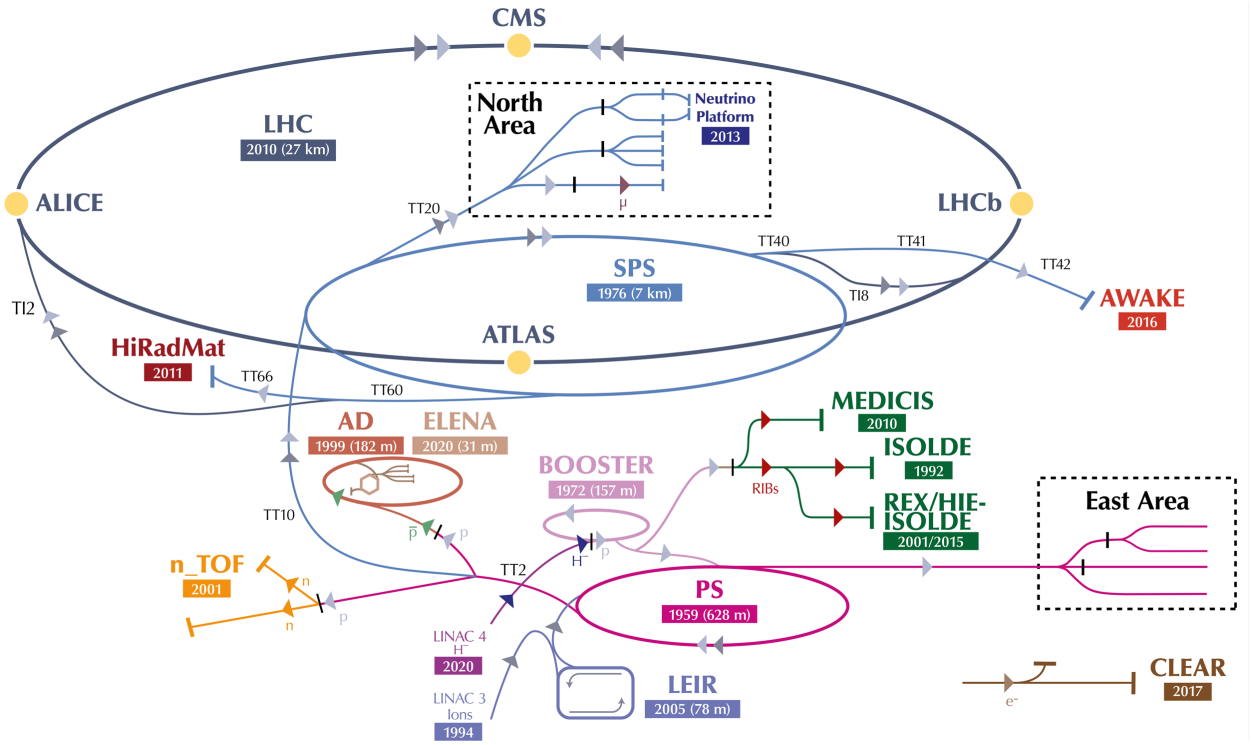


Figure 2.1: The CERN accelerator complex. Shown along the main beam line are the four main experiments of the LHC: ATLAS, CMS, ALICE, LHCb.[50][12]

2.2 The ATLAS Detector

2.2.1 Detector Overview

The ATLAS detector has a hermetic, forward-backward symmetric construction built to be a versatile probe of the SM. It was designed to be able to perform precision measurements and carry out the physics program mentioned in the previous section. A diagram of the detector is shown in Figure 2.2 [51]. Detector elements most important for this analysis were the pixel detector, semiconductor tracker and muon spectrometers explained in the following sections. I worked on the upgrade and validation of the new pixel detector readout chip as described in Appendix A.

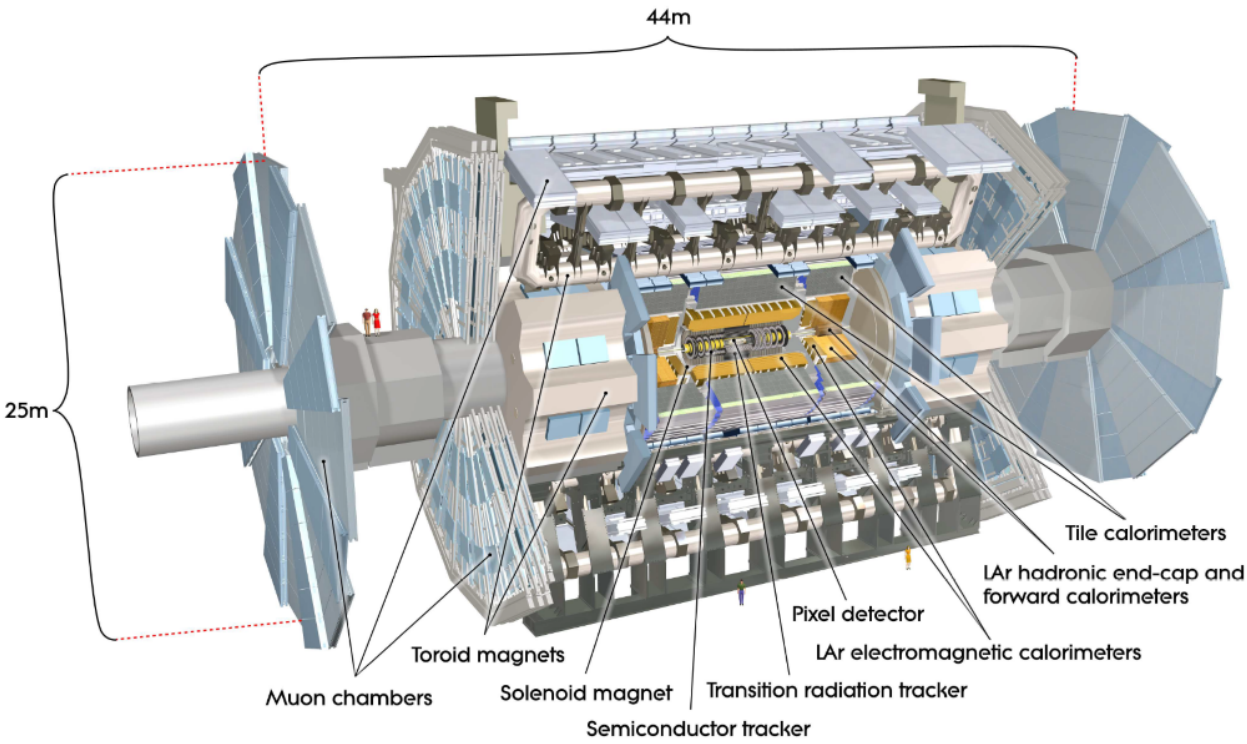


Figure 2.2: The ATLAS detector. Shown are all of the detector elements necessary to measure and reconstruct the particles produced from the *pp* collisions. [51]

The ATLAS Coordinate System

The coordinate system and nomenclature used to describe the ATLAS detector and the particles emerging from the p-p collisions are briefly summarised here. The nominal interaction point is defined as the origin of the coordinate system, while the beam direction defines the z-axis and the x-y plane is transverse to the beam direction. The positive x-axis is defined as pointing from the interaction point to the center of the LHC ring and the positive y-axis is defined as pointing upwards. The azimuthal angle ϕ is measured as usual around the beam axis, and the polar angle θ is the angle from the beam axis. The pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$ (in the case of massive objects such as jets, the rapidity $y = 1/2 \ln[(E + p_z)/(E - p_z)]$ is used). The transverse momentum p_T, the transverse energy E_T, and the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ are defined in the x-y plane unless stated otherwise. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

Overview of Detector Elements

The ATLAS Detector is made up of three major detecting components: the inner detector, the calorimeters, and the muon spectrometer. The inner detector is surrounded by a 2 T solenoidal magnet. Together they are used to measure charged particle trajectories and their transverse momentum. The inner detector consists of the Pixel detector, the Semiconductor Strips (SCT) detector and the Transition Radiation Tracker (TRT). Further out in radius are the electromagnetic (EM) and hadronic calorimeters. These calorimeters measure the energy of the particles passing through our detector; the EM calorimeter is optimized to measure electron and photon energy, the hadronic calorimeter measures hadronic particle energies. The final detector element at highest radius is the Muon Spectrometer. Surrounded by a toroidal magnet, the muon spectrometer measures the momentum of muons that traverse the full inner detector and calorimeters of the detector. Each detector element is explained in greater detail in the next few sections.

2.2.2 Inner Detector

Particle momentum and vertex measurements are achieved with a combination of discrete, high-resolution semiconductor pixel and strip detectors in the inner part of the tracking volume. The straw-tube tracking detectors allow for electron identification with its capability to generate and detect transition radiation. The schematic of the full inner detector is shown in Figure 2.3. The overall tracking resolution is $\sigma_{p_T}/p_T = 0.05\%\ pT \bigoplus 1\%$.

The Pixel Detector

The Pixel detector is made up of four barrel layers and 2x3 end-cap disks and contains a total 92M readout channels. The layer closest to the beam line is called the Insertable B-layer. The overall pixel detector envelope covers radius $31 < r < 242$ mm and longitudinal distance $0 < |z| < 3092$ mm.

The IBL is a barrel of semiconductor (pixel) sensors inserted into the ATLAS detector in 2013 and located closest to the beam line at radius 334 mm [53]. The IBL was designed to increase tracking robustness, tracking precision and withstand larger luminosities and radiation doses. Its proximity to the beam line and the use of small feature sizes in the pixel sensors result in a greater tracking, vertexing, and heavy physics tagging performance. The IBL is a cylinder made up of 14 pixel staves tilted by 14deg to ensure full ϕ coverage. Each stave contains 20 IBL pixel modules, each of which is made up of a silicon sensor bump bonded to one of four readout chips resulting in 224 total modules. Two sensor technologies used in this layer are double-chip planar sensors and 3D single-chip sensors. The planar sensor is a 200 μm thick n^+-in-n silicon sensor. The 3D sensor is a 230 μm thick produced with the double-side double type Columns process with n^+ columns from the front side and p^+ columns from the back side. Both sensors have a 250x50 μm^2 pixel size resulting in

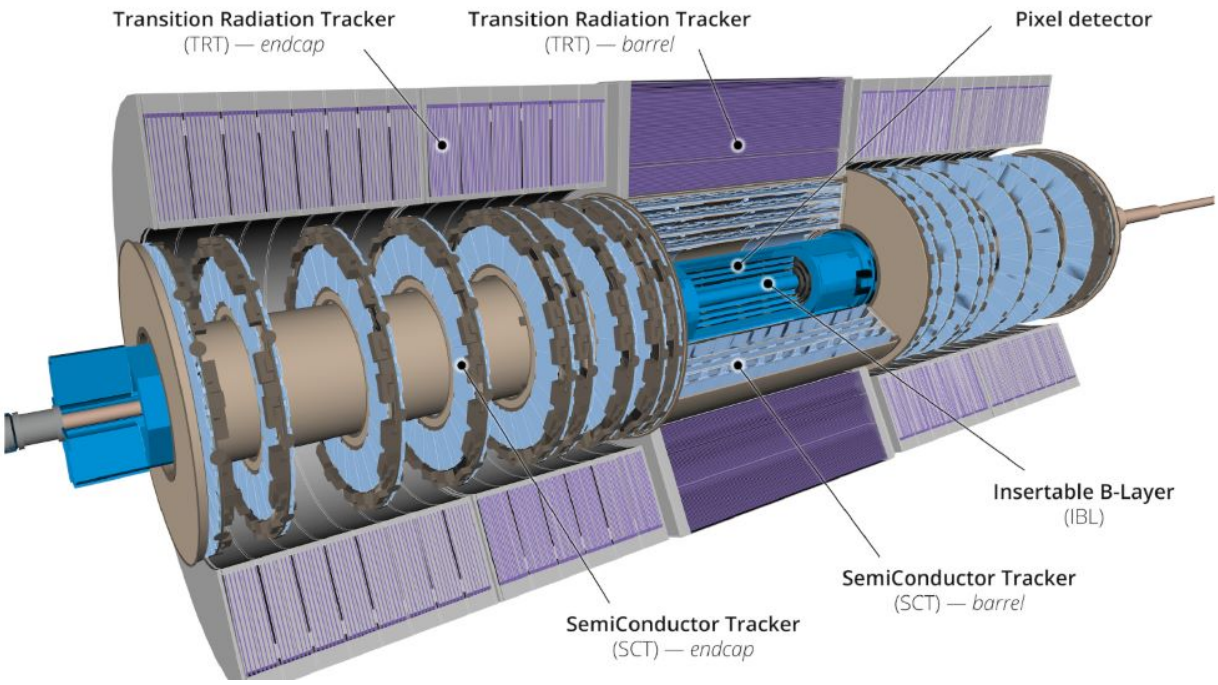


Figure 2.3: Schematic view of the inner detector [52].

6.02M pixels total in the IBL. The Front-End Readout Chip (FE-I4B) is used to read out the charge deposited into the sensors. It has 26880 pixel cells with a pixel size of 250x50 μm^2 distributed over 80 columns and 336 rows which are connected to the sensors by bump bonding each individual cell. The FE-I4B uses a 130 nm feature size bulk CMOS process.

The three pixel outer layers all have identical 400x50 μm^2 pixel sensors and all use the single-side planar sensor technology with a 250 μm pitch. The pixel array are distributed over 18 rows and 160 columns. The read out chip used in these layers (FE-I3) is the older iteration of the FE-I4B technology used in the IBL and is built on a 250 nm feature size bulk CMOS process.

The intrinsic accuracies in the barrel are 10 μm (in rϕ) and 115 μm (in z) and in the disks are 10 μm (in rϕ) and 115 μm (in r). The result of the inclusion of the IBL results in a greater than 50% increase in resolution of the transverse and longitudinal impact parameters of low pT tracks.

The Strips Detector

The Strips SCT detector is a silicon micro strip detector made up of 4 barrel layers and 2x9 endcap layers and contains 6.3M readout channels. The overall strips detector envelope covers radius $251 < r < 560$ mm and longitudinal distance $0 < |z| < 2797$ mm.

The sensors use a single-sided p^+-in-n silicon technology with AC-coupled readout strips. The sensor thickness is 285±15 μm, strip pitch of 80 μm with two 6 cm-long sensors daisy chained chosen for the barrel sensors and radial strips and a mean pitch of ∼80 μm for the trapezoidal end-cap sensors. The readout chips used are fabricated in radiation tolerant bi-CMOS DMILL technology.

The intrinsic accuracies per module in the barrel are 17 μm (rϕ) and 580 μm (z) and in the discs are 17 μm (rϕ) and 580 μm (r).

The Transition Radiation Tracker Detector

The transition radiation tracker detector comprises several layers of gas-filled straw tubes interleaved with transition radiation material and contains 350K readout channels. The overall TRT detector envelope covers radius $554 < r < 1106$ mm and longitudinal distance $0 < |z| < 2744$ mm. Polymide straw drift tubes of 4 mm diameter have 31 μm Tungsten wire anodes plated with 0.5-0.7 μm gold. Cathodes are operated at ∼1530 V to give the proper gain to the 70% Xe, 27% Co, 3% O gas mixture filling the tubes. Low energy transition radiation (TR) photons are absorbed in the Xe-based gas mixture, and yield much larger signal amplitudes than minimum-ionising charged particles. The distinction between TR and tracking signals is obtained on a straw-by-straw basis using separate low and high thresholds in the front-end electronics.

The 300K thin-walled proportional-mode drift tubes provide on average 30 (r,ϕ) points with 130 μm resolution for charged particle tracks with $|\eta| < 2$ and $p_T > 0.5$ GeV, contributing to the combined tracking system p_T resolution. The mean electron collection time of ∼40 ns was observed for the TRT.

2.2.3 Calorimeters

The EM calorimeter measures the energies and positions of charged and neutral electromagnetically interacting particles while the Hadronic calorimeter measures the strongly interacting particles. Both calorimeters are designed to fully stop particles produced in the collision to force them to deposit all of their energy and therefore measure the full energy they contained. The ATLAS calorimeters are sampling calorimeters; they consist of layers of "absorbing" high-density materials that stop incoming particles, interleaved with layers of "active" media that measure the particle energies. ATLAS uses two sampling calorimeter technologies: LAr for the electromagnetic calorimeters and all of the endcap and forward calorimeters, and scintillating Tiles for hadron calorimetry in the central region. The full schematic of the EM and Hadronic Calorimeters is shown in Figure 2.4.

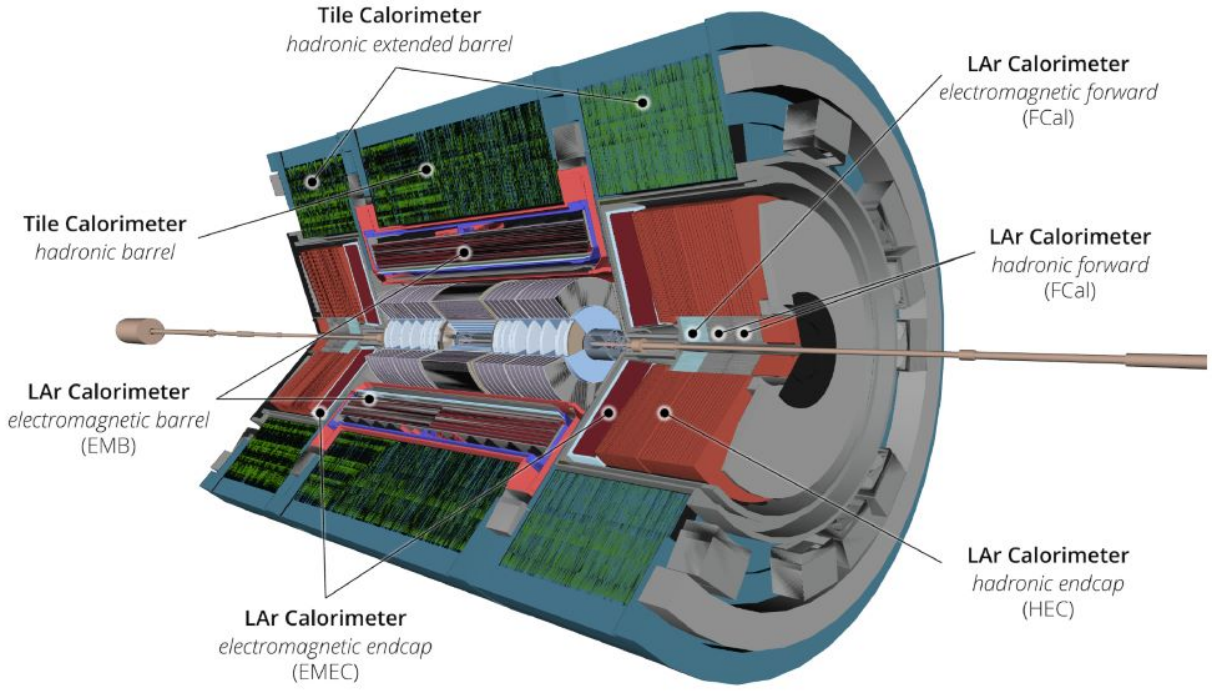


Figure 2.4: Schematic view of the Electromagnetic and Hadronic Calorimeters [52].

Electromagnetic Calorimeter

High granularity liquid-argon (LAr) electromagnetic sampling calorimeters cover the pseudorapidity range $|\eta| < 3.2$ and contains a total of ∼170K readout channels. The Liquid Argon (LAr) system consists of several subsystems, namely the LAr Electromagnetic Barrel Calorimeter (EMB), the LAr Electromagnetic Endcap Calorimeter (EMEC), the LAr Hadronic Endcap Calorimeter (HEC), and the LAr Forward Calorimeter (FCal). the LAr Calorimeter system measures the energy of electrons, photons, τ leptons, and jets (clusters of particles) as they are slowed by the dense calorimeter material. Each physics object contains distinct signatures in the calorimeter which contributes to the identification of which particle was present for a particular collision. Missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) is also calculated using the calorimeter by vectorally adding all the energy contributions deposited into the calorimeter and calculating the resulting p_T. A non-zero result of this vector-sum in p_T is defined as missing momentum and gives the $E_{\mathrm{T}}^{\mathrm{miss}}$ amount of the event.

The EM Calorimeter uses an accordion structure for the absorbers and the electrodes of the barrel and end-cap calorimeters. Where alternating layers of LAr and lead plate absorbers are stacked providing the measuring and stopping material to cascade incoming particles and fully absorb their energy. A module has three layers or layers in depth (front, middle and back as viewed from the interaction point). The front layer is read out at the

low-radius side of the electrode, whereas the middle and back layers are read out at the high-radius side of the electrode. The readout granularity is different for all three layers of modules being more granular the closer to the beamline and lowering its granularity further out. The inner layer has a granularity of $\Delta\phi \times \Delta\eta = 0.0245$ x 0.0031 while the outermost layer has a granularity of 0.0245 x 0.05. The EMEC end-cap calorimeter consists of two co-axial wheels. Each end-cap wheel is further divided into eight wedge-shaped modules. The overall energy resolution is $\sigma_E/E = 10\%/\sqrt{E} \bigoplus 0.7\%$.

Hadronic Calorimeter

The hadronic calorimetry in the range $|\eta| < 1.7$ is provided by a scintillator-tile calorimeter, which is separated into a large barrel and two smaller extended barrel cylinders, one on either side of the central barrel and contain a total of ∼15K read out channels. In the end-caps ($|\eta| > 1.5$), LAr technology is also used for the hadronic calorimeters, matching the outer $|\eta|$ limits of end-cap electromagnetic calorimeters. The LAr forward calorimeters provide both electromagnetic and hadronic energy measurements, and extend the pseudorapidity coverage to $|\eta| = 4.9$. The barrel hadronic calorimeter is a sampling calorimeter using steel as the absorber and scintillator as the active medium. It is located in the region, $|\eta| <$ 1.7, outside the liquid argon electromagnetic calorimeter. The scintillating tiles are paired with wavelength-shifting fibre readout on the tile edges that feed into photomultiplier tubes to measure particle energies. The overall energy resolution is $\sigma_E/E = 50\%/\sqrt{E} \bigoplus 3\%$ in the barrel region and $\sigma_E/E = 100\%/\sqrt{E} \bigoplus 10\%$ in the end-cap region.

2.2.4 Muon Spectrometer

The air-core toroid system, with a long barrel and two inserted end-cap magnets, generates strong bending power in a large volume within a light and open structure. The full schematic of the muon spectrometer is shown in Figure 2.5. The Muon Spectrometer is made up of multiple sub-detectors: Monitored Drift Tubes (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin-Gap Chambers (TGC). All detectors combine to provide a measurement coverage of $|\eta| < 2.7$. Multiple-scattering effects are thereby minimised, and excellent muon momentum resolution is achieved with three layers of high precision tracking chambers. The muon instrumentation includes, as a key component, trigger chambers with timing resolution of the order of 1.5-4 ns. The overall muon momentum resolution is $\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV.

2.2.5 Triggers and Data Acquisition

The Trigger and Data Acquisition system (TDAQ) is designed to reduce the readout of the detector from saving data every collision at a rate of 40 MHz to saving and reading out the data only for particularly interesting events. This decision is based on a Trigger Menu

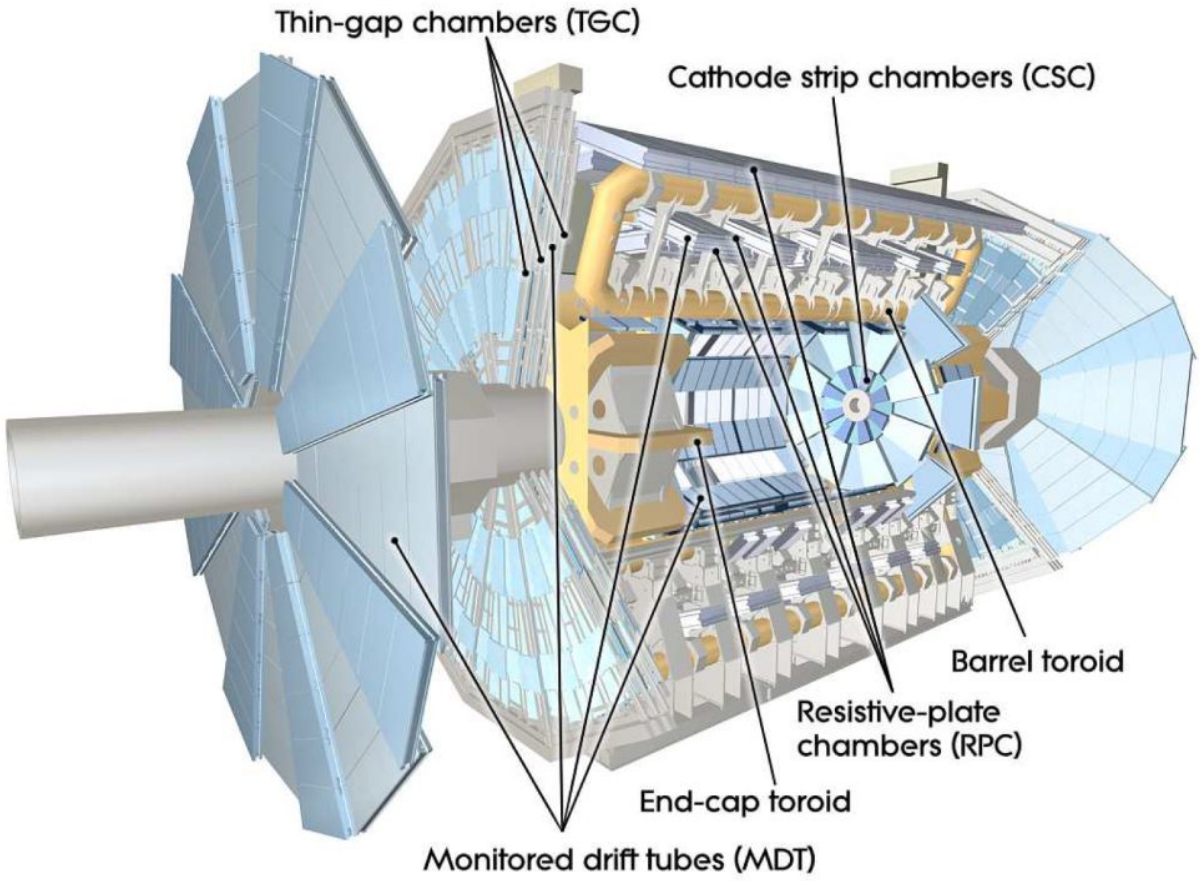


Figure 2.5: Schematic view of the Muon Spectrometer [52].

that is designed to save events that contain either certain physics objects (leptons, photons, τ leptons, etc.), contain objects which pass kinematic criteria (p_T, E, etc.) or meet event-level criteria ($E_{\mathrm{T}}^{\mathrm{miss}}$, jet multiplicity, etc.). The full TDAQ system is shown in Figure 2.6.

The triggers are split into two levels: the hardware-only based trigger (L1) and the software based high level trigger (HLT). The L1 trigger decision is formed by the Central Trigger Processor (CTP) which receives input from the L1 Calorimeter Trigger, the L1 Muon Trigger and the L1 Topological trigger amongst other detector systems. The L1 trigger accepts events at a rate up to the maximum detector readout rate of 100 kHz within a latency of 2.5 μs. Once the L1 trigger fires, the data from these detectors are sent out to the readout system where it buffers the data waiting for a decision to send to data storage from the HLT system. The HLT is a software based trigger which utilizes CPUs to first implement dedicated fast trigger algorithms and subsequently more complicated CPU-intensive algorithms for the

final data selection. HLT readout rates averaged 1.2 kHZ (storage rate ∼1.2 GB/s) during the Run 2 data taking campaign. Once an HLT decision is taken, the data is sent to an offline CERN computing center.

The DAQ system reads out the data from the various sub-detectors, serializes it, and sends it to a off-site computing cluster when a trigger has been executed.

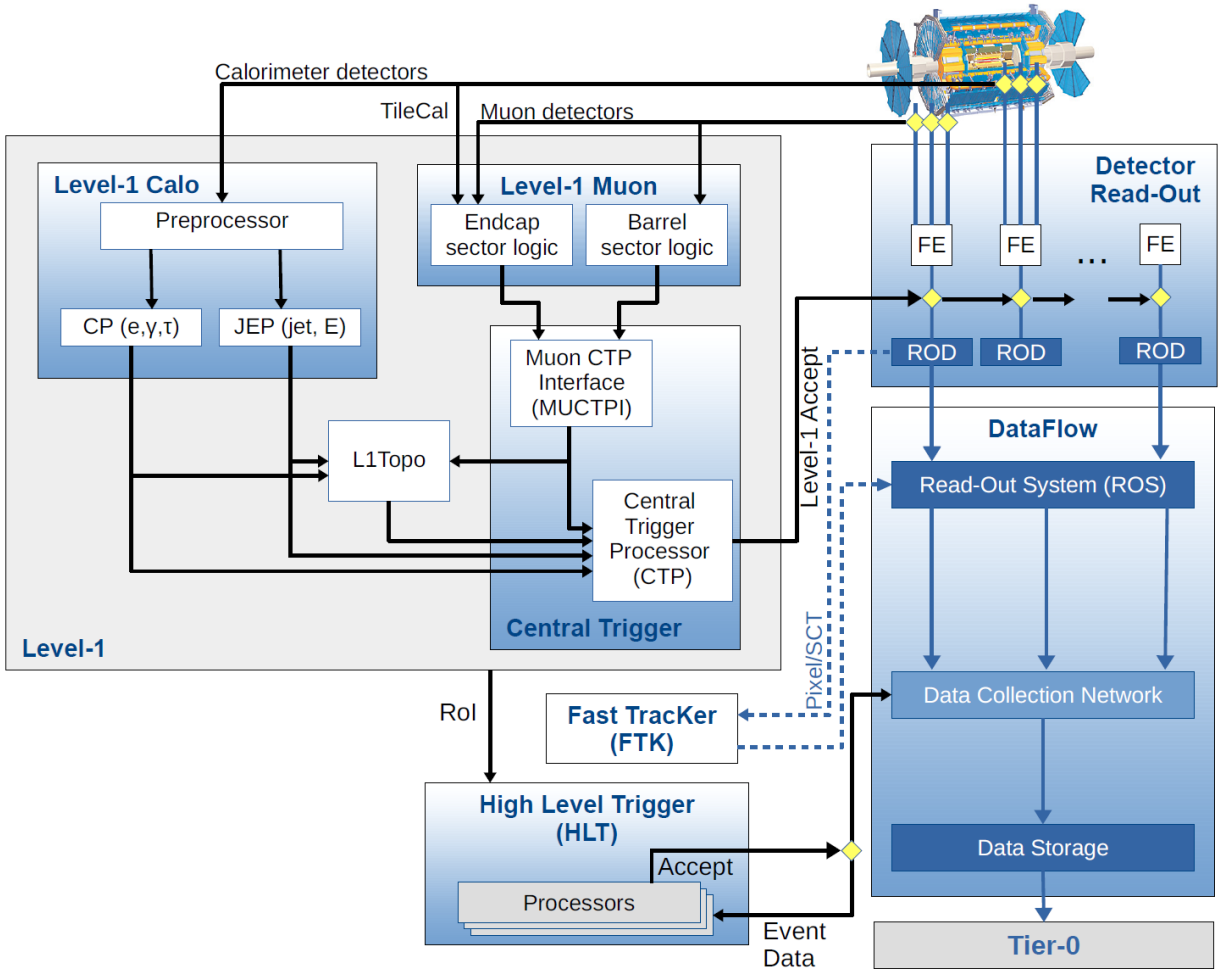


Figure 2.6: The ATLAS TDAQ system in Run 2 showing the components relevant for triggering and the detector readout and data flow. [54].

Chapter 3

Data and Monte Carlo Samples

This chapter describes the data used for both measurements and Monte Carlo samples generated to provide the physics modeling we compare against. Chapter 3.1 gives the description of the data collected by the ATLAS collaboration along with the triggers used to readout this data. Chapter 3.2 give the description of the Monte Carlo simulation and generation configurations used to model the signal and background processes.

3.1 Data

The data used for these analyses were taken by the ATLAS detector between 2015-2018 known as Run 2 of the LHC. The collision energy of the pp collisions was $\sqrt{s} =$ 13 TeV throughout the data taking period with pileup conditions ranging $\langle\mu\rangle \sim 20-60$. Pileup is the mean number of actual protons that collide each bunch crossing and is shown in Figure 3.1 for Run 2.

The delivered luminosity for Run 2 was 140 fb^{-1} after data quality requirements were applied [56] which ensured full detector functionality throughout the data-taking period. The total delivered luminosity for all years of Run 1 and Run 2 data taking at ATLAS is shown in Figure 3.2. The uncertainty of the Run 2 delivered integrated luminosity was measured to be 0.83% [57], obtained using the LUCID-2 detector [58], inner detector and calorimeters. The absolute luminosity scale was determined using van der Meer scans and extrapolated to physics data-taking using complementary measurements from several detectors.

Events used for the analysis presented here were recorded and read out using either single-electron or single-muon triggers. The minimum p_T threshold ranged during data-taking from 24 GeV to 26 GeV for electrons and from 20 GeV to 26 GeV for muons. Low p_T threshold triggers ($< 60\,\text{GeV}$ for electrons and $< 50\,\text{GeV}$ for muons) include isolation requirements to distinguish them from pileup tracks. For electrons, the requirement is $p_T^{\text{iso}}(\Delta R_{\text{var}} < 0.2)/p_T < 0.10$, where $p_T^{\text{iso}}(\Delta R_{\text{var}} < 0.2)$ is the scalar sum of transverse momenta of tracks within a variable-size cone, ΔR_{var}, around the electron. The cone size has a maximum value of 0.2

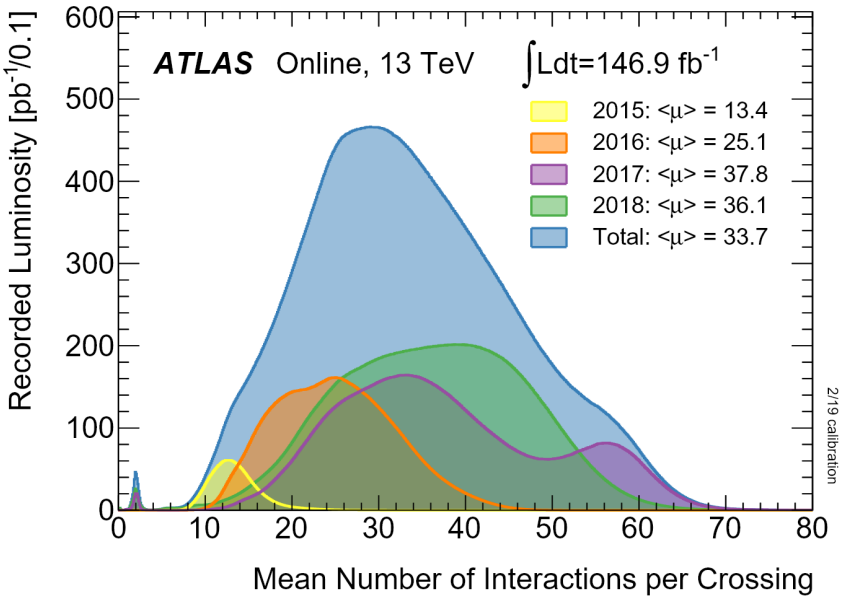


Figure 3.1: Luminosity-weighted distribution of the mean number of interactions per crossing for the 2015-2018 pp collision data at $\sqrt{s} =$ 13 TeV. [55].

and decreases as a function of electron's p_T as $10GeV/p_T[GeV]$[59]. The muon isolation criterion is constructed by summing the p_T of ID tracks with $p_T^{\text{track}} > 1GeV$ around the muon candidate satisfying $\Delta z < 6mm$, with Δz being the distance of the track from the primary vertex in the z-direction. Higher pileup events in 2017 required this selection to be tightened to $\Delta z < 2mm$, which allowed the loosening of the isolation criterion for data-taking in 2018. The muon isolation cut in 2017-18 is defined as $p_T^{\text{iso}}(\Delta z)/p_T < 0.07$, where $p_T^{\text{iso}}(\Delta z)$ is the scalar sum of transverse momenta of additional nearby tracks [60]. Triggers with higher p_T thresholds of 60 GeV and 140 GeV for electrons and 50 GeV for muons are added to increase the selection efficiency.

3.2 Simulated Monte Carlo Samples

Monte Carlo (MC) simulations are used to model the signal and all backgrounds except multijet. Samples produced with various MC generators are processed using a full detector simulation [61] based on GEANT [62] and then reconstructed using the same algorithms as the data. Pileup is modeled by overlaying each simulated hard-scattering event with inelastic pp events generated with Pythia8 (8.186) [63] using the `NNPDF2.3lo` set of PDFs [64] and a set of tuned parameters called the A3 tune [65]. The MC events are weighted to reproduce the distribution of the average number of interactions per bunch crossing ($\langle\mu\rangle$) observed

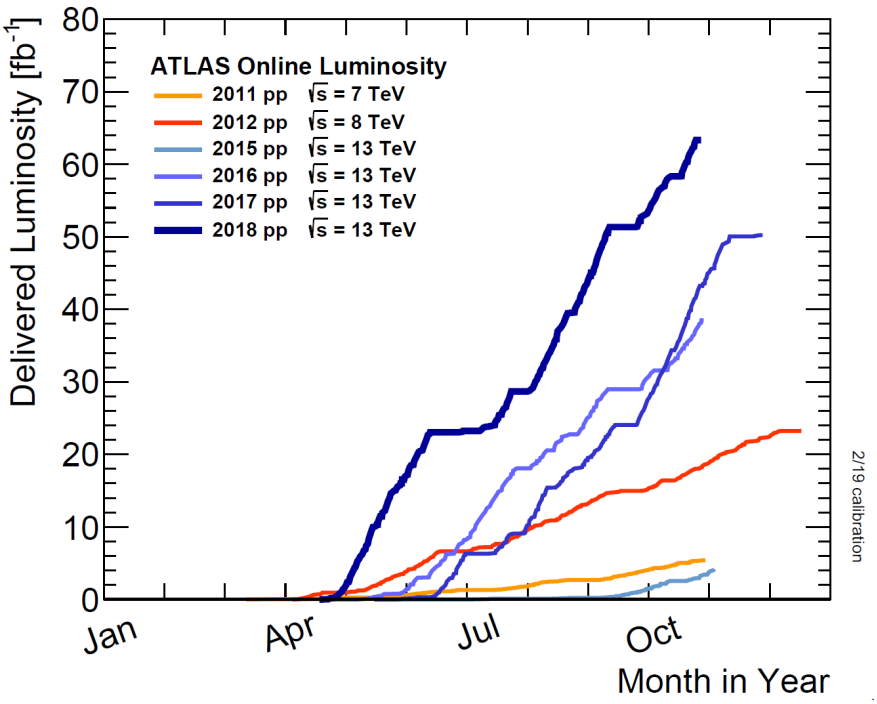


Figure 3.2: Delivered Luminosity of the ATLAS Detector for each year of Run 1 and Run 2 data taking campaigns. [52].

in the data, scaled up by a factor of 1.03 ± 0.04 to improve agreement between data and simulation in the visible inelastic pp cross-section [66]. A re-weighting procedure is applied to all MC samples to correct the charmed hadron production fractions to the world-average values [67, 68]. The change in the individual charmed meson production fractions is as large as 20%, depending on the MC configuration. An overview of all signal and background processes and the generators used to model them is given in Table 3.1 for the $W+D^{(*)}$ analysis and Table 3.2 for the $Z+D^{(*)}$ analysis. Further information about the relevant generators configurations is provided below. Processes with more than one jet, known as multi-leg processes, can have different numbers of jets in each event. To improve the accuracy of calculations, samples with different jet multiplicities are often merged. In such multi-leg samples, the QCD accuracy for each jet multiplicity is specified in the table.

3.2.1 Background $V+$jets samples

Three generator configurations are used to model inclusive vector boson (W or Z) plus jet production. These samples are used to estimate the $V+D^{(*)}$ backgrounds and the corresponding experimental and theory systematic uncertainties.

Sherpa: The nominal MC generator used for this analysis is SHERPA [2.2.11] [69]. NLO-

Table 3.1: The generator configurations used to simulate the signal and background processes for the $W+D^{(*)}$ analysis. The acronyms ME, PS and UE stand for matrix element, parton shower and underlying event, respectively. The column "HF decay" specifies which software package is used to model the heavy-flavor decays of bottom and charmed hadrons. For multi-leg samples where different jet multiplicities are merged, the QCD accuracy for each jet multiplicity is specified.

Process	ME generator	QCD accuracy	ME PDF	PS generator	UE tune	HF decay
W+jets (background modeling)						
W+jets	Sherpa [2.2.11]	0–2j@NLO+3–5j@LO	NNPDF3.0nnlo	Sherpa	Default	Sherpa
W+jets	MadGraph5_aMC@NLO (CKKW-L)	0–4j@LO	NNPDF3.0nlo	Pythia8	A14	EvtGen
W+jets	MadGraph5_aMC@NLO (FxFx)	0–3j@NLO	NNPDF3.1nnlo_luxqed	Pythia8	A14	EvtGen
$W+D^{(*)}$ (signal modeling and theory predictions)						
$W+D^{(*)}$	Sherpa [2.2.11]	0–1j@NLO+2j@LO	NNPDF3.0nnlo	Sherpa	Default	EvtGen
$W+D^{(*)}$	MadGraph5_aMC@NLO (NLO)	NLO	NNPDF3.0nnlo	Pythia8	A14	EvtGen
$W+D^{(*)}$	MadGraph5_aMC@NLO (FxFx)	0–3j@NLO	NNPDF3.1nnlo_luxqed	Pythia8	A14	EvtGen
Backgrounds						
Z+jets	Sherpa [2.2.11]	0–2j@NLO+3–5j@LO	NNPDF3.0nnlo	Sherpa	Default	Sherpa
$t\bar{t}$	PowhegBox [v2]	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
Single-t, Wt	PowhegBox [v2]	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
Single-t, t-channel	PowhegBox [v2]	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
Single-t, s-channel	PowhegBox [v2]	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
$t\bar{t}V$	MadGraph5_aMC@NLO	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
Diboson fully leptonic	Sherpa [2.2.2]	0–1j@NLO+2–3j@LO	NNPDF3.0nnlo	Sherpa	Default	Sherpa
Diboson hadronic	Sherpa [2.2.1]	0–1j@NLO+2–3j@LO	NNPDF3.0nnlo	Sherpa	Default	Sherpa

accurate matrix elements (ME) for up to two partons, and LO-accurate matrix elements for between three and five partons, are calculated in the five-flavor scheme using the Comix [70] and OpenLoops [71, 72, 73] libraries. The b- and c-quarks are treated as massless at matrix-element level and massive in the parton shower. The Hessian NNPDF3.0nnlo PDF set [74] is used. The default Sherpa parton shower [75] based on Catani–Seymour dipole factorization and the cluster hadronization model [76] is used. The samples are generated using a dedicated set of tuned parameters developed by the Sherpa authors and use the NNPDF3.0nnlo set. The NLO matrix elements for a given jet multiplicity are matched to the parton shower (PS) using a color-exact variant of the MC@NLO algorithm [77]. Different jet multiplicities are then merged into an inclusive sample using an improved CKKW matching procedure [78, 79] which is extended to NLO accuracy using the MEPS@NLO prescription [80]. The merging scale Q_{cut} is set to 20 GeV.

Uncertainties from missing higher orders in Sherpa samples are evaluated [81] using seven variations of the QCD renormalization (μ_{R}) and factorization (μ_{F}) scales in the matrix elements by factors of 0.5 and 2, avoiding variations in opposite directions. The strong coupling constant α_{S} is varied by ± 0.001 to assess the effect of its uncertainty. Additional details of the use of these samples are available in Ref. [82].

MadGraph5_aMC@NLO (CKKW-L): $V+$jets production is simulated with LO-

Table 3.2: The generator configurations used to simulate the signal and background processes for the $Z+D^{(*)}$ analysis. The acronyms ME, PS and UE stand for matrix element, parton shower and underlying event, respectively. The column "HF decay" specifies which software package is used to model the heavy-flavor decays of bottom and charmed hadrons. For multi-leg samples where different jet multiplicities are merged, the QCD accuracy for each jet multiplicity is specified.

Process	ME generator	QCD accuracy	ME PDF	PS generator	UE tune	HF decay
Z+jets (background modeling)						
Z+jets	Sherpa [2.2.11]	0–2j@NLO+3–5j@LO	NNPDF3.0nnlo	Sherpa	Default	Sherpa
Z+jets	MadGraph5_aMC@NLO (FxFx)	0–3j@NLO	NNPDF3.1nnlo_luxqed	Pythia8	A14	EvtGen
$Z+D^{(*)}$ (signal modeling and theory predictions)						
$Z+D^{(*)}$	Sherpa [2.2.11]	0–2j@NLO+3–5j@LO	NNPDF3.0nnlo	Sherpa	Default	EvtGen
$Z+D^{(*)}$	MadGraph5_aMC@NLO (FxFx)	0–3j@NLO	NNPDF3.1nnlo_luxqed	Pythia8	A14	EvtGen
Backgrounds						
W+jets	Sherpa [2.2.11]	0–2j@NLO+3–5j@LO	NNPDF3.0nnlo	Sherpa	Default	Sherpa
$t\bar{t}$	PowhegBox [v2]	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
Single-t, Wt	PowhegBox [v2]	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
Single-t, t-channel	PowhegBox [v2]	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
Single-t, s-channel	PowhegBox [v2]	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
$t\bar{t}V$	MadGraph5_aMC@NLO	NLO	NNPDF3.0nlo	Pythia8	A14	EvtGen
Diboson fully leptonic	Sherpa [2.2.12]	0–1j@NLO+2–3j@LO	NNPDF3.0nnlo	Sherpa	Default	Sherpa
Diboson hadronic	Sherpa [2.2.11]	0–1j@NLO+2–3j@LO	NNPDF3.0nnlo	Sherpa	Default	Sherpa

accurate matrix elements for up to four partons with MadGraph5_aMC@NLO [2.2.2] [38]. The matrix-element calculation is interfaced with Pythia8 [8.186] for the modeling of the parton shower, hadronization and underlying event. To remove overlap between the matrix element and the parton shower, the CKKW-L merging procedure [83, 84] is applied with a merging scale of $Q_{\text{cut}} = 30GeV$ and a jet-clustering radius parameter of 0.2. In order to better model the region of large jet p_T, the strong coupling α_S is evaluated at the scale of each splitting to determine the weight. The matrix-element calculation is performed with the NNPDF3.0nlo PDF set [74] with $\alpha_S = 0.118$. The calculation is done in the five-flavor scheme with massless b- and c-quarks. Cross-sections are calculated using a diagonal CKM matrix. Heavy-quark masses are reinstated in the Pythia8 shower. The values of μ_R and μ_F are set to one half of the transverse mass of all final-state partons and leptons. The A14 tune [85] of Pythia8 is used with the NNPDF2.3lo PDF set with $\alpha_S = 0.13$. The decays of bottom and charmed hadrons are performed by EvtGen [1.7.0] [86].

MadGraph5_aMC@NLO (FxFx): The MadGraph5_aMC@NLO [2.6.5] [38] program is used to generate weak bosons with up to three additional partons in the final state at NLO accuracy. The scales μ_R and μ_F are set to one half of the transverse mass of all final-state partons and leptons. Cross-sections are calculated using a diagonal CKM matrix. The showering and subsequent hadronization are performed using Pythia8 [8.240] with the A14 tune and the NNPDF2.3lo PDF set with $\alpha_S = 0.13$. The different jet multiplicities are

merged using the FxFx NLO matrix-element and parton-shower merging prescription [87]. Pythia8 [8.186] is used to model the parton shower, hadronization and underlying event.

The calculation uses a five-flavor scheme with massless b- and c-quarks at the matrix-element level, and massive quarks in the Pythia8 shower. At the event-generation level, the jet transverse momentum is required to be at least 10 GeV, with no restriction on the absolute value of the jet pseudorapidity. The PDF set used for event generation is `NNPDF3.1nnlo_luxqed`. The merging scale is set to $Q_{\text{cut}} = 20GeV$. Scale variations where μ_{R} and μ_{F} are varied independently by a factor of 2 or 0.5 in the matrix element are included as generator event weights. The decays of bottom and charmed hadrons are performed by EvtGen [1.7.0].

3.2.2 $V+D^{(*)}$ Signal samples

Only a small subset of events in the inclusive $V+$jets samples pass the $V+D^{(*)}$ fiducial requirements. This, coupled with the branching ratios of 9.2% (2.5%) to the D^+ (D^{*+}) decay mode of interest, means that even very large $V+$jets samples provide statistically inadequate measurements of the $V+D^{(*)}$ fiducial efficiency. Filtered signal samples are therefore used to enhance the statistical precision. The generated events are filtered to require the presence of a single lepton with $p_T > 15GeV$ and $|\eta| < 2.7$ and either a D^{*+} or a D^+ meson with $p_T > 7GeV$ and $|\eta| < 2.3$. EvtGen [1.7.0] is used to force all D^0 mesons to decay through the mode $D^0 \to K^-\pi^+$ and all D^+ mesons to decay through the mode $D^+ \to K^-\pi^+\pi^+$ (plus charge conjugates). EvtGen describes this three-body D^+ decay using a Dalitz plot amplitude that includes contributions from the $\overline{K}^{*0}(892)$, $\overline{K}^{*0}(1430)$, $\overline{K}^{*0}(1680)$ and $\kappa(800)$ resonances, as measured by CLEO-c [88].

These samples are used for signal modeling, for calculating the detector response matrix and fiducial efficiencies with small statistical uncertainties, and for determining the $V+D^{(*)}$ signal mass distribution used in the statistical analysis described in Chapters 8.1.1 and 8.2.1. The MG_aMC@NLO+Py8 (NLO) simulation described below is also used to calculate the theory predictions with the up-to-date PDF sets in Chapter 10 for the $W+D^{(*)}$ analysis. Three such filtered samples are used:

Sherpa [2.2.11] $V+D^{(*)}$: To reduce the per-event CPU time for the generation of the $V+D^{(*)}$ signal data sets, Sherpa [2.2.11] is configured to have lower perturbative accuracy than for the inclusive $V+$jets samples described above. Events are generated with NLO-accurate matrix elements for up to one jet, and LO-accurate matrix elements for two partons, in the five-flavor scheme. Other Sherpa parameters are set to the same values as for the baseline inclusive samples and uncertainties are evaluated using the same variations in QCD scale and α_{S} as for the baseline. The production cross-section for this configuration differs from that of the inclusive sample by ∼2%. The two configurations show no significant differences in kinematic distributions associated with the $D^{(*)}$ meson or V boson.

MG_aMC@NLO+Py8 (NLO) $W+D^{(*)}$: MadGraph5_aMC@NLO [2.9.3] is used to

generate the $W+c$-jet process at NLO accuracy. A finite charm quark mass of $m_c = 1.55GeV$ is used to regularize the cross-section, and a full CKM matrix is used to calculate the hard-scattering amplitudes. The values of μ_{R} and μ_{F} are set to half of the transverse mass of all final-state partons and leptons. The PDF set used for event generation is `NNPDF3.0nnlo` with $\alpha_{\mathrm{S}} = 0.118$. The matrix-element calculation is interfaced with Pythia8 [8.244] for the modeling of the parton shower, hadronization, and underlying event and the A14 tune is employed. Scale variations where μ_{R} and μ_{F} are varied independently by a factor of 2 or 0.5 in the matrix element are included as generator event weights. This sample was only used in the $W+D^{(*)}$ measurement.

MadGraph FxFx $V+D^{(*)}$: Events are generated using the same Pythia8 configuration as used for the inclusive MadGraph FxFx sample, but with the event-level filtering and configuration described above.

3.2.3 Top quark pair production background samples

The production of $t\bar{t}$ events is modeled using the PowhegBox [v2] [89, 90, 91, 92] generator which provides matrix elements at NLO in the strong coupling constant α_{S} with the `NNPDF3.0nlo` PDF and the h_{damp} parameter[1] set to $1.5\,m_{\mathrm{top}}$ [93]. The functional form of μ_{R} and μ_{F} is set to the default scale $\sqrt{m_{\mathrm{top}}^2 + p_T^2}$ where p_T is the transverse momentum of the top quark obtained using the underlying Born kinematics. Top quarks are decayed at LO using MadSpin [94, 95] to preserve all spin correlations. The events are interfaced with Pythia8 [8.230] for the parton shower and hadronization, using the A14 tune and the `NNPDF2.3lo` PDF set. The decays of bottom and charmed hadrons are simulated using EvtGen [1.6.0].

The NLO $t\bar{t}$ inclusive production cross-section is corrected to the theory prediction at NNLO in QCD including the re-summation of next-to-next-to-leading logarithmic (NNLL) soft-gluon terms calculated using Top++[2.0] [96, 97, 98, 99, 100, 101, 102].

Powheg+Herwig [7.04] and MadGraph5_aMC@NLO+Pythia8 [8] $t\bar{t}$ samples are used to estimate the systematic uncertainty due to the choice of MC model as explained in the following and the details of the configurations used are provided below.

$t\bar{t}$ Powheg+Herwig [7.04]: The impact of using a different parton shower and a different hadronization model is evaluated by comparing the nominal $t\bar{t}$ sample with another event sample produced with the PowhegBox [v2] generator using the `NNPDF3.0nlo` parton distribution function. Events in the latter sample are interfaced with Herwig [7.04] [103, 104], using the H7UE set of tuned parameters [104] and the `MMHT2014LO` PDF set [105]. The decays of bottom and charmed hadrons are simulated using EvtGen [1.6.0] [86].

[1]The h_{damp} parameter controls the transverse momentum p_T of the first additional emission beyond the leading-order Feynman diagram in the parton shower and therefore regulates the high-p_T emission against which the $t\bar{t}$ system recoils.

$t\bar{t}$ MadGraph5_aMC@NLO+Pythia8 [8]: The uncertainty in the matching of NLO matrix elements to the parton shower is assessed by comparing the Powheg sample with events generated with MadGraph5_aMC@NLO [2.6.0] interfaced with Pythia8 [8.230]. The MadGraph5_aMC@NLO calculation used the `NNPDF3.0nlo` set of PDFs and Pythia8 used the A14 tune and the `NNPDF2.3lo` set of PDFs. The decays of bottom and charmed hadrons are simulated using EvtGen 1.6.0.

3.2.4 Wt-channel single-top background samples

Single-top Wt associated production is modeled using the PowhegBox [v2] generator which provides matrix elements at NLO in the strong coupling constant α_S in the five-flavor scheme with the `NNPDF3.0nlo` parton distribution function set. The functional form of μ_R and μ_F is set to the default scale $\sqrt{m_{top}^2 + p_T^2}$. The diagram removal scheme [106] is employed to handle the interference with $t\bar{t}$ production [93]. Top quarks are decayed at LO using MadSpin to preserve all spin correlations. The events are interfaced with Pythia8 [8.230] using the A14 tune and the `NNPDF2.3lo` PDF set. The decays of bottom and charmed hadrons are simulated using EvtGen [1.6.0]. The inclusive cross-section is corrected to the theory prediction calculated at NLO in QCD with NNLL soft gluon corrections [107, 108].

3.2.5 t-channel and s-channel single-top background samples

Single-top t-channel (s-channel) production is modeled using the PowhegBox [v2] generator at NLO in QCD using the four-flavor (five-flavor) scheme and the corresponding `NNPDF3.0nlo` set of PDFs. The events are interfaced with Pythia8 [8.230] using the A14 tune and the `NNPDF2.3lo` set of PDFs.

The uncertainty due to initial-state radiation (ISR) is estimated by simultaneously varying the h_{damp} parameter and μ_R and μ_F, and choosing the Var3c up and down variants of the A14 tune as described in Ref. [109]. The impact of final-state radiation (FSR) is evaluated by halving and doubling the renormalization scale for emissions from the parton shower.

3.2.6 $t\bar{t} + V$ background samples

The production of $t\bar{t}V$ events, where V denotes either W, Z, or $\ell^+\ell^-$ produced through Z/γ interference, is modeled using the MadGraph5_aMC@NLO [2.3.3] [38] generator at NLO with the `NNPDF3.0nlo` parton distribution function. The events are interfaced with Pythia8 [8.210] using the A14 tune and the `NNPDF2.3lo` PDF set. The uncertainty due to ISR is estimated by comparing the nominal $t\bar{t}V$ sample with two additional samples, which have the same settings as the nominal one, but with the Var3 up or down variation of the A14 tune.

3.2.7 Diboson background samples

Samples of diboson final states (VV) are simulated with the Sherpa 2.2.11 or 2.2.12 [69] generator depending on the process (see Tables 3.1 and 3.2), including off-shell effects and Higgs boson contributions, where appropriate. Fully leptonic final states and semileptonic final states, where one boson decays leptonically and the other hadronically, are generated using matrix elements at NLO accuracy in QCD for up to one additional parton and at LO accuracy for up to three additional parton emissions. Samples for the gluon-loop-induced processes $gg \to VV$ are generated using LO-accurate matrix elements for up to one additional parton emission for both the cases of fully leptonic and semileptonic final states. The matrix-element calculations are matched and merged with the Sherpa parton shower based on Catani–Seymour dipole factorization using the MEPS@NLO prescription. The virtual QCD corrections are provided by the OpenLoops library. The `NNPDF3.0nnlo` set of PDFs is used along with the dedicated set of tuned parton-shower parameters developed by the Sherpa authors.

Matrix element to parton shower matching [77] is employed for different jet multiplicities, which are then merged into an inclusive sample using an improved CKKW matching procedure which is extended to NLO accuracy using the MEPS@NLO prescription. These simulations are NLO-accurate for up to one additional parton and LO-accurate for up to three additional partons. The virtual QCD correction for matrix elements at NLO accuracy is provided by the OpenLoops library. The calculation is performed in the G_μ scheme [110], ensuring an optimal description of pure electroweak interactions at the electroweak scale.

Chapter 4

Object Selection

This chapter lays out how the physics objects used in the analysis are selected. The selection and categorization of $Z+D^{(*)}$ and $W+D^{(*)}$ candidate events depend on the reconstruction and identification of the physics objects used: electrons, muons, tracks, and jets. These objects are chosen because they are part of the final state being measured, or they are necessary to distinct our signal from backgrounds. The selection criteria imposed on these objects are done to reduce backgrounds whilst increasing the signal. "Leptons" for both analyses include electrons and muons, but exclude τ-leptons. This chapter is split into object selection common to both analyses ($V+D^{(*)}$), and to each analysis where object selection differ. Chapter 4.1 gives the selection criteria for objects used in both analyses, Chapters 4.2 and 4.3 go over the objects selected for the $W+D^{(*)}$ and $Z+D^{(*)}$ analyses respectively.

4.1 $V+D^{(*)}$ Common Object Selection

4.1.1 Primary Vertex Selection

Proton-proton interaction vertices are reconstructed from charged-particle tracks with $p_T(\text{track}) > 500$ MeV in the ID. The presence of at least one such vertex with a minimum of two associated tracks is required, and the vertex with the largest sum of p_T^2 of associated tracks is chosen as the primary vertex (PV).

4.1.2 Track Selection

Tracks used in the electron and muon reconstruction are required to be associated with the PV, using constraints on the transverse impact parameter significance ($|d_0^{\text{BL}}/\sigma(d_0^{\text{BL}})|$) and on the longitudinal impact parameter (z_0^{BL}). The transverse impact parameter of the track is defined as the distance of closest approach of the track to the measured beamline position. The transverse impact parameter significance must satisfy $|d_0^{\text{BL}}/\sigma(d_0^{\text{BL}})| < 3.0$ for

muons and $|d_0^{\mathrm{BL}}/\sigma(d_0^{\mathrm{BL}})| < 5.0$ for electrons. The longitudinal impact parameter of the track is the longitudinal distance along the beamline between the point where $|d_0^{\mathrm{BL}}/\sigma(d_0^{\mathrm{BL}})|$ is measured and the primary vertex. Tracks are required to have $|z_0^{\mathrm{BL}} \sin\theta| < 0.5$ mm, where θ is the polar angle of the track with respect to the beamline axis.

4.1.3 Electron Selection

Electron candidates are reconstructed from an isolated energy deposit in the electromagnetic calorimeter matched to a track in the ID and must pass the tight likelihood-based working point [111]. The identification of prompt electrons relies on a likelihood discriminant constructed from quantities measured in the ID and the calorimeter. Electrons identified with Tight identification imposes a stringent selection on the likelihood discriminant. Electrons must be in the fiducial pseudorapidity region of $|\eta| < 2.47$, excluding the transition region $1.37 < |\eta| < 1.52$ between the calorimeter barrel and endcaps. The `tight` electrons are required to meet the "tight" isolation criteria [111], based on a combination of the track-based and calorimeter-based isolation. Isolation is defined as the sum of the p_T of all the particles which lie within a certain isolation cone around the particle of interest. Promptly produced leptons from the hard scatter, like those originating from our signal, tend to be isolated. The track-based isolation is $p_T^{\mathrm{iso}}(\Delta R_{\mathrm{var}} < 0.2)/p_T < 0.06$, with a variable cone size as defined in Chapter 3.1. The tracks are required to have $p_T^{\mathrm{track}} > 1\ GeV$ and are required to be associated with the primary vertex. The calorimeter-based isolation is $E_{\mathrm{T}}^{\mathrm{cone20}}/p_T < 0.06$, where $E_{\mathrm{T}}^{\mathrm{cone20}}$ is the sum of the transverse energy of positive-energy topological clusters whose barycenter falls within a $\Delta R < 0.2$ cone centred around the electron, corrected for the energy leakage, pileup, and underlying event, as described in Ref. [111]. Electron energy scale is calibrated following the procedure given in Ref. [111].

4.1.4 Muon Selection

Muon candidates are reconstructed in the region $|\eta| < 2.5$ by matching tracks in the MS with those in the ID. Muons are identified using the "Tight" ("Medium") quality criteria [112] for the W+$D^{(*)}$ (Z+$D^{(*)}$) analysis, characterized by the numbers of hits in the ID and MS subsystems. The global re-fitting algorithm [113] is used to combine the information from the ID and MS sub-detectors. The `tight` muons are required to pass the "tight" isolation working point, based on a combination of the track-based and particle-flow-based [114] isolation. The requirement is $(p_T^{\mathrm{iso}}(\Delta R_{\mathrm{var}} < 0.3)+0.4\times E_{\mathrm{T}}^{\mathrm{neflow20}})/p_T < 0.045$, where the track-based isolation uses a variable cone size as defined in Chapter 3.1, with a maximum size of $\Delta R = 0.3$. The tracks are required to have $p_T^{\mathrm{track}} > 500$ MeV and are required to be associated with the primary vertex. The $E_{\mathrm{T}}^{\mathrm{neflow20}}$ is the sum of the transverse energy of neutral-charge particle-flow objects in a cone of size $\Delta R < 0.2$ around the muon [112]. Muon momentum calibration is performed using the prescription in Ref. [113].

4.1.5 Overlap Removal

To avoid cases where the detector response to a single physical object is reconstructed as two different final-state objects, e.g., an electron reconstructed as both an electron and a jet, an overlap removal strategy is used. If the two calorimeter energy clusters from two electron candidates overlap, the electron with the highest E_{T} is retained. If a reconstructed electron and muon share the same ID track, the muon is rejected if it is calorimeter-tagged, meaning the muon is identified as a reconstructed ID track that extrapolates to the calorimeter energy deposit of a minimum-ionizing particle without a MS signal [112]; otherwise the electron is rejected. Next, jets within $\Delta R = 0.2$ of electrons are removed. In the last step, electrons and muons within $\Delta R = 0.4$ of any remaining jet are removed. In the case where a track can be either a pion or a kaon, we default to using the kaon mass hypothesis and assigning it to be a kaon unless otherwise noted.

4.2 $W+D^{(*)}$ Object Selection

4.2.1 Lepton Categories

Three different categories of leptons are used in this analysis: `baseline`, `loose`, and `tight`. These categories differ only in their isolation requirements. `Baseline` leptons are required to have $p_T > 20$ GeV. `Loose` and `tight` leptons are required to have $p_T > 30$ GeV. `Tight` leptons are required to meet isolation requirements and are used in our signal region. `Loose` leptons have no isolation requirements which enhances the fraction of leptons originating from non-prompt multijet processes. Anti-`tight` leptons are required to pass the `loose` requirements, but fail the `tight` requirements. Anti-`tight` leptons are used in the data-driven multijet production estimation described in Chapter 7.1.3. Full electron and muon selection criteria are summarized in Table 4.1.

Table 4.1: Lepton categories used in $W+D^{(*)}$ analysis.

	Electrons			Muons		
Features	`baseline`	`loose`	`tight`	`baseline`	`loose`	`tight`
p_T	> 20 GeV	> 30 GeV		> 20 GeV	> 30 GeV	
$\lvert\Delta z_0^{\mathrm{BL}} \sin(\theta)\rvert$	< 0.5 mm			< 0.5 mm		
$\lvert d_0^{\mathrm{BL}}/\sigma(d_0^{\mathrm{BL}})\rvert$	< 5			< 3		
Pseudorapidity	$(\lvert\eta\rvert < 1.37) \lvert\rvert (1.52 < \lvert\eta\rvert < 2.47)$			$\lvert\eta\rvert < 2.5$		
Identification	Tight			Tight		
Isolation	No	Yes		No	Yes	

4.2.2 Jet Selection

Jets are reconstructed from particle-flow objects [114] using the anti-k_T [115, 116] jet-reconstruction algorithm with a distance parameter $R = 0.4$. Candidate jets are required to have $p_T > 20$ GeV and $|\eta| < 5.0$. The jet energy scale (JES) calibration restores the jet energy to that of jets reconstructed at the particle level, as described in Ref. [117]. The jets from pileup interactions are suppressed using the Jet Vertex Tagger algorithm (JVT) [118].

4.2.3 Jet Flavor Tagging

Jets with $|\eta| < 2.5$ and $p_T > 20$ GeV containing b-hadrons are identified by a deep neural network tagger, DL1r [119, 120, 121], that uses displaced tracks, secondary vertices and decay topologies. The chosen working point has 70% efficiency for identifying $b-$jet in a simulated $t\bar{t}$ sample and the measured rejection factor (i.e. the inverse misidentification efficiency) for c-jets (light-jets) is about 11 (600) [121]. The $b-$jet are defined according to the presence of b-hadrons with $p_T > 5$ GeV within a cone of size $\Delta R = 0.3$ around the jet axis. If a b-hadron is not found and a c-hadron is found, then the jet is labeled a c-jet. Light-jets are all the rest.

4.2.4 Missing Transverse Momentum

The missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) in the events is calculated as the negative vector sum of the selected high-p_T calibrated objects (jets and `baseline` electrons and muons), plus a "soft term" reconstructed from tracks not associated with any of the calibrated objects [122, 123].

4.3 $Z+D^{(*)}$ Object Selection

4.3.1 Lepton Categories

Two different categories of leptons are used: `loose`, and `tight`. These categories differ only in their isolation requirements. `tight` leptons are required to meet isolation requirements and are used in our signal region. `loose` leptons have no isolation requirements which enhances the amount of leptons originating from non-prompt multijet processes. `loose` leptons are used in the data-driven multijet production estimation described in Chapter 7.2.3. Both `tight` and `loose` leptons are required to have $p_T > 27$ GeV. Full electron and muon selection criteria are summarized in Table 4.2.

Table 4.2: Lepton categories used in $Z+D^{(*)}$ analysis.

	Electrons		Muons	
Features	`loose`	`tight`	`loose`	`tight`
p_T (l_1)	> 27 GeV		> 27 GeV	
p_T (l_2)	> 27 GeV		> 27 GeV	
$\lvert\Delta z_0^{\mathrm{BL}} \sin(\theta)\rvert$	< 0.5 mm		< 0.5 mm	
$\lvert d_0^{\mathrm{BL}}/\sigma(d_0^{\mathrm{BL}})\rvert$	< 5		< 3	
Pseudorapidity	$(\lvert\eta\rvert < 1.37)$ || $(1.52 < \lvert\eta\rvert < 2.47)$		$\lvert\eta\rvert < 2.5$	
Identification	Tight		Medium	
Isolation	No	Yes	No	Yes

Chapter 5

Charm Meson Reconstruction

This chapter explains the construction and selection criteria of the $D^{(*)}$ mesons used in both analyses. Events containing c-quarks in the final state are identified by explicitly reconstructing charmed mesons in charged, hadronic decay channels. Two charmed hadron decay channels are used: $D^+ \to K^-\pi^+\pi^+$ and $D^{*+} \to D^0\pi \to (K^-\pi^+)\pi^+$ (and charge conjugates). The invariant mass distribution $m(D^+)$ (mass difference $m(D^{*+}-D^0)$) used in the fit for the D^+ (D^{*+}) channel is described in Chapters 8.1.1 and 8.2.1. Charm meson reconstruction and selection is identical to both $W+D^{(*)}$ and $Z+D^{(*)}$ analyses. Selection criteria were optimized to give the best signal significance (estimated as $S^2/(S+B)$) for OS-SS where the background under the peak was determined using the mass sidebands in the data and the variation of the peak size with respect to the selection cuts was estimated using MC. Chapter 5.1 explains how each meson is reconstructed, Chapters 5.2 and 5.3 give the selection criteria of the D^+ and D^* mesons, Chapter 5.4 summarizes selection criteria which aim to reduce backgrounds.

5.1 Meson Reconstruction Overview

Charm mesons are identified through the final state particles they decay to. In the case of both D^+ and D^* mesons, the decays result in three tracks corresponding to the two pions and the kaon. Selection on the tracks used to reconstruct the $D^{(*)}$ meson candidates depends on reducing candidates reconstructed due to pileup events and not the hard scatter being measured. ID tracks must satisfy $|\eta| < 2.5$ and $|z_0 \sin\theta| < 5$ mm with the Loose track quality requirement is applied [124]. The D^+ (D^0) candidate is reconstructed using ID tracks with $p_T > 800(600)$ MeV. In order to ensure these tracks come from the same $D^{(*)}$ meson decay point a geometric separation of $\Delta R < 0.6$ between the tracks is required. Tracks identified as belonging to `baseline` leptons used for the Z boson or W boson candidates are excluded in this reconstruction.

D^+ candidates are required to have three tracks with total charge = ± 1. The two tracks

with the same charge are assigned the charged pion mass and the remaining track is assigned the kaon mass. The D^0 candidates are required to have two tracks with total charge = 0. One track is assigned the charged pion mass and the other is assigned the charged kaon mass. Both possible choices for the mass assignment are retained until matching to the prompt pion from the D^{*+} decay is performed. Tracks from the D^+ (D^0) candidate are fitted to a common secondary vertex (SV), with a fit χ^2 required to be $\chi^2 < 8.0$ (10.0). To reduce the contribution from pileup and from b-hadron decays, the transverse impact parameter of the $D^{(*)}$ candidate's flight path with respect to the PV is required to satisfy $|d_0| < 1.0$ mm and the candidate is required to have a 3D impact parameter significance $\sigma_{3D} < 4.0$, where σ_{3D} is the distance of closest approach of the candidate's flight path to the PV divided by the uncertainty in that distance. These selection criteria and those described below were determined by optimizing the signal significance, using MC predictions to estimate the signal, and mass sidebands to estimate the background.

5.2 D^+ Meson Reconstruction

Several requirements are placed on the D^+ candidates to reduce combinatorial background coming from the many tracks produced in pp collisions. The angle between the kaon track in the rest frame of the D^+ candidate and the line of flight of the D^+ candidate in the center-of-mass frame is required to satisfy $\cos\theta^*(K) > -0.8$. The distance between the SV and the PV in the transverse plane is required to satisfy $L_{xy} > 1.1$ mm for D^+ candidates with $p_T < 40$ GeV and $L_{xy} > 2.5$ mm for D^+ candidates with $p_T > 40$ GeV. Due to the many possible decays the $D^{(*)}$ mesons can undergo that do not match the signal decay modes kinematic requirements are applied to ensure orthogonality to other $D^{(*)}$ decays with similar final states. The contamination from $D^{*+} \to D^0\pi^+ \to (K^-\pi^+)\pi^+$, which has the same final-state content as the $D^+ \to K^-\pi^+\pi^+$ channel, is reduced by requiring $m(K\pi\pi) - m(K\pi) > 160$ MeV. Background from the $D_s^\pm \to \phi\pi^\pm \to (K^+K^-)\pi^\pm$ channel, with one of the kaons misidentified as a pion, is removed by requiring the mass of each pair of oppositely charged particles, assuming the kaon mass hypothesis, to be $m(K^+K^-) > |m_\phi - 8|$ MeV. The world-average mass of the ϕ meson from the Particle Data Group (PDG) database [125], $m_\phi = 1019.455$ MeV, is used. An explicit removal of decay modes including the ϕ meson is performed; such removal is possible without significantly reducing the signal efficiency because the ϕ peak is narrow. Backgrounds from other charm decay modes are included in the fit templates as part of the $V{+}c^{\text{match}}$ category, for example $D_s \to K^*K$. The reliability on this estimate relies on the MC using correct production fractions and correct charm branching ratios. As explained in Chapter 3.2, production fraction have been re-weighted to the world average. The branching fractions for relevant charm decays used in EVTGEN have been compared to the most recent PDG averages have found to be acceptable. Finally, a requirement is placed on the invariant mass of the D^+ candidates, $1.7\,\text{GeV} < m(D^+) < 2.2\,\text{GeV}$.

5.3 D^* Reconstruction

D^{*+} candidates are reconstructed by combining D^0 candidates with prompt tracks that are assigned the charged pion mass. Only combinations where the pion in the D^0 candidate has the same charge as the prompt pion are selected. The small mass difference between the D^{*+} and D^0 mesons restricts the phase space of this associated prompt pion, which has low momentum in the D^0 rest frame and is referred to as the slow pion. Slow pion tracks are required to have $p_T > 500$ MeV and a transverse impact parameter of $|d_0| < 1.0$ mm with respect to the primary vertex. An $L_{xy} > 0$ mm requirement is applied to D^0 candidates. The mass of the D^0 candidate must be within 40 MeV of the PDG world-average value of the D^0 mass, $m_{D^0} = 1864.83$ MeV [125]. Additionally, the angular separation between the slow pion and the D^0 meson must be small, $\Delta R(\pi_{\text{slow}}, D^0) < 0.3$. An invariant mass cut of $140\,\text{MeV} < m(D^{*+} - D^0) < 180\,\text{MeV}$ is imposed.

5.4 $D^{(*)}$ Meson Background Reduction

Combinatorial background is produced primarily from pileup events and from events containing light jets. This background is reduced by requiring $D^{(*)}$ candidates to be isolated. The transverse momenta of tracks in a cone of size $\Delta R = 0.4$ around the $D^{(*)}$ candidate are summed, and the sum is required to be less than the p_T of the $D^{(*)}$. Background from semi-leptonic B meson decays is reduced by requiring $\Delta R(D^{(*)}, \ell) > 0.3$. Finally, the $D^{(*)}$ candidates are required to have $8\,\text{GeV} < p_T < 150\,\text{GeV}$ and $|\eta| < 2.2$. This upper p_T cut is applied to reject the background from fake $D^{(*)}$ mesons at high momentum. The predicted fraction of $D^{(*)}$ mesons with $p_T(D^{(*)}) > 150\,\text{GeV}$ is small and has no significant impact on the signal reconstruction efficiency. The η cut is applied to avoid the edge of the ID, where the amount of the detector material increases rapidly and thus reduces the reconstruction efficiency and degrades the resolution. The full set of selection requirements for the $D^{(*)}$ candidates is summarized in Table 5.1.

In order to increase statistics and due to a difference in MC modeling in the $Z+D^{(*)}$ analysis, the 3D Impact significance selection is not used, the isolation parameter is loosened to a cut of 2 and the D^0 mass window is widened to 60 GeV. This "looser" selection introduces a larger amount of non-prompt D mesons originating from B meson decays.

Table 5.1: $D^{(*)}$ object selection criteria. For D^{*+} candidates the cuts related to SV reconstruction are applied to the corresponding D^0 candidate.

$D^{(*)}$ cut	D^+ cut value	D^{*+} cut value ($D^0\pi \to (K\pi)\pi$)
N_{tracks} at SV	3	2
SV charge	± 1	0
SV fit quality	$\chi^2 < 8$	$\chi^2 < 10$
Track p_T	$p_T > 800$ MeV	$p_T > 600$ MeV
Track angular separation	$\Delta R < 0.6$	$\Delta R < 0.6$
Flight length	$L_{xy} > 1.1$ mm $\left(p_T(D^+) < 40 \text{ GeV}\right)$ $L_{xy} > 2.5$ mm $\left(p_T(D^+) \geq 40 \text{ GeV}\right)$	$L_{xy} > 0$ mm
SV impact parameter	$\lvert d_0 \rvert < 1$ mm	$\lvert d_0 \rvert < 1$ mm
SV 3D impact significance*	$\sigma_{\text{3D}} < 4.0$, —	$\sigma_{\text{3D}} < 4.0$, —
Combinatorial background rejection	$\cos\theta^*(K) > -0.8$	—
Isolation*	$\Sigma {p_T}_{\text{tracks}}^{\Delta R<0.4} / p_T(D^+) <$ 1.0(2.0)	$\Sigma {p_T}_{\text{tracks}}^{\Delta R<0.4} / p_T(D^{*+}) <$ 1.0(2.0)
$D_s^\pm \to \phi\pi^\pm$ rejection	$m(K^+K^-) > \lvert m_\phi - 8 \rvert$ MeV	—
D^{*+} background rejection	$m(K\pi\pi) - m(K\pi) >$ 160 MeV	—
D^0 mass*	—	$\lvert m_{K\pi} - m_{D^0} \rvert < 40$ MeV (60 MeV)
π_{slow} p_T	—	$p_T > 500$ MeV
π_{slow} angular separation	—	$\Delta R(\pi_{\text{slow}}, D^0) < 0.3$
π_{slow} d_0	—	$\lvert d_0 \rvert < 1$ mm
QCD background rejection	$\Delta R(D^+, \ell) > 0.3$	$\Delta R(D^{*+}, \ell) > 0.3$
$D^{(*)}$ p_T	8 GeV $< p_T(D^+) <$ 150 GeV	8 GeV $< p_T(D^{*+}) <$ 150 GeV
$D^{(*)}$ η	$\lvert \eta(D^+) \rvert < 2.2$	$\lvert \eta(D^{*+}) \rvert < 2.2$
Invariant mass	1.7 GeV $< m(D^+) <$ 2.2 GeV	140 MeV $< m(D^{*+}-D^0) <$ 180 MeV

Chapter 6

Event Selection

This chapter reviews the selection criteria implemented on the composite objects and event level variables of the analyses which pass our fiducial definition. Event selection for both analyses rely on the reconstruction of the leptons, V boson, jets, $E_{\mathrm{T}}^{\mathrm{miss}}$ and $D^{(*)}$ meson. The subsequent selection satisfying the criteria defined in Chapters 4 and 5 and passing the single-lepton triggers as discussed in Chapter 3 must also be met. Further selections required in the truth objects to perform truth level unfolding are explained in Chapter 8. Chapters 6.1 and 6.2 review the event selection utilized for the $W+D^{(*)}$ and $Z+D^{(*)}$ analyses respectively. The choice of the minimum p_T of the charged lepton was chosen to reduce the amount of multijet background which is poorly determined from the data. The selection on the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} variables were chosen to reduce the background of QCD production and to provide a symmetric selection on the charged lepton and the neutrino.

6.1 $W+D^{(*)}$ Event Selection

Event selection of the $W+D^{(*)}$ analysis is described for both reconstruction level objects and truth level objects. The full selection on both reconstruction level and truth level objects is given in Tables 6.1a and 6.1b.

6.1.1 Reconstructed Event Selection

Reconstruction of W bosons is based on their leptonic decays to either an electron ($W \to e\nu$) or a muon ($W \to \mu\nu$). The lepton is explicitly measured in the detector and the presence of a neutrino is inferred from calculating $E_{\mathrm{T}}^{\mathrm{miss}}$ in an event. Events are required to have exactly one `tight` lepton with $p_T >$ 30 GeV and $|\eta| < 2.5$. Events with additional leptons are rejected. To reduce the multijet background and enhance the W boson signal purity, additional requirements are imposed: $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV and $m_{\mathrm{T}} > 60$ GeV, where the W boson transverse mass (m_{T}) is defined as $\sqrt{2p_T(\mathrm{lep})E_{\mathrm{T}}^{\mathrm{miss}}(1-\cos(\Delta\phi))}$ and $\Delta\phi$ is the

Table 6.1: Tables summarizing the event selection in the $W+D^{(*)}$ analysis: (a) fit regions used in the statistical analysis and (b) the "truth" fiducial selection. The $W+D^{(*)}$ signal is defined by performing the OS-SS subtraction as described in the text.

(a)

Detector-level selection		
Requirement	$W+D^{(*)}$ SR	Top CR
$N(b\text{-jet})$	0	≥ 1
$E_{\mathrm{T}}^{\mathrm{miss}}$	$> 30\ GeV$	
m_{T}	$> 60\ GeV$	
Lepton p_T	$> 30\ GeV$	
Lepton $\lvert\eta\rvert$	< 2.5	
$N(D^{(*)})$	≥ 1	
$D^{(*)}\ p_T$	$> 8\ GeV$ and $< 150\ GeV$	
$D^{(*)}\ \lvert\eta\rvert$	< 2.2	

(b)

Truth fiducial selection	
Requirement	$W+D^{(*)}$
$N(b\text{-jet})$	—
$E_{\mathrm{T}}^{\mathrm{miss}}$	—
m_{T}	—
Lepton p_T	$> 30GeV$
Lepton $\lvert\eta\rvert$	< 2.5
$N(D^{(*)})$	≥ 1
$D^{(*)}\ p_T$	$> 8GeV$
$D^{(*)}\ \lvert\eta\rvert$	< 2.2

azimuthal separation between the lepton and the missing transverse momentum. Candidate $D^{(*)}$ mesons are reconstructed using a secondary-vertex fit as described in Chapter 5. Any number of $D^{(*)}$ meson candidates satisfying these criteria are selected, which accounts for the production of multiple mesons in a single event. Only events with one or more $D^{(*)}$ candidates are selected.

Events passing this selection are used to extract the $W+D^{(*)}$ observables with a profile likelihood fit defined in Chapters 8.1.1 and 8.2.1. Selected events are categorized according to the b-jet multiplicity to separate the $W+D^{(*)}$ signal process from the $t\bar{t}$ background with $W \to cs$ decays. The ID tracks associated with the reconstructed $D^{(*)}$ candidates can also be associated with a jet mis-tagged as a b-jet. To reduce this inefficiency, b-jets are required to be geometrically separated from reconstructed $D^{(*)}$ mesons by satisfying $\Delta R(b\text{-jet}, D^{(*)}) > 0.4$. This ensures avoiding the categorization of these $W+D^{(*)}$ signal events as events with one more more b-jets. Events with exactly zero such b-tagged jets are classified as the $W+D^{(*)}$ signal region (SR) and events with one or more b-tagged jets comprise the Top control region (CR). This stratification results in about 80% of the $t\bar{t}$ background events populating the Top CR whilst retaining about 99% of $W+D^{(*)}$ signal events in the $W+D^{(*)}$ SR, effectively reducing the amount of $t\bar{t}$ background. Collectively, the $W+D^{(*)}$ SR and Top CR are called the "fit regions". The yield of signal events is about 5% of the $t\bar{t}$ background yield in the Top CR. These requirements are summarized in Table 6.1a.

This analysis exploits the charge correlation of the W boson and the charm quark to enhance the signal and reduce the backgrounds. The signal has a W boson and a $D^{(*)}$

meson of opposite charge, while most backgrounds are symmetric in charge. Therefore, the signal is extracted by measuring the difference between the numbers of opposite-sign (OS) and same-sign (SS) $W+D^{(*)}$ candidates, which is referred to as OS-SS. This strategy results in a reduction of background events that survive the OS-SS subtraction. While the signal-to-background ratio is about unity in the OS region, the OS-SS $W+D^{(*)}$ signal is an order of magnitude larger than the remaining background after the subtraction.

6.1.2 Truth Fiducial Selection

The $W+D^{(*)}$ measurement is unfolded to a "truth" fiducial region defined at MC particle level to have exactly one "truth" lepton with $p_T(\ell) > 30$ GeV and $|\eta(\ell)| < 2.5$. The lepton must originate from a W boson decay, with τ decays excluded from the fiducial region. Lepton momenta are calculated using "dressed" leptons, where the four-momenta of photons radiated from the final-state leptons within a cone of $\Delta R = 0.1$ around the lepton are added to the four-momenta of leptons. Truth $D^{(*)}$ mesons are selected by requiring $p_T(D^{(*)}) > 8$ GeV and $|\eta(D^{(*)})| < 2.2$. The OS-SS subtraction is also applied to the truth fiducial events. This removes any charge-symmetric processes, which are expected to originate mostly from gluon splitting in the final state. The $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} requirements and b-jet veto are not applied in the fiducial selection. The truth fiducial selection is summarized in Table 6.1b. The fiducial efficiency is defined as the fraction of $W+D^{(*)}$ signal events from the truth fiducial region that pass the detector-level reconstruction and requirements in Table 6.1a. In the unfolding, events where the reconstructed objects pass the event selection but the truth objects fail the truth fiducial requirements are treated as fakes; cases where the reconstructed objects fail the reconstruction fiducial selection but the truth objects pass the truth selection are treated as inefficiencies.

6.2 $Z+D^{(*)}$ Event Selection

Event selection of the $Z+D^{(*)}$ analysis is described for both reconstruction level objects and truth level objects. The full selection on both reconstruction level and truth level objects is given in Tables 6.2a and 6.2b.

6.2.1 Reconstructed Event Selection

Reconstruction of Z bosons is based on their leptonic decays to either a pair of electrons ($Z \to ee$) or a pair of muons ($Z \to \mu\mu$). Events are required to have exactly two `tight` leptons with $p_T >$ 27 GeV and $|\eta| < 2.5$. The pair of leptons are also required to be oppositely charged. Events with additional leptons are rejected. Candidate $D^{(*)}$ mesons are reconstructed using a secondary-vertex fit as described in Chapter 5. Any number of $D^{(*)}$

Table 6.2: Tables summarizing the event selection in the $Z+D^{(*)}$ analysis: (a) fit regions used in the statistical analysis and (b) the "truth" fiducial selection.

(a)

Detector-level selection	
Requirement	$Z+D^{(*)}$ SR
Lepton 1 p_T	$> 27\ GeV$
Lepton 2 p_T	$> 27\ GeV$
Lepton 1 $\|\eta\|$	< 2.5
Lepton 2 $\|\eta\|$	< 2.5
$m(Z)$	> 76 and $< 106\ GeV$
$N(D^{(*)})$	≥ 1
$D^{(*)}$ p_T	> 8 and $< 150\ GeV$
$D^{(*)}$ $\|\eta\|$	< 2.2

(b)

Truth fiducial selection	
Requirement	$Z+D^{(*)}$
Lepton 1 p_T	$> 27GeV$
Lepton 2 p_T	$> 27GeV$
Lepton 1 $\|\eta\|$	< 2.5
Lepton 2 $\|\eta\|$	< 2.5
$m(Z)$	> 76 and $< 106\ GeV$
$N(D^{(*)})$	≥ 1
$D^{(*)}$ p_T	> 8 and $< 150\ GeV$
$D^{(*)}$ $\|\eta\|$	< 2.2

meson candidates satisfying these criteria are selected, which accounts for the production of multiple mesons in a single event. Only events with one or more $D^{(*)}$ candidates are selected.

Events selected in this way are used to extract the $Z+D^{(*)}$ observables with a profile likelihood fit defined in Chapter 8.2.1. No explicit $\Delta R(b\text{-jet}, D^{(*)})$ selection is performed for this analysis. Events passing the above selection are classified as the $Z+D^{(*)}$ signal region (SR). These requirements are summarized in Table 6.2a.

6.2.2 Truth Fiducial Selection

The $Z+D^{(*)}$ measurement is unfolded to a "truth" fiducial region defined at MC particle level to have exactly two "truth" leptons with $p_T(\ell) > 27$ GeV and $|\eta(\ell)| < 2.5$. The leptons must originate from a Z boson decay, with τ decays excluded from the fiducial region. The leptons must also be same flavor (ee or $\mu\mu$) and be oppositely charged. Lepton momenta are calculated using "dressed" leptons, where the four-momenta of photons radiated from the final-state leptons within a cone of $\Delta R = 0.1$ around the lepton are added to the four-momenta of leptons. Truth $D^{(*)}$ mesons are selected by requiring $p_T(D^{(*)}) > 8\,\text{GeV}$ and $< 150\,\text{GeV}$ and $|\eta(D^{(*)})| < 2.2$. The truth fiducial selection is summarized in Table 6.2b. The fiducial efficiency is defined as the fraction of $Z+D^{(*)}$ signal events from the truth fiducial region that pass the detector-level reconstruction and requirements in Table 6.2a. In the unfolding, events where the reconstructed objects pass the event selection but the truth objects fail the truth fiducial requirements are treated as fakes; cases where the reconstructed objects fail the reconstruction fiducial selection but the truth objects pass the truth selection

are treated as inefficiencies.

Chapter 7

Signal and Background Modeling

This chapter explains how the signal and background mass templates are constructed. MC samples are used to construct signal and background mass templates, except for the multijet background, which is determined using a data-driven method (Chapters 7.1.3 and 7.2.3). Generally, SHERPA (2.2.11) MC samples are used to model events containing a single V boson and one or more reconstructed $D^{(*)}$ meson candidates because they provide the highest precision when simulating QCD processes and the highest statistical power among the available samples. For specific purposes, MG+Py8 (CKKW-L), MG_aMC@NLO+Py8 (FxFx) and MG_aMC@NLO+Py8 (NLO) MC samples are used in conjunction with SHERPA (2.2.11) to account for shortcomings in SHERPA (2.2.11) modeling of $D^{(*)}$ meson decays as described in Chapters 7.1.1 and 7.1.2. Chapters 7.1 and 7.2 describe the event templates utilized for the $W+D^{(*)}$ and $Z+D^{(*)}$ analyses respectively.

7.1 $W+D^{(*)}$ Templates

MC truth information is used to categorize the MC $W+D^{(*)}$ candidate according to the origin of the tracks used to reconstruct the $D^{(*)}$ meson:

- $W+D^{(*)}$ signal: If all tracks originate from the signal charmed hadron species (D^*) and are assigned in the reconstruction to the correct particle species ($K^{\mp}\pi^{\pm}\pi^{\pm}$), then that reconstructed $D^{(*)}$ candidate is labeled as $W+D^{(*)}$ signal.

- $W+c^{\text{match}}$: If all tracks originate either from a different charmed hadron species (D^0, D_s, or c-baryon) or from a different decay mode of a signal charmed meson (e.g. $D^{*+} \to D^0\pi^+ \to (K^-\pi^+)\pi^+$), the reconstructed $D^{(*)}$ candidate is labeled as $W+c^{\text{match}}$.

- $W+c^{\text{mis-match}}$: If at least one but not all tracks belong to a single charmed hadron, the reconstructed $D^{(*)}$ candidate is labeled as $W+c^{\text{mis-match}}$.

- W+jets: if none of the tracks are matched to a particle originating from a charmed particle, the $D^{(*)}$ candidate is labeled W+jets. This is the combinatorial background from the underlying event and pileup.

Additional background categories modeled using MC simulation are:

- Top: Processes containing top quarks ($t\bar{t}$, single-t, $t\bar{t}X$) are jointly represented by the "Top" category, which is dominated by the $t\bar{t}$ process.
- Other: Events from diboson and Z+jets processes are combined into the "Other" category.

The signal and background samples used in the $W+D^{*+}$ fits are given in Table 7.1. The rates at which c-quarks hadronize into different species of weakly decaying charmed hadrons in the MC samples are re-weighted to the world-average values [68]. The weights improve agreement between data and MC simulation by modifying the signal and background normalizations and the shapes of the $W+D^{(*)}$ background templates by changing the relative contribution of each species. The normalization of the background templates changes by up to 3%, depending on the $D^{(*)}$ species.

Table 7.1: Single-W-boson MC samples employed to create mass templates used in the $W+D^{(*)}$ fits. The "Normalization" and "Shape" columns indicate the source used to calculate the corresponding property. "LIS" refers to the Loose Inclusive Selection explained in the text. The MC configurations used to model these backgrounds are described in Chapter 3.2. Preferentially, SHERPA (2.2.11) samples are used for signal and background modeling. There are some exceptions to account for the shortcomings as explained in the text (e.g. incorrect D^{*+} decay with in SHERPA (2.2.11)).

Category	Normalization	$m(D^{*+}-D^0)$ shape
$W+D^{*+}$	SHERPA (2.2.11)	MG_aMC@NLO+Py8 (NLO)
$W+c^{\text{match}}$	SHERPA (2.2.11)	SHERPA (2.2.11)
$W+c^{\text{mis-match}}$	SHERPA (2.2.11)	LIS SHERPA (2.2.11)
W+jets	MG+Py8 (CKKW-L)	LIS MG+Py8 (CKKW-L)

7.1.1 Signal Modeling

The SHERPA (2.2.11) $W+D^{*+}$ signal sample with EvtGen decays is the default MC in this analysis and would nominally be used for the modeling of the mass difference template. However, because the width of the D^{*+} meson is set incorrectly in SHERPA (2.2.11), the mass shape is taken from the MG_aMC@NLO+Py8 (NLO) $W+D^{*+}$ signal sample instead.

The normalization is taken from SHERPA (2.2.11) because it provides the best available statistical power for calculating the fiducial efficiency.

7.1.2 Modeling Backgrounds with Monte Carlo Simulations

The W+c^{match} and W+$c^{\text{mis-match}}$ backgrounds are modeled using SHERPA (2.2.11). The W+jets background is modeled using MG+Py8 (CKKW-L) in the D^{*+} channel because its descriptions of this background yield and invariant mass difference shape is closer to the data before the fit. These background MC samples suffer from large statistical uncertainties. A Loose Inclusive Selection (LIS) method was developed to reduce these uncertainties. The LIS method is based on the observation that, for these backgrounds, the $D^{(*)}$ meson mass shapes are the same for both W boson charges and do not depend on the $E_{\text{T}}^{\text{miss}}$ and m_{T} cuts. Therefore, the LIS can be used to construct mass templates inclusively and without $E_{\text{T}}^{\text{miss}}$ and m_{T} cuts. These inclusive mass distributions are then used for both W boson charges. The LIS W+jets background is then fitted with a parametric function. This parametric function is then used to generate the template histogram which is used in the W+D^{*+} fit.

7.1.3 Data-Driven Multijet Background Estimation

Multijet backgrounds arise if one or more constituents of a jet are misidentified as a prompt lepton. In the electron channel, multijet events pass the electron selection due to having misidentified hadrons, converted photons or semi-leptonic heavy-flavor decays. In the muon channel, muons from heavy-flavor hadron decays are the dominant source. Collectively, these backgrounds are called "fake and non-prompt leptons". MC-based predictions for the normalization and composition of these backgrounds suffer from large uncertainties due to difficulty in modeling this background. The background rate is therefore determined using the data-driven Matrix Method [126].

The Matrix Method takes advantage of the fact that fake and non-prompt leptons (F) are less well isolated than real leptons (R). Leptons can be split independently in two ways: by origin (R and F), or by the `tight` (T) and `loose` reconstruction criteria defined in Table 4.1. Leptons satisfying the `loose` but not the `tight` criteria are labeled as anti-`tight` (!T). While the amount of R and F leptons (N_{R} and N_{F}) are not directly measurable in data, they can be related to the measurable numbers of `tight` and anti-`tight` leptons (N_{T} and $N_{!\text{T}}$) via the efficiency r (f) for a `loose` real (fake) lepton to also be `tight`:

$$\begin{pmatrix} N_{\text{T}} \\ N_{!\text{T}} \end{pmatrix} = \begin{pmatrix} r & f \\ 1-r & 1-f \end{pmatrix} \begin{pmatrix} N_{\text{R}} \\ N_{\text{F}} \end{pmatrix},$$

This expression is inverted to give an expression for the number of fake and non-prompt leptons in the W+$D^{(*)}$ SR, dependent on measurable quantities:

$$N_{\mathrm{T}}^{\mathrm{fake}} = \frac{f}{r-f}\left((r-1)N_{\mathrm{T}} + rN_{\mathrm{!T}}\right).$$

This Matrix Method relation is applied bin-by-bin to estimate the multijet background yield in the variable of interest. The real-lepton efficiency r is determined from the data in auxiliary measurements [111, 113] and extrapolated to the W+$D^{(*)}$ SR using MC samples. The real-lepton efficiency is estimated in 3 (4) bins in η for electrons (muons) and in p_T bins of 6 GeV width. The fake-lepton efficiency f is computed from the data in a dedicated region enriched in fake and non-prompt leptons, called the Fake CR. This region, orthogonal to the W+$D^{(*)}$ SR, is selected by inverting the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} requirements to $E_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV and $m_{\mathrm{T}} < 40$ GeV. These requirements reduce the contribution of real leptons originating from W boson decays.

To further increase the Fake CR's purity in fake and non-prompt leptons, processes with real leptons are estimated from MC simulation and subtracted from both `tight` and anti-`tight` subsets of Fake CR. The OS-SS subtraction is not performed for the calculation of the fake-lepton efficiencies because the multijet background is largely symmetric in OS and SS events. The number of `tight` leptons divided by the sum of `tight` and anti-`tight` gives the fake-lepton efficiency. The efficiency is estimated in 3 (4) bins in η for electrons (muons) and in p_T bins of 5 GeV to 20 GeV width, depending on the available sample size. The fake-lepton efficiency, in the Fake CR, is in the range 50%–90% or 10%–70% for electrons and muons respectively.

$$f = N_{\mathrm{T}}/(N_{\mathrm{T}} + N_{\mathrm{!T}}).$$

Systematic uncertainties in the multijet estimation arise from several sources. Statistical uncertainties in the determination of the real- and fake-lepton efficiencies lead to systematic uncertainties of approximately 10% to 20% in the overall multijet yield. Uncertainties in the size of the real-lepton contamination in the Fake CR are assessed using two methods. First, the change in rate due to varying the QCD renormalization and factorization scales in MC samples is obtained. Second, the difference between the prompt rates determined using MG+Py8 (CKKW-L) or SHERPA (2.2.11) W+jets MC samples is evaluated. These two variations together result in relative uncertainties on the multijet yield of ∼30% for the D^{*+} channel.

An additional systematic uncertainty is derived to account for the dependence of fake-lepton efficiencies on $E_{\mathrm{T}}^{\mathrm{miss}}$, which may arise from the different composition of fake background processes depending on the $E_{\mathrm{T}}^{\mathrm{miss}}$ (e.g. misidentified hadrons or semileptonic heavy-flavor decays), the correlation between the lepton isolation variables and $E_{\mathrm{T}}^{\mathrm{miss}}$, and the tendency of misidentified objects (e.g. jets misidentified as electrons) to give rise to $E_{\mathrm{T}}^{\mathrm{miss}}$ due to incorrect assumption about the object type in their energy calibration. To estimate this, the Fake CR's $E_{\mathrm{T}}^{\mathrm{miss}}$ cut is inverted to require $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV while its m_{T} cut is retained to ensure orthogonality with the W+$D^{(*)}$ SR. This process provides an independent

estimate of the multijet background. Differences between the multijet background yields in the $W+D^{(*)}$ SR obtained with these two choices of Fake CR cuts are ∼60% for D^{*+}. While this multijet background estimate has large systematic uncertainties, the multijet yield in the $W+D^{(*)}$ SR is only up to 1% of the signal yield in the electron channel and negligible in the muon channel. Thus the multijet background uncertainties are subdominant when estimating the overall background yield.

Figure 7.1 demonstrates the extrapolation of the multijet background from the Fake CR to the $W+D^{(*)}$ SR. Without the OS-SS subtraction, most of the D mesons in the Top background originate from B meson decays. The central values of the fake-lepton efficiencies are calculated in the $m_{\mathrm{T}} < 40$ GeV region, but with the $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement inverted ($E_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV). The figure instead shows the events with the $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV requirement corresponding to the $W+D^{(*)}$ SR selection. The prediction disagrees with the data at low m_{T} due to an $E_{\mathrm{T}}^{\mathrm{miss}}$ dependence in the fake-lepton efficiencies that is not directly accounted for in the parameterization. A systematic uncertainty is introduced, as described above, by calculating the fake-lepton efficiencies with the $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV requirement and taking the full difference between the two multijet predictions as the uncertainty. Since this is the largest systematic uncertainty in the multijet background, the data is almost exactly covered by the one-standard-deviation variation in this region. Furthermore, the multijet prediction and the uncertainties are extrapolated into the $W+D^{(*)}$ SR with the $m_{\mathrm{T}} > 60$ GeV requirement. To validate the extrapolation, the prediction is evaluated in a validation region (VR) with an m_{T} requirement of 40 GeV $< m_{\mathrm{T}} <$ 60 GeV. Figure 7.1 shows that the prediction in the VR is in agreement with the data within the systematic uncertainties, indicating that the multijet background is modeled well enough.

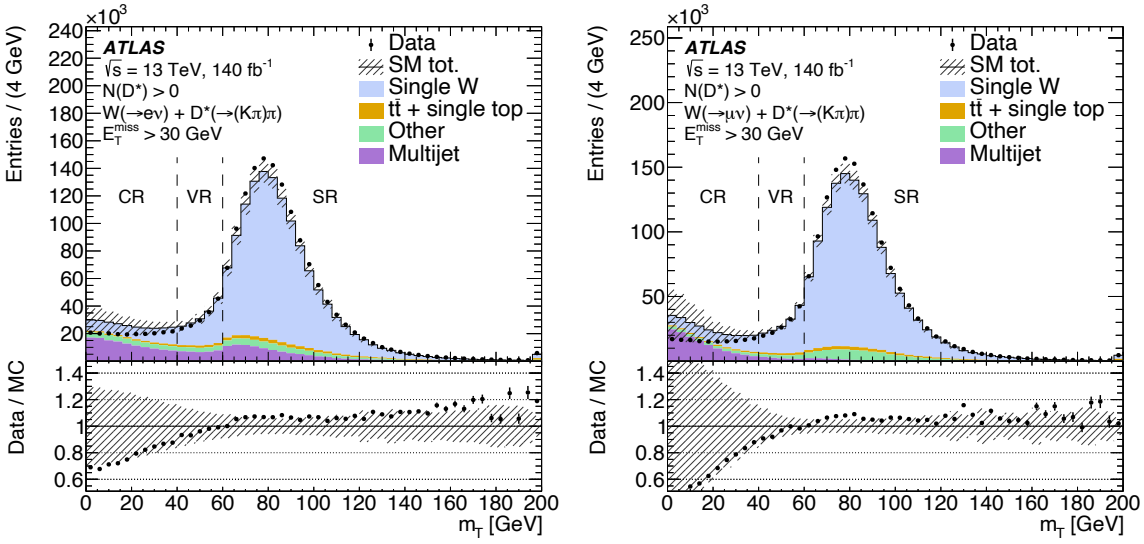
7.2 $Z+D^{(*)}$ Templates

MC truth information is used to categorize the MC $Z+D^{(*)}$ events according to the origin of the tracks used to reconstruct the $D^{(*)}$ meson candidate:

- $Z+D^{(*)}$ signal: If all tracks originate from the signal charmed hadron species (D^+ or D^*) and are assigned in the reconstruction to the correct particle species ($K^{\mp}\pi^{\pm}\pi^{\pm}$), then that reconstructed $D^{(*)}$ candidate is labeled as $Z+D^{(*)}$ signal.

- $Z+c^{\mathrm{match}}$: If all tracks originate either from a different charmed hadron species (D^0, D_s, or c-baryon) or from a different decay mode of a signal charmed meson (e.g. $D^+ \to \phi\pi^+ \to (K^+K^-)\pi^+$), the reconstructed $D^{(*)}$ candidate is labeled as $Z+c^{\mathrm{match}}$.

- $Z+c^{\text{mis-match}}$: If at least one but not all tracks belong to a single charmed hadron, the reconstructed $D^{(*)}$ candidate is labeled as $Z+c^{\text{mis-match}}$.



(a) (b)

Figure 7.1: Modeling distributions of the m_{T} variable using the Matrix Method to estimate the multijet background. The distributions are (a) m_{T} in the D^{*+} electron channel, (b) m_{T} in the D^{*+} muon channel. The uncertainty band includes all Matrix Method systematic uncertainties, $E_{\mathrm{T}}^{\mathrm{miss}}$ systematic uncertainties, and QCD scale variations. The "Single W" component includes all contributions from Table 7.1. Dashed vertical lines indicate the m_{T} values defining the control, validation, and signal regions (CR, VR, and SR) as explained in the text. The last bin also includes the events with $m_{\mathrm{T}} > 200$ GeV.

- Z+jets: if none of the tracks are matched to a particle originating from a charmed particle, the $D^{(*)}$ candidate is labeled Z+jets. This is the combinatorial background from the underlying event and pileup.

Additional background categories modeled using MC simulation are:

- Top: Processes containing top quarks ($t\bar{t}$, single-t, $t\bar{t}X$) are jointly represented by the "Top" category, which is dominated by the $t\bar{t}$ process.
- Diboson: Events from diboson processes.

The signal and background samples used in the Z+D^{+} and Z+D^{*+} fits are given in Table 7.2. The rates at which c-quarks hadronize into different species of weakly decaying charmed hadrons in the MC samples are re-weighted to the world-average values [68]. The weights improve agreement between data and MC simulation by modifying the signal and background normalizations and the shapes of the Z+$D^{(*)}$ background templates by changing

the relative contribution of each species. The normalization of the background templates changes by up to 3%, depending on the $D^{(*)}$ species.

Table 7.2: Single-Z-boson MC samples employed to create mass templates used in the Z+$D^{(*)}$ fits. The "Normalization" and "Shape" columns indicate the source used to calculate the corresponding property. $m(D^{(*)})$ stands for $m(D^+)$ in the D^+ channel and $m(D^{*+}-D^0)$ in the D^* channel. The MC configurations used to model these backgrounds are described in Chapter 3.2. Preferentially, MG_aMC@NLO+Py8 (FxFx) samples are used for signal and background modeling.

Category	Normalization	$m(D^{(*)})$ shape
Z+$D^{(*)}$ (D^+ channel)	MG_aMC@NLO+Py8 (FxFx)	
Z+$D^{(*)}$ (D^* channel)	MG_aMC@NLO+Py8 (FxFx)	
Z+c^{match} (D^+ channel)	MG_aMC@NLO+Py8 (FxFx)	
Z+c^{match} (D^* channel)	MG_aMC@NLO+Py8 (FxFx)	
Z+$c^{\text{mis-match}}$	MG_aMC@NLO+Py8 (FxFx)	
Z+jets	MG_aMC@NLO+Py8 (FxFx)	

7.2.1 Signal Modeling

The MG_aMC@NLO+Py8 (FxFx) Z+D^+ signal sample with EvtGen decays is used for the modeling of the mass template both the D^+ and D^* channel. In both channels the normalization is taken from MG_aMC@NLO+Py8 (FxFx) because it provides the best available statistical power for calculating the fiducial efficiency.

7.2.2 Modeling Backgrounds with Prompt Monte Carlo

The Z+c^{match} background in the D^+ channel is modeled using MG_aMC@NLO+Py8 (FxFx) because the EvtGen decay tables and models used with MG_aMC@NLO+Py8 (FxFx) provide a better description of the D meson decay rates and kinematics than those implemented in SHERPA (2.2.11). MG_aMC@NLO+Py8 (FxFx) is used in the D^{*+} channel.

The Z+$c^{\text{mis-match}}$ and Z+jets backgrounds are modeled using MG_aMC@NLO+Py8 (FxFx) in both the D^+ and D^{*+} channels.

7.2.3 Data-Driven Multijet Background Estimation

Due to the requirement of two `tight` leptons in the definition of the Z+$D^{(*)}$ SR, the multijet background in this analysis is expected to be small. MC-based predictions for the

normalization and composition of these backgrounds suffer from large uncertainties due to difficulty in modeling this background. The Matrix Method is not used to estimate this background for this analysis due to smaller statistics leading to much larger uncertainties. Instead the data-driven "ABCD" Method is used which splits the $Z+D^{(*)}$ SR and Fake CR through the inversion of two selections: isolation and charge quality of the pair of leptons (OS or SS). These four sections make up the "ABCD" regions and are shown in Table 7.3.

Table 7.3: The definition of the $Z+D^{(*)}$ SR and Fake CRs used in the multijet estimation.

Region	Opposite Sign	Same Sign
isolated	A	B
non-isolated	C	D

This method relies on the concept that the make-up of the physics processes contributing to the multijet background are identical for both OS and SS events. Similar to the Matrix Method, inverting isolation requirements increases the number of fakes in a region, increasing the statistics necessary to measure the multijet background. Therefore the assumption is made that the ratio of events in region A and B are the same as the ratio of events in C and D:

$$\frac{N_A}{N_B} = \frac{N_C}{N_D}.$$

The yield of multijet in each region is calculated by subtracting out the prompt MC from data. The total multijet yield in the $Z+D^{(*)}$ SR can then be calculated by solving for region A:

$$N_A^{qcd} = N_B \cdot \frac{N_C}{N_D}.$$

The B and D control regions are constructed by requiring exactly two same-sign leptons in an event. The C and D control regions are constructed by implementing no isolation requirement on the two leptons. In order to increase statistics in the Fake CRs (B,C,D), a wider mass window was used for the Z boson. While the analysis requires the reconstructed Z mass to fall within 76 and 106 GeV (Chapter 6.2.1), our Fake CRs require the Z mass to fall within 40 and 160 Gev. The final multijet estimation requires a further scaling based on the events that pass both the nominal selection and wider mass window selection.

Using the ABCD method, the overall amount of multijet background in our $Z+D^{(*)}$ SR is at the sub-percent level and negligible in this analysis. It is therefore not included in our fit and extraction of cross-sections (Chapter 8).

Chapter 8

Cross-Section Fit and Extraction

This chapter explains the fitting strategy and procedure used for both analyses. A statistical fitting procedure based on the standard profile-likelihood formalism used in LHC experiments [127, 128] is used to extract the observables from the data with corresponding uncertainties:

- W+$D^{(*)}$ Analysis
  - absolute fiducial cross-sections: $\sigma_{\text{fid}}^{\text{OS-SS}}(W^-+D^{(*)})$ and $\sigma_{\text{fid}}^{\text{OS-SS}}(W^++D^{(*)})$,
  - the cross-section ratio: $R_c^{\pm} = \sigma_{\text{fid}}^{\text{OS-SS}}(W^++D^{(*)})/\sigma_{\text{fid}}^{\text{OS-SS}}(W^-+D^{(*)})$,
  - differential cross-sections for OS-SS $W^-+D^{(*)}$ and $W^++D^{(*)}$.
- Z+$D^{(*)}$ Analysis
  - absolute fiducial cross-section: $\sigma_{\text{fid}}(Z+D^+)$,
  - differential cross-sections for Z+$D^{(*)}$.

The likelihood fit enables the estimation of background normalization and constraining of systematic uncertainties in situ by extracting the information from the data in mass peak sidebands and control regions. It is a crucial ingredient in achieving percent-level precision in the V+$D^{(*)}$ cross-section measurements. The formalism of the profile likelihood fit is given in Chapters 8.1.1 and 8.2.1, Chapter 8.1.2 explains how the "OS-SS " subtraction is incorporated, Chapters 8.1.3 and 8.2.2 introduce the measurement of normalized differential cross-sections, and Chapters 8.1.4 and 8.2.3 define the bin edges of the measured differential variables.

8.1 W+$D^{(*)}$ Fit Configuration

8.1.1 The profile likelihood fit

A binned likelihood function, $\mathcal{L}(\vec{\sigma},\vec{\theta})$, is constructed as the product of Poisson probability terms for each bin of the input mass distributions dependent on the number of data events and the expected signal and background yields. The product over the mass bins is performed for each differential bin, in bins of either $p_T(D^{(*)})$ or $|\eta(\ell)|$. The invariant mass difference ($m(D^{*+}-D^0)$) bins used for the D^* channel fit are narrower in the peak region (with about 8 bins) and wider in the tails, where the shape is more uniform (up to 4 bins). Only a single bin is fit in each Top CR. The integrated W+$D^{(*)}$ SR invariant mass difference distributions are shown in Figure 10.1 in Chapter 10. The impact of systematic uncertainties is included via nuisance parameters, $\vec{\theta}$. These likelihood fits are performed separately for the $p_T(D^{(*)})$ and $|\eta(\ell)|$ distributions. A likelihood equation describing this fitting procedure is given in Equations (8.1)-(8.4):

$$\mathcal{L}(\vec{\sigma},\vec{\theta}) = \prod_{\alpha}\left(\prod_{i}^{W^-\,\mathrm{OS}}\mathcal{L}(\vec{\sigma},\vec{\theta})_i^{\alpha\,\mathrm{OS}} \times \prod_{i}^{W^-\,\mathrm{SS}}\mathcal{L}(\vec{\theta})_i^{\alpha\,\mathrm{SS}} \times \prod_{i}^{W^+\,\mathrm{OS}}\mathcal{L}(\vec{\sigma},\vec{\theta})_i^{\alpha\,\mathrm{OS}} \times \prod_{i}^{W^+\,\mathrm{SS}}\mathcal{L}(\vec{\theta})_i^{\alpha\,\mathrm{SS}}\right) \times \mathcal{L}^{\mathrm{constr.}}, \tag{8.1}$$

$$\mathcal{L}(\vec{\sigma},\vec{\theta})_i^{\alpha\,\mathrm{OS}} = f\left(N_i^{\alpha}|\gamma_i^{\alpha}\cdot\left(\sum_{\beta}\left[\sigma_{\mathrm{fid}}^{\beta}\cdot r^{\alpha\beta}(\vec{\theta})\cdot\mathscr{P}_i^{\alpha\beta}(\vec{\theta})\right]\cdot\mathscr{L}(\theta_{\mathrm{lumi}})\cdot B_{D^{(*)}}+\mathscr{B}_i^{\alpha}(\vec{\theta},\mu_{\mathrm{Top}})\right) + \mathscr{C}_i^{\alpha}\right), \tag{8.2}$$

$$\mathcal{L}(\vec{\theta})_i^{\alpha\,\mathrm{SS}} = f\big(N_i^{\alpha}|\gamma_i^{\alpha}\cdot\mathscr{B}_i^{\alpha}(\vec{\theta},\mu_{\mathrm{Top}})+\mathscr{C}_i^{\alpha}\big), \tag{8.3}$$

$$\mathcal{L}^{\mathrm{constr.}} = \prod_{t} g(\theta_t)\times\prod_{\alpha,\,i} f(\gamma_i^{\alpha}), \tag{8.4}$$

where the index i represents the bins of the $D^{(*)}$ mass distribution (either OS or SS) both in the W+$D^{(*)}$ SR ($n_{\mathrm{bjet}}=0$), and the single bin used in the Top CR ($n_{\mathrm{bjet}}>0$). Indices α and β represent the detector-level and truth differential bins respectively, and the index t represents the nuisance parameters $\vec{\theta}$. The expression $f(k|\lambda)=\lambda^k e^{-\lambda}/k!$ is the Poisson probability density function. Furthermore,

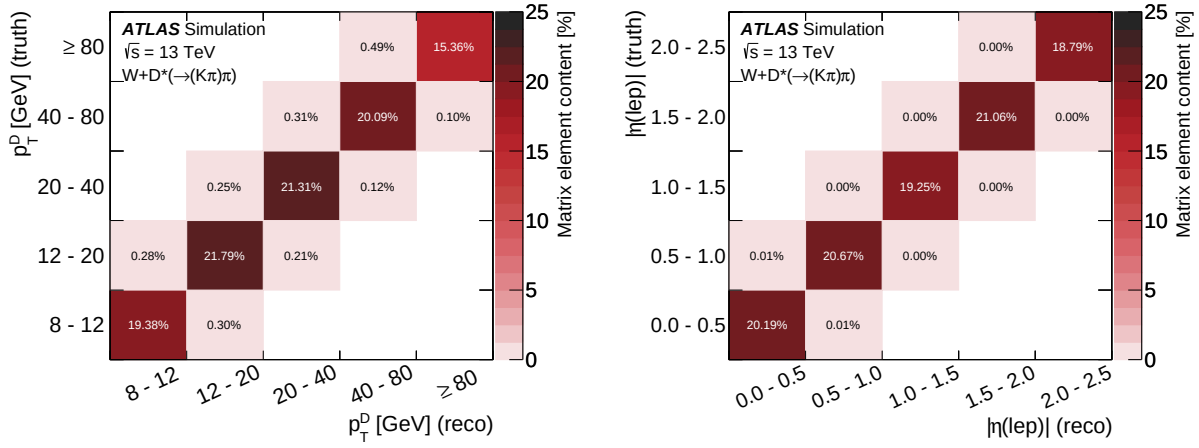
- N_i^α is the number of observed events in mass bin i and reconstructed differential bin α,
- $\sigma_{\text{fid}}^\beta$ is the fiducial cross-section in differential bin β (one parameter per differential bin and W boson charge),
- $r^{\alpha\beta}(\vec{\theta})$ is the detector response matrix, defined as the fraction of $W{+}D^{(*)}$ events produced in truth fiducial bin β that also satisfy the $W{+}D^{(*)}$ SR reconstruction criteria in bin α,
- $\mathscr{P}_i^{\alpha\beta}(\vec{\theta})$ is the i-th bin of the mass shape distribution of the signal sample corresponding to truth differential bin β in reconstructed differential bin α (a separate invariant mass distribution for every non-zero bin in Figure 8.1)
- $\mathscr{L}(\theta_{\text{lumi}})$ is the integrated luminosity,
- $B_{D^{(*)}}$ is the branching ratio of the D^{*+} decaying into $K\pi\pi$ (Ref. [125]),
- $\mathscr{B}_i^\alpha(\vec{\theta}, \mu_{\text{Top}})$ is the total number of background events in mass bin i and reconstructed differential bin α, including the $W{+}D^{(*)}$ signal events failing the truth fiducial selection (Table 6.1b),
- μ_{Top} is the normalization factor for the top quark background,
- $\mathscr{C}_i^\alpha$ is the "common floating component" in mass bin i and reconstructed differential bin α (mathematical construct to enable likelihood minimization in OS-SS, described further in Chapter 8.1.2),
- $\vec{\theta}$ represents all nuisance parameters that are profiled in the likelihood fit,
- γ_i^α parameters are the Poisson-constrained parameters accounting for the MC statistical uncertainties in the combined signal-plus-background mass templates, following the simplified Beeston–Barlow technique [129].

The signal sample is distributed among the detector-level differential bins using the detector response matrix. The detector response matrix $r^{\alpha\beta}(\vec{\theta})$ is defined as the fraction of $W{+}D^{(*)}$ events produced in truth fiducial bin β that also satisfy the $W{+}D^{(*)}$ SR reconstruction criteria in bin α. The values of the response matrix depend on the nuisance parameters $\vec{\theta}$ representing the systematic uncertainties. Signal $W{+}D^{(*)}$ events failing the truth fiducial selection given in Table 6.1b are treated as background events in the likelihood fit.

The nuisance parameters $\vec{\theta}$ have Gaussian constraints $g(\theta)$ in the likelihood with a mean of 0 and a standard deviation that corresponds to the one-standard-deviation variations of the associated systematic uncertainties, determined from auxiliary measurements (e.g. lepton calibration described in Chapter 4). The γ_i^α parameters are centered around 1 and may

deviate from unity within the corresponding Poisson constraints reflecting the combined signal-plus-background statistical uncertainty in the invariant mass templates. The post-fit (profiled) uncertainty is determined by performing a likelihood scan for each nuisance parameter individually and taking the 68% CL interval around the best-fit value of the nuisance parameter as the uncertainty. None of the nuisance parameters that have a large impact on the observables deviate significantly from their initial values and the post-fit uncertainty remains close to unity as shown in Chapter 10.

Response matrices for the D^{*+} channel is shown in Figure 8.1 for differential $p_T(D^{(*)})$ and $|\eta(\ell)|$ bins for nominal values of the nuisance parameters. Differential cross-sections extracted in this way correspond to unfolding with matrix inversion. No regularization techniques are used because the detector response matrices are nearly diagonal and because the statistical uncertainties are sufficiently low.



(a) (b)

Figure 8.1: The W+$D^{(*)}$ detector response matrix in differential $p_T(D^{(*)})$ bins: (a) W+D^{*+}, and in differential $|\eta(\ell)|$ bins: (b) W+D^{*+}. The detector response matrix is calculated with SHERPA (2.2.11) W+$D^{(*)}$ samples. The detector response matrices are normalized to unity such that the sum of all elements is 100 %. The last $p_T(D^{(*)})$ bin has an upper cut of 150 GeV at the detector level, while there is no upper cut at the truth level.

8.1.2 The OS-SS subtraction

A fitting procedure exploiting the charge correlation between the W boson and the $D^{(*)}$ meson was developed to perform the OS-SS subtraction within the likelihood fit. Instead of using OS-SS distributions in the fit, both the OS and SS regions enter the likelihood function and a common floating component is added in both regions. The additional component

has one free parameter per invariant mass bin. This parameter is correlated between the corresponding OS and SS regions. The common floating component is configured to absorb all charge-symmetric processes, which effectively translates the maximization of separate OS and SS likelihoods into a maximization of the OS-SS likelihood. This is done because the OS-SS event yields do not follow the Poisson distributions, which is a requirement for the data yields in the profile likelihood fit. Furthermore, this fitting procedure ensures that the yields of the individual signal and background components remain positive in the fit even though their OS-SS difference could be negative.

The method used to extract the OS-SS $\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W+D^{(*)})$ cross-section from a simultaneous fit to OS and SS regions with the common floating component is demonstrated in Figure 8.2 for the second bin of the $p_T(D^{(*)})$ distribution in the D^+ channel as the working example. The pre-fit OS, SS, and OS-SS distributions are shown at the left-hand side of Figure 8.2 and the corresponding post-fit distributions are at the right-hand side. The $W+D^+$ signal sample is split into three components (labelled bin 1, bin 2 and bin 3), which corresponding to the diagonal and two off-diagonal elements immediately above and beneath the diagonal in Figure 8.1. Since all other non-diagonal elements are zero, signal samples corresponding to truth fiducial bins 4 and 5 are not included. The common floating component is shown with the gray histograms named "Ch. Symm." in the legend. The initial pre-fit values of the common floating component are arbitrary because every bin has a corresponding free parameter in the fit. This component is merely a mathematical construct to translate the minimization of separate OS and SS negative log likelihoods into a minimization in OS-SS. The initial values in both the OS and SS regions are set to the difference between the data and the MC prediction in the SS region (different results were not observed with other initial values). The plots illustrate the effectiveness of the OS-SS subtraction; the backgrounds are almost symmetric in OS and SS regions, so the resulting OS-SS distributions are largely dominated by the $W+D^{(*)}$ signal. The "SM Tot." line represents the sum of all signal and background samples. The corresponding pre-fit uncertainty bands include MC statistical uncertainties only and the post-fit uncertainty bands include the total uncertainty extracted from the fit. The gray histograms represent the charge-symmetric common floating component and the three histograms associated with the signal samples are the truth bins of the $p_T(D^{(*)})$ differential distribution.

8.1.3 Normalized differential cross-section

Normalized differential cross-sections are generally more powerful than absolute differential cross-sections in distinguishing between the observed data and the theory predictions since overall systematic uncertainties such as those in the integrated luminosity and branching ratio cancel out in the normalized differential cross-sections. To extract the normalized differential cross-sections and the corresponding uncertainties, the fit is performed to extract the four normalized cross-sections and the total fiducial cross-section, $\sigma_{\text{fid}}^{\text{tot.}}$, instead of extracting the five absolute differential cross-sections. This is done because by definition, the

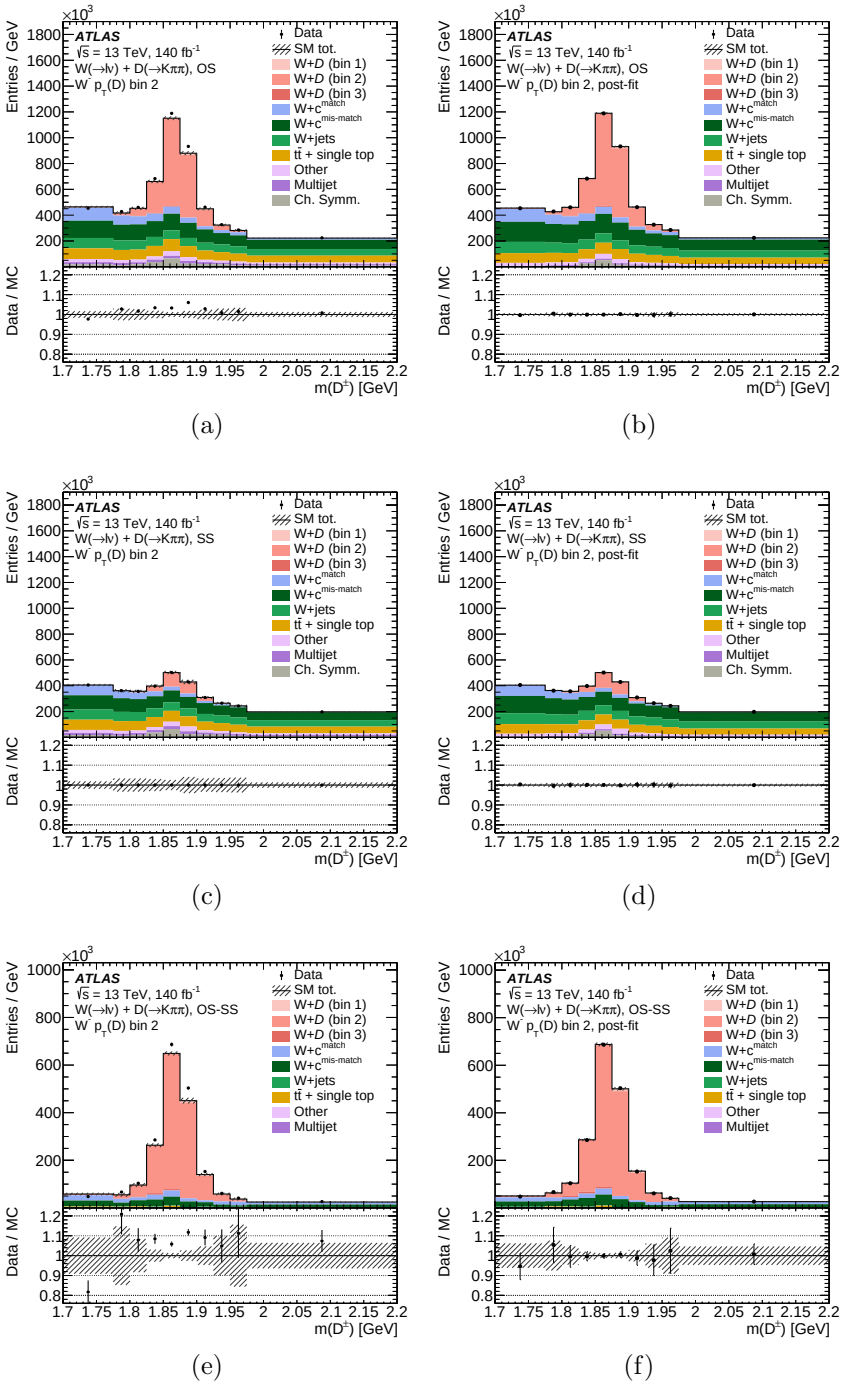


Figure 8.2: A demonstration of the OS-SS W+$D^{(*)}$ cross-section fit. Pre-fit $m(D^+)$ distributions for the W^-+D^+ $p_T(D^+)$ bin 2: (a) OS, (c) SS, and (e) OS-SS. The corresponding post-fit distributions: (b) OS, (d) SS, and (f) OS-SS.

sum of all normalized differential cross-sections adds up to the total fiducial cross-section, introducing a degeneracy in the fits. A substitution of the free parameters in the likelihood fit is made as shown in Equation (8.5):

$$\begin{aligned} &\sigma^1_{\text{fid}} \rightarrow \sigma^{\text{tot.}}_{\text{fid}} \times \sigma^1_{\text{rel}}, \\ &\sigma^2_{\text{fid}} \rightarrow \sigma^{\text{tot.}}_{\text{fid}} \times \sigma^2_{\text{rel}}, \\ &\cdots \\ &\sigma^N_{\text{fid}} \rightarrow \sigma^{\text{tot.}}_{\text{fid}} \times \left[1 - \sum_{i=1}^{N-1} \sigma^i_{\text{rel}}\right], \end{aligned} \tag{8.5}$$

where σ^i_{fid} is the absolute fiducial cross-section in truth differential bin i and σ^i_{rel} is the corresponding normalized differential cross-section. The value of N is five in all cases and the sum of all normalized differential cross-sections is defined to be one. This substitution is performed separately for each charge, W^++$D^{(*)}$ and W^-+$D^{(*)}$. Furthermore, a similar substitution is made for the $R_c^{\pm}$ parameter. The normalization factor for the $\sigma^{\text{OS}-\text{SS}}_{\text{fid}}(W^+$+$D^{(*)})$ total fiducial cross-section is replaced by the expression shown in Equation (8.6):

$$\sigma^{\text{tot.}}_{\text{fid}}(W^+\text{+}D^{(*)}) \rightarrow R_c^{\pm} \times \sigma^{\text{tot.}}_{\text{fid}}(W^-\text{+}D^{(*)}), \tag{8.6}$$

The free parameters in the fit after these substitutions are $\sigma^1_{\text{rel}}, \ldots, \sigma^{N-1}_{\text{rel}}$ for each charge (8 parameters in total), $R_c^{\pm}$, and $\sigma^{\text{tot.}}_{\text{fid}}(W^-$+$D^{(*)})$. The central values of all additional observables can be deduced from these free parameters; however, systematic uncertainties can only be calculated for the parameters directly included in the fit (with a likelihood scan). To achieve this, several fits with different substitutions of parameters are performed:

1. $\sigma^1(W^-$+$D^{(*)})$, …, $\sigma^N(W^-$+$D^{(*)})$, $\sigma^1(W^+$+$D^{(*)})$, …, $\sigma^N(W^+$+$D^{(*)})$,
2. $\sigma^1_{\text{rel}}(W^-$+$D^{(*)})$, …, $\sigma^{N-1}_{\text{rel}}(W^-$+$D^{(*)})$, $\sigma^1_{\text{rel}}(W^+$+$D^{(*)})$, …, $\sigma^{N-1}_{\text{rel}}(W^+$+$D^{(*)})$, $R_c^{\pm}$, $\sigma^{\text{tot.}}_{\text{fid}}(W^-$+$D^{(*)})$,
3. $\sigma^2_{\text{rel}}(W^-$+$D^{(*)})$, …, $\sigma^N_{\text{rel}}(W^-$+$D^{(*)})$, $\sigma^2_{\text{rel}}(W^+$+$D^{(*)})$, …, $\sigma^N_{\text{rel}}(W^+$+$D^{(*)})$, $\sigma^{\text{tot.}}_{\text{fid}}(W^+$+$D^{(*)})$, $R_c^{\pm}$.

These three fits allow a precise determination of the central values and systematic uncertainties of all observables, including absolute and normalized differential cross-sections. In all cases the number of free parameters is the same and the minimization procedure reaches the same minimum, yielding identical results.

8.1.4 Differential cross-section bins

The bin edges of the five differential $p_T(D^{(*)})$ bins are given in Table 8.1. The last bin starts at 80 GeV and has no upper limit. The number of bins and the bin edges were chosen

such that the expected data statistical uncertainty is about 1%–2% in the first four bins. The available MC sample sizes also play an important role in determining the bin size; up to a 1% statistical uncertainty is present in the diagonal elements of the detector response matrix. Similarly to the $p_T(D^{(*)})$ fits, five bins are chosen in $|\eta(\ell)|$ to provide percent-level precision. Furthermore, the absolute value of the pseudorapidity is used to further reduce the statistical uncertainty because there is no additional discriminating power in measuring the sign of the pseudorapidity. The $|\eta(\ell)|$ bin edges are also given in Table 8.1.

With five differential bins per W boson charge there are 10 differential cross-sections in total represented with the free parameters $\vec{\sigma}$ in the likelihood fit. Regions with both charges of the W boson are included in the fit at the same time in order to extract the cross-section ratio $R_c^{\pm}$. SRs in the $n_{\text{bjet}} = 0$ category are split between the two W charges, into OS and SS events and into the five differential bins: $[W^-, W^+] \times [\text{OS, SS}] \times 5 = 20$ regions. The $n_{\text{bjet}} > 0$ CRs are split in the same way with the exception of differential bins since the normalization of the backgrounds from top-quark production is extracted from the data only inclusively. The relative contribution of the top-quark background in the $W+D^{(*)}$ SR is small (negligible for D^*), so the modeling of the differential spectrum in the top-quark background simulation has a negligible impact on the result. The regions used in the fit are summarized in Table 8.2.

Table 8.1: The differential $p_T(D^{(*)})$ and $|\eta(\ell)|$ bins used in the measurement. The last $p_T(D^{(*)})$ bin has no upper limit.

Bin number	1	2	3	4	5
$p_T(D^{(*)})$ bin edges [GeV]	[8, 12]	[12, 20]	[20, 40]	[40, 80]	[80, ∞)
$\|\eta(\ell)\|$ bin edges	[0.0, 0.5]	[0.5, 1.0]	[1.0, 1.5]	[1.5, 2.0]	[2.0, 2.5]

Table 8.2: A schematic of the signal and control regions (SR and CR) used in the fit. The bin numbers correspond to either the $p_T(D^{(*)})$ or $|\eta(\ell)|$ differential bins listed in Table 8.1. The table indicates that the invariant mass difference distribution is fit in each $W+D^{(*)}$ SR, while only a single bin is fit in the Top CR.

	$W+D^{(*)}$ SR ($n_{\text{bjet}} = 0$)				Top CR ($n_{\text{bjet}} > 0$)			
W charge	W^-		W^+		W^-		W^+	
$D^{(*)}$ charge	OS	SS	OS	SS	OS	SS	OS	SS
Bin 1	Fit the $m(D^{*+}-D^0)$ distribution				Fit total yield			
Bin 2								
Bin 3								
Bin 4								
Bin 5								

8.2 $Z+D^{(*)}$ Fit Configuration

8.2.1 The profile likelihood fit

A binned likelihood function, $\mathcal{L}(\vec{\sigma},\vec{\theta})$, is constructed as the product of Poisson probability terms for each bin of the input mass distributions, based on the number of data events and the expected signal and background yields. The product over the mass bins is performed for each differential bin, in bins of $p_T(D^{(*)})$. The reconstructed invariant mass of the D^+ meson, $m(D^+)$, is used as input in the D^+ channel and the mass difference $m(D^{*+}-D^0)$ is used in the D^* channel because it has better resolution than the D^* invariant mass. The invariant mass bins in the $Z+D^{(*)}$ SR are narrower in the peak region (with about 8 bins) and wider in the tails, where the shape is more uniform (up to 4 bins). The integrated $Z+D^{(*)}$ SR invariant mass distributions are shown in Figure 10.6 in Chapter 10. The impact of systematic uncertainties is included via nuisance parameters, $\vec{\theta}$. Separate likelihood fits are performed for the D^+ and D^{*+} channels. A likelihood equation describing this fitting procedure is given in Equations (8.7)-(8.9):

$$\mathcal{L}(\vec{\sigma},\vec{\theta}) = \prod_{\alpha}\prod_{i}\mathcal{L}(\vec{\sigma},\vec{\theta})_i^{\alpha} \times \mathcal{L}^{\text{constr.}}, \tag{8.7}$$

$$\mathcal{L}(\vec{\sigma},\vec{\theta})_i^{\alpha} = f\left(N_i^{\alpha}|\gamma_i^{\alpha}\cdot\left(\sum_{\beta}\left[\sigma_{\text{fid}}^{\beta}\cdot r^{\alpha\beta}(\vec{\theta})\cdot\mathscr{P}_i^{\alpha\beta}(\vec{\theta})\right]\cdot\mathscr{L}(\theta_{\text{lumi}})\cdot B_{D^{(*)}}+\mathscr{B}_i^{\alpha}(\vec{\theta})\right)\right), \tag{8.8}$$

$$\mathcal{L}^{\text{constr.}} = \prod_{t} g(\theta_t)\times\prod_{\alpha,i} f(\gamma_i^{\alpha}), \tag{8.9}$$

where the index i represents the bins of the $D^{(*)}$ mass distribution. Indices α and β represent the detector-level and truth differential bins respectively, and the index t represents the nuisance parameters $\vec{\theta}$. The expression $f(k|\lambda) = \lambda^k e^{-\lambda}/k!$ is the Poisson probability density function. Furthermore,

- N_i^{α} is the number of observed events in mass bin i and reconstructed differential bin α,
- $\sigma_{\text{fid}}^{\beta}$ is the fiducial cross-section in differential bin β (one parameter per differential bin),
- $r^{\alpha\beta}(\vec{\theta})$ is the detector response matrix, defined as the fraction of $Z+D^{(*)}$ events produced in truth fiducial bin β that also satisfy the $Z+D^{(*)}$ SR reconstruction criteria in bin α,
- $\mathscr{P}_i^{\alpha\beta}(\vec{\theta})$ is the i-th bin of the mass shape distribution of the signal sample corresponding to truth differential bin β in reconstructed differential bin α (a separate invariant mass distribution for every non-zero bin in Figure 8.3)

- $\mathscr{L}(\theta_{\text{lumi}})$ is the integrated luminosity,
- $B_{D^{(*)}}$ is the branching ratio of either the D^+ or D^{*+} decaying into $K\pi\pi$ (Ref. [125]),
- $\mathscr{B}_i^{\alpha}(\vec{\theta})$ is the total number of background events in mass bin i and reconstructed differential bin α, including the $Z{+}D^{(*)}$ signal events failing the truth fiducial selection (Table 6.2b),
- $\vec{\theta}$ represents all nuisance parameters that are profiled in the likelihood fit,
- γ_i^{α} parameters are the Poisson-constrained parameters accounting for the MC statistical uncertainties in the combined signal-plus-background mass templates, following the simplified Beeston–Barlow technique [129].

The signal sample is distributed among the detector-level differential bins using the detector response matrix. The detector response matrix $r^{\alpha\beta}(\vec{\theta})$ is defined as the fraction of $Z{+}D^{(*)}$ events produced in truth fiducial bin β that also satisfy the $Z{+}D^{(*)}$ SR reconstruction criteria in bin α. The values of the response matrix depend on the nuisance parameters $\vec{\theta}$ representing the systematic uncertainties.

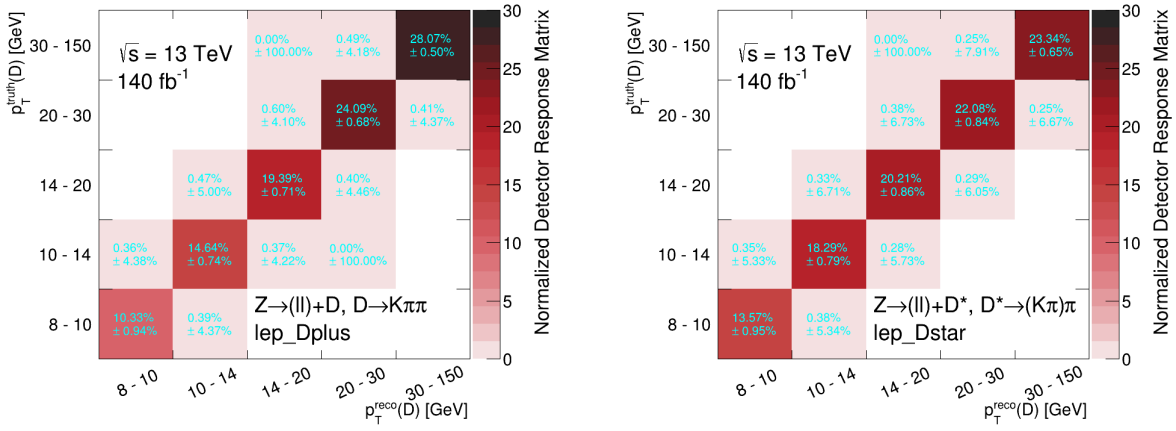
Signal $Z{+}D^{(*)}$ events failing the truth fiducial selection given in Table 6.2b are treated as background events in the likelihood fit.

The nuisance parameters $\vec{\theta}$ have Gaussian constraints $g(\theta)$ in the likelihood with a mean of 0 and a standard deviation that corresponds to the one-standard-deviation variations of the associated systematic uncertainties, determined from auxiliary measurements (e.g. lepton calibration described in Chapter 4). The γ_i^{α} parameters are centered around 1 and may deviate from unity within the corresponding Poisson constraints reflecting the combined signal-plus-background statistical uncertainty in the invariant mass templates. The post-fit (profiled) uncertainty is determined by performing a likelihood scan for each nuisance parameter individually and taking the 68% CL interval around the best-fit value of the nuisance parameter as the uncertainty. None of the nuisance parameters that have a large impact on the observables deviate significantly from their initial values and the post-fit uncertainty remains close to unity as shown in Chapter 10.

Response matrices for the D^+ and D^{*+} channels are shown in Figure 8.3 for differential $p_T(D^{(*)})$ bins for nominal values of the nuisance parameters. Differential cross-sections extracted in this way correspond to unfolding with matrix inversion. No regularization techniques were used because the detector response matrices are nearly diagonal and because the statistical uncertainties are sufficiently low.

8.2.2 Normalized differential cross-section

Normalized differential cross-sections are generally more powerful than absolute differential cross-sections in distinguishing between the observed data and the theory predictions



(a) (b)

Figure 8.3: The Z+$D^{(*)}$ detector response matrix in differential $p_T(D^{(*)})$ bins: (a) Z+D^{+}, (b) Z+D^{*+}. The detector response matrix is calculated with MG_aMC@NLO+Py8 (FxFx) samples. The detector response matrices are normalized to unity such that the sum of all elements is 100 %.

since overall systematic uncertainties such as those in the integrated luminosity and branching ratio cancel out in the normalized differential cross-sections. To extract the normalized differential cross-sections and the corresponding uncertainties, the fit is performed to extract the four normalized cross-sections and the total fiducial cross-section, $\sigma_{\text{fid}}^{\text{tot.}}$, instead of extracting the five absolute differential cross-sections. By default, a substitution of the free parameters in the likelihood fit is made as shown in Equation (8.10):

$$\begin{aligned} \sigma_{\text{fid}}^{1} &\rightarrow \sigma_{\text{fid}}^{\text{tot.}} \times \sigma_{\text{rel}}^{1}, \\ \sigma_{\text{fid}}^{2} &\rightarrow \sigma_{\text{fid}}^{\text{tot.}} \times \sigma_{\text{rel}}^{2}, \\ &\cdots \\ \sigma_{\text{fid}}^{N} &\rightarrow \sigma_{\text{fid}}^{\text{tot.}} \times \left[1 - \sum_{i=1}^{N-1} \sigma_{\text{rel}}^{i}\right], \end{aligned} \tag{8.10}$$

where σ_{fid}^{i} is the absolute fiducial cross-section in truth differential bin i and σ_{rel}^{i} is the corresponding normalized differential cross-section. The value of N is five in all cases. By definition, the sum of all normalized differential cross-sections is one.

The free parameters in the fit after these substitutions are $\sigma_{\text{rel}}^{1}, \ldots, \sigma_{\text{rel}}^{N-1}$ for (4 parameters in total) and $\sigma_{\text{fid}}^{\text{tot.}}(Z$+$D^{(*)})$. The central values of all additional observables can be deduced from these free parameters; however, systematic uncertainties can only be calculated for the

parameters directly included in the fit (with a likelihood scan).To achieve this, several fits with different substitutions of parameters are performed:

1. $\sigma^1(Z{+}D^{(*)}), \ldots, \sigma^N(Z{+}D^{(*)})$
2. $\sigma^1_{\text{rel}}(Z{+}D^{(*)}), \ldots, \sigma^{N-1}_{\text{rel}}(Z{+}D^{(*)}), \sigma^{\text{tot.}}_{\text{fid}}(Z{+}D^{(*)})$
3. $\sigma^2_{\text{rel}}(Z{+}D^{(*)}), \ldots, \sigma^N_{\text{rel}}(Z{+}D^{(*)}), \sigma^{\text{tot.}}_{\text{fid}}(Z{+}D^{(*)})$.

These three fits allow a precise determination of the central values and systematic uncertainties of all observables, including absolute and normalized differential cross-sections. In all cases the number of free parameters is the same and the minimization procedure reaches the same minimum, yielding identical results.

8.2.3 Differential cross-section bins

The bin edges of the five differential $p_T(D^{(*)})$ bins are given in Table 8.3. The number of bins and the bin edges were chosen such that the expected data statistical uncertainty is about 4%–7% in the first four bins. The available MC sample sizes also play an important role in determining the bin size; up to a 1% statistical uncertainty is present in the diagonal elements of the detector response matrix. All off-diagonal elements contribute less than 1 % each to their respective bins.

With five differential bins there are 5 differential cross-sections in total represented with the free parameters $\vec{\sigma}$ in the likelihood fit. The regions used in the fit are summarized in Table 8.4.

Table 8.3: The differential $p_T(D^{(*)})$ bins used in the measurement.

Bin number	1	2	3	4	5
$p_T(D^{(*)})$ bin edges [GeV]	[8, 10]	[10, 14]	[14, 20]	[20, 30]	[30, 150]

Table 8.4: A schematic of the signal region used in the fit. The bin numbers correspond to the $p_T(D^{(*)})$ differential bins listed in Table 8.3. The table indicates that the invariant mass distribution is fit in each Z+$D^{(*)}$ SR, with $m(D^{(*)})$ standing for $m(D^+)$ in the D^+ channel and $m(D^{*+}-D^0)$ in the D^* channel.

Bin	Z+$D^{(*)}$ SR
Bin 1	Fit the $m(D^{(*)})$ distribution
Bin 2	
Bin 3	
Bin 4	
Bin 5	

Chapter 9

Systematic Uncertainties

This chapter describes all the systematic uncertainties present in both analyses and used in the likelihood fits. The measurements in both $Z+D^{(*)}$ and $W+D^{(*)}$ analyses are affected by several sources of systematic uncertainty. Uncertainties that affect both analyses are referred to with the inclusive V boson designation. The first category, related to detector-interaction and reconstruction processes, includes uncertainties in lepton and jet reconstruction efficiencies, energy resolution, and energy scale, in lepton identification, isolation, and trigger efficiencies, in b-jet tagging efficiencies, and in the total integrated luminosity and pileup re-weighting. These uncertainties affect the $V+D^{(*)}$ signal efficiencies by altering the detector response matrix, yields of the background processes estimated with MC simulation, and the signal and background invariant mass templates used in the profile likelihood. These uncertainties are correlated between all samples and regions in the likelihood fit and are generally derived from auxiliary measurements.

9.1 $V+D^{(*)}$ Systematic Uncertainties

Charged leptons: Electron and muon reconstruction, isolation, identification, and trigger efficiencies, and the energy/momentum scale and resolution are derived from data using large samples of $J/\psi \to \ell\ell$ and $Z \to \ell\ell$ events [111, 113]. Systematic variations of the MC efficiency corrections and energy/momentum calibrations applied to MC samples are used to estimate the signal selection uncertainties.

Pileup and luminosity: The uncertainty in the integrated luminosity is 0.83% [57], which is obtained using the LUCID-2 detector [58] for the primary luminosity measurements. MC samples are re-weighted to have the number of pileup vertices match the pileup distribution measured in the Run 2 data. To account for the uncertainty in the pileup estimation, variations of the re-weighting are applied to the MC samples. In addition to affecting the background yields, it also has a small impact on the resolution of the reconstructed D^+ meson mass peak and the $m(D^{*+}-D^0)$ mass difference.

Background MC modeling: The implementation of the background modeling uncertainties varies between the backgrounds. For $V+c^{\text{match}}$, $V+c^{\text{mis-match}}$, and $V+$jets backgrounds, SHERPA (2.2.11) QCD scale, PDF, and α_{S} variations are used. Among the three, the QCD scale uncertainty generally has the largest effect and leads to a 10%–30% uncertainty in the yield of the corresponding background process, depending on the differential bin. The uncertainty is constrained in the likelihood fit by the small statistical uncertainties in the tails of the invariant mass distributions in the D^+ and D^{*+} channels, reducing its impact on the observables. As in the case of the signal process, these uncertainties are correlated between the differential bins. An additional modeling uncertainty is included by taking the full difference between SHERPA (2.2.11) and MG_aMC@NLO+Py8 (NLO) or MG_aMC@NLO+Py8 (FxFx) predicted background yields. To be conservative, this uncertainty is taken to be uncorrelated between the differential bins. This avoids the assumption that either of the simulations have an a priori perfect description of the shape of the differential variable (i.e. $p_T(D^{(*)})$ or $|\eta(\ell)|$), and provides more flexibility in the likelihood fit.

Internal event weight variations in the MG_aMC@NLO+Py8 (NLO) 2.3.3 $t\bar{t}$ simulation are used to determine the effect of the PDF uncertainty on the top quark background. The uncertainty due to initial-state radiation is estimated by simultaneously varying the h_{damp} parameter and the μ_r and μ_f scales, and choosing the Var3c up and down variants of the A14 tune as described in Ref. [109]. The impact of final-state radiation is evaluated by halving and doubling the renormalization scale for emissions from the parton shower. Uncertainties in the $t\bar{t}$ ME calculation and PS are estimated by replacing the nominal $t\bar{t}$ prediction with two alternative simulations: PowhegBox+Herwig 7.04 and MadGraph5_aMC@NLO+Pythia8 and taking the full difference as a systematic uncertainty. For other small backgrounds (Z+jets in $W+D^{(*)}$ analysis and diboson events for both analyses) a conservative 20% uncertainty in their yields is used. Due to the high purity of the $V+D^{(*)}$ signal process in the $V+D^{(*)}$ SRs selection, background modeling uncertainties are subdominant in the statistical analysis.

Charm hadronization: The $V+c^{\text{match}}$ and $V+c^{\text{mis-match}}$ backgrounds in the D^+ channel have large contributions from weakly decaying charmed mesons incorrectly reconstructed as $D^+ \to K^-\pi^+\pi^+$ (e.g. $D_s^\pm \to \phi\pi^\pm \to (K^+K^-)\pi^\pm$ or $D_s \to K^*K(K^* \to K\pi)$ reconstructed as $D^+ \to K^-\pi^+\pi^+$). Two sources of associated systematic uncertainty are included: uncertainties in the charmed hadron production fractions and uncertainties in the charmed hadron branching ratios. Charmed hadron production fractions in the MC samples are re-weighted to the world-average values as described in Chapters 7.1.2 and 7.2.2 . Following the procedure in Ref. [68], three eigenvector variations of the event weights are derived to describe the correlated experimental systematic uncertainties associated with the measurements of the charmed hadron production fractions. The uncertainty affects the relative background yield by up to 3% and also the shape of the background invariant mass distribution because the different charmed hadron species populate different ranges of the reconstructed D^+ invariant mass. These uncertainties are evaluated by generating large single-particle samples of D^+

and D^{*+} decays with the same p_T and η distributions as the baseline $W+D^{(*)}$ MC samples. These are named "Single Particle Gun"(SPG) samples. These SPG samples are simulated multiple times with different simulation parameters, mirroring the procedure in Ref. [124]. The impact of the uncertainties in the charmed hadron branching ratios is estimated in a conservative way by generating SPG D^+ samples with all branching ratios shifted simultaneously in a correlated manner to cover the systematic uncertainties in charmed hadron decays reported in Ref.[125]. The relative change in the background yield and shape of the $V+c^{\text{match}}$ background with respect to the nominal SPG configuration is propagated to the SHERPA (2.2.11) MC sample and implemented in the statistical analysis. The size of the uncertainty is up to 5%. Both sources of charmed hadronization uncertainty related to background processes were found to have a negligible impact on all observables.

Finite size of MC samples: The binomial uncertainties in the $V+D^{(*)}$ fiducial efficiencies calculated with the SHERPA (2.2.11) or MG_aMC@NLO+Py8 (FxFx) MC samples are included in the likelihood fit through the use of nuisance parameters which can affect the post-fit yield of the signal and thus the measured cross-section. One parameter is included per non-zero entry of the detector response matrix. The statistical uncertainty in the diagonal elements is less than 1%, while the uncertainty in the off-diagonal elements can exceed 10% depending on the measurement. The size of the off-diagonal elements are very small compared to the diagonal entries. This results in the corresponding statistical uncertainty having a negligible impact. Statistical uncertainties associated with the bins of the invariant mass distributions are implemented as constrained "γ" parameters in the likelihood fit as explained in Chapters 8.1.1 and 8.2.1. One γ parameter is defined per invariant mass bin and their impact on the observables is of the order 2%.

SV reconstruction: Mismodeling on the amount and location of ID material introduces uncertainties in the secondary vertex reconstruction efficiency. This is introduced through the possible mismodeling of hadronic interactions in the GEANT simulations and from possible differences between the impact parameter resolutions in data and MC. These are the simulation parameters modified for the SPG sample production: passive material in the whole ID scaled up by 5%, passive material in the IBL scaled up by 10%, and passive material in the Pixel detector services scaled by 25%. An additional SPG sample where the FTFP_BERT physics model is changed to the QGSP_BIC model was generated [62].

Comparing the baseline and altered material distributions and their corresponding efficiencies is used to construct the uncertainty due to material differences. The change in $D^{(*)}$ reconstruction efficiency is calculated separately for charge and meson mode (D^+ vs. D^*). A parametrization as a function of $p_T(D^{(*)})$ and $\eta(D^{(*)})$ is performed. Physics model choice impact is negligible while the effect of the ID material variation varies by 1%–4%. The uncertainty is largest for low $p_T(D^{(*)})$ and high $\eta(D^{(*)})$. The $t\bar{t}$ background and the signal have distinct $p_T(D^{(*)})$ spectra, thus their tracking efficiency systematic uncertainties are uncorrelated in the fit.

ID material variations are found to affect the shape of the D^+ (D^*) invariant mass ($m(D^{*+}-D^0)$) distributions. The size of the effect is calculated by fitting these distributions

to a double-sided Crystal Ball function [130, 131]. The result of these fits are widths and positions of the peak extracted from the width and mean of the Gaussian contribution to this function. The position of the D^+ (D^{*+}) peak shifted up to 0.2 MeV (0.05 MeV). The peak width is calculated as the difference between the squares of the nominal width and the width obtained from each variation. Up to 4.0 MeV (0.2 MeV) of smearing of the peak width is found for the D^+ (D^{*+}) peak. The peak position and resolution variations are implemented in the likelihood fit as shape uncertainties. This ensures there is no impact on the signal yield.

ID track impact-parameter resolution differences between simulation and data after the ID alignment is performed([132]) introduce an additional systematic uncertainty. Minimum-bias data is used to estimate this uncertainty. This is extrapolated to higher p_T with muon tracks from Z boson decays [124]. It is propagated to the measurements by generating D^+ and D^{*+} SPG samples by performing the SV fit for the $D^{(*)}$ reconstruction where the impact parameters of the ID tracks are smeared beforehand. The $D^{(*)}$ reconstruction efficiency relative change was found to be 1.5%(5%) for low (high-p_T) $D^{(*)}$ mesons. ID material variation and ID track impact-parameter resolution derived systematic uncertainties are some of the largest sources of uncertainty in the analyses.

9.2 $W+D^{(*)}$ Systematic Uncertainties

Jets and missing transverse momentum: Jet energy scale and energy resolution uncertainties affect the signal efficiency and background yields indirectly by altering the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ in $W+D^{(*)}$ events and hence the selection efficiency of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts. Systematic variations of the jet energy calibration are applied to MC samples to estimate signal section uncertainties using the methodology described in Ref. [117]. In total, there are 20 independent jet energy scale variations and 8 independent jet energy resolution variations. None of the single variations have an impact of more than 1% on the signal selection efficiency. Similarly, variations in $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction are derived specifically for the soft-term estimation following the methodology in Ref. [123]. Furthermore, a single nuisance parameter is included to model the uncertainty in the JVT selection efficiency.

Flavor tagging: The uncertainty in the calibration of the b-tagging efficiencies and mis-tag rates is derived from data using samples of di-leptonic $t\bar{t}$ events for b-jets and c-jets [119, 133] and a data sample enriched in light-flavor jets for light-jets [134]. Since the majority of $V+D^{(*)}$ signal events have no additional b-tagged jets, these variations have a negligible impact on the signal efficiency. Nevertheless, the variations in b-tagging efficiency have an impact of up to 10% on the relative yields of the top quark backgrounds in the $W+D^{(*)}$ SR and Top CR for the $W+D^{(*)}$ analysis.

Signal modeling: The signal modeling uncertainty is derived by comparing the fiducial region efficiencies for the signal SHERPA (2.2.11), MG_aMC@NLO+Py8 (FxFx), and MG_aMC@NLO+Py8 (NLO) $W+D^{(*)}$ simulations. In each differential bin, the maxi-

mum difference between the nominal MC simulation (SHERPA (2.2.11)) and either of the MG_aMC@NLO+Py8 (NLO) simulations is taken and a symmetric systematic uncertainty is applied in the two directions. The uncertainty is correlated between the differential bins and W boson charges. It accounts for the fact that the choice of MC simulation for unfolding affects the measured values of the observables because of differences in the ME calculation, PS simulation, and heavy-flavor quark fragmentation and hadronization. The uncertainty ranges from 1% to 4%, depending on the bin, and is generally one of the largest uncertainties in the analysis. The relatively large difference in fiducial efficiency between SHERPA (2.2.11) and MG_aMC@NLO+Py8 (NLO) simulations arises from the modeling of the correlation between W boson and $D^{(*)}$ meson kinematics when the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts are applied at the detector level. Including the same $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts in the truth fiducial definition would reduce the uncertainty; however, it would give rise to a large background from signal W+$D^{(*)}$ events that fail the truth $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} selection, but pass the detector-level selection due to the poor $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, ultimately increasing the total uncertainty.

Additional uncertainties are considered by varying the QCD scales, the PDFs, α_{S}, and the virtual EW corrections in SHERPA (2.2.11). The PDF variations, α_{S} uncertainty, and EW corrections were found to have a negligible effect on the fiducial efficiency. The effect of QCD scale uncertainties is defined by the envelope of variations resulting from changing the renormalization and factorization scales by factors of two with an additional constraint of $0.5 \leq \mu_r/\mu_f \leq 2$. In most differential bins the effect was found to be smaller than the corresponding difference between SHERPA (2.2.11) and MG_aMC@NLO+Py8 (NLO). Lastly, the uncertainties in the $D^+ \to K^-\pi^+\pi^+$ and $D^{*+} \to D^0\pi^+ \to (K^-\pi^+)\pi^+$ branching ratios [125] are applied as uncertainties of 1.7% and 1.1%, respectively, in the signal yield in the likelihood fit.

Multijet estimation: Chapter 7.1.3 describes how the multijet template is estimated as well as its uncertainties. A lack of systematics in the Fake CR as well as a non-trivial dependence of the CR to the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut value used in its construction results in large uncertainties (>50%). The multijet background is a relatively small contribution of the W+$D^{(*)}$ SR in the OS–SS distributions because it is largely symmetric between OS and SS regions. Due to this, the impact of these uncertainties on the measured observables is negligible.

9.3 Z+$D^{(*)}$ Systematic Uncertainties

Signal modeling: The signal modeling uncertainty is derived by comparing the fiducial region efficiencies for the signal SHERPA (2.2.11), MG_aMC@NLO+Py8 (FxFx), Z+$D^{(*)}$ simulations. In each differential bin, the maximum difference between the nominal MC simulation and the alternative simulation is taken and a symmetric systematic uncertainty is applied in the two directions. It accounts for the fact that the choice of MC simulation for unfolding affects the measured values of the observables because of differences in the ME calculation, PS simulation, and heavy-flavor quark fragmentation and hadronization. The

uncertainty ranges from 2% to 5%, depending on the bin, and is generally one of the largest uncertainties in the analysis.

Additional uncertainties are considered by varying the QCD scales and the PDFs in MG_aMC@NLO+Py8 (FxFx). The PDF variations were found to have a negligible effect on the fiducial efficiency. The effect of QCD scale uncertainties is defined by the envelope of variations resulting from changing the renormalization and factorization scales by factors of two with an additional constraint of $0.5 \leq \mu_r/\mu_f \leq 2$. Lastly, the uncertainties in the $D^+ \to K^-\pi^+\pi^+$ and $D^{*+} \to D^0\pi^+ \to (K^-\pi^+)\pi^+$ branching ratios [125] are applied as uncertainties of 1.7% and 1.1%, respectively, in the signal yield in the likelihood fit.

Chapter 10

Results

The results and comparison to theory are presented in this chapter, Chapter 10.1 for the $W+D^{(*)}$ analysis and Chapter 10.2 for the $Z+D^{(*)}$ analysis.

10.1 $W+D^{(*)}$ Results

Post-fit comparisons between the data and MC distributions for the D^{*+} channels are shown in Figure 10.1 separately for the W^- and W^+ channels. Most of the data points are within the resulting 1σ systematic uncertainty band. The $W+D^{(*)}$ SR post-fit yields obtained with the likelihood fit are given in Table 10.1. Yields are shown for both the $p_T(D^{(*)})$ and $|\eta(\ell)|$ fits. The systematic uncertainties in the integrated yields are slightly lower in the $p_T(D^{(*)})$ fits than in the $|\eta(\ell)|$ fits because the dominant systematic uncertainties depend more strongly on $p_T(D^{(*)})$ and are therefore more constrained in the fit.

The resulting cross-sections $\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W+D^{*+}) \times B(W \to \ell\nu)$ and $R_c^{\pm}$ are presented in Table 10.2. The results presented here are obtained using the $p_T(D^{(*)})$ fit; results from the differential $|\eta(\ell)|$ fit are compatible. Ratios of cross-sections obtained in the D^+ and D^{*+} channels are consistent with predictions obtained using the world-average production fractions, $\sigma(W+D^{*+})/\sigma(W+D^+) = 1.010 \pm 0.034$, where the 3.4% uncertainty is obtained using the (correlated) uncertainties in the D^* and D^+ production fractions [67]. A combined value of $R_c^{\pm}(D^{(*)})$ is derived from the individual measurement of $R_c^{\pm}(D^{*+})$. Systematic uncertainties are largely uncorrelated between the channels.

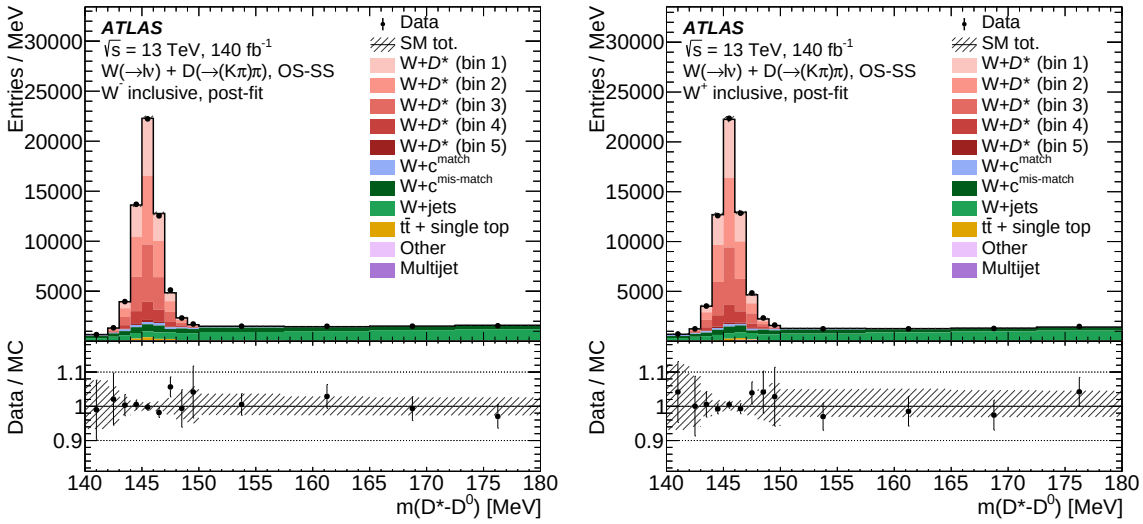
A "Ranking Plot" shows the impact of the post-fit systematic uncertainty pulls and constraints on the total fiducial cross-section for the $p_T(D^{(*)})$ fits. Shown in Figure 10.2 are the 20 nuisance parameters (NPs) that have the most impact on the measured cross-section and its uncertainty. These parameters are ordered by decreasing impact. The γ parameters are the Poisson-constrained parameters accounting for the MC statistical uncertainties in the combined signal-plus-background mass template and are centered around 1 by definition. No large deviations from the initial values are seen on the NPs with the greatest impact on the

Table 10.1: Post-fit yields in the OS-SS W+D^{*+} SR from the $p_T(D^{*+})$ differential fit. The data statistical uncertainty is calculated as $\sqrt{N_{\mathrm{OS}} + N_{\mathrm{SS}}}$. Uncertainties in individual SM components are the full post-fit systematic uncertainties.

	OS-SS W+D^{*+} SR ($p_T(D^{*+})$ fit)		OS-SS W+D^{*+} SR ($\lvert\eta(\ell)\rvert$ fit)	
Sample	W^-+D^{*+}	W^++D^{*-}	W^-+D^{*+}	W^++D^{*-}
$W^{\pm}$+$D^{*\mp}$ (bin 1)	13 670 ± 280	13 880 ± 260	12 640 ± 260	12 980 ± 230
$W^{\pm}$+$D^{*\mp}$ (bin 2)	17 210 ± 250	16 950 ± 280	12 470 ± 260	12 910 ± 280
$W^{\pm}$+$D^{*\mp}$ (bin 3)	15 000 ± 200	14 890 ± 200	10 370 ± 220	10 250 ± 200
$W^{\pm}$+$D^{*\mp}$ (bin 4)	5402 ± 89	5139 ± 95	9500 ± 230	9120 ± 240
$W^{\pm}$+$D^{*\mp}$ (bin 5)	822 ± 45	744 ± 41	6900 ± 290	6390 ± 290
W+c^{match}	2800 ± 530	2730 ± 530	3060 ± 450	2690 ± 480
W+$c^{\text{mis-match}}$	15 900 ± 1700	14 000 ± 1600	16 400 ± 1400	14 200 ± 1400
W+jets	35 600 ± 1800	32 000 ± 1700	35 600 ± 1800	31 900 ± 1700
$t\bar{t}$ + single top	1580 ± 200	1320 ± 180	1480 ± 180	1350 ± 160
Other	1710 ± 540	650 ± 480	1480 ± 480	510 ± 420
Multijet	−90 ± 190	−20 ± 200	−160 ± 220	−120 ± 240
Total SM	109 600 ± 1100	102 200 ± 1500	109 700 ± 1000	102 200 ± 1000
Data	109 690 ± 900	102 320 ± 970	109 690 ± 900	102 320 ± 970

Table 10.2: Measured fiducial cross-sections times the single-lepton-flavor W boson branching ratio and the cross-section ratios. $R_c^{\pm}(D^{(*)})$ is obtained by combining the individual measurements of $R_c^{\pm}(D^{+})$ and $R_c^{\pm}(D^{*+})$ as explained in the text.

Channel	$\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}(W$+$D^{*+}) \times B(W \to \ell\nu)$ [pb]
W^-+D^{*+}	51.1 ± 0.4 (stat.) $^{+1.9}_{-1.8}$ (syst.)
W^++D^{*-}	50.0 ± 0.4 (stat.) $^{+1.9}_{-1.8}$ (syst.)
	$R_c^{\pm} = \sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}(W^+$+$D^{(*)})/\sigma_{\mathrm{fid}}^{\mathrm{OS-SS}}(W^-$+$D^{(*)})$
$R_c^{\pm}(D^{*+})$	0.980 ± 0.010 (stat.) ± 0.013 (syst.)
$R_c^{\pm}(D^{(*)})$	0.971 ± 0.006 (stat.) ± 0.011 (syst.)



(a) (b)

Figure 10.1: Post-fit OS-SS W+$D^{(*)}$ signal and background predictions compared with data: (a) W^-+D^{*+} channel, and (b) W^++D^{*-} channel. The "SM Tot." line represents the sum of all signal and background samples and the corresponding hatched band shows the full post-fit systematic uncertainty. The five bins associated with the signal samples are the truth bins of the $p_T(D^{(*)})$ differential distribution.

observable. The signal mass-peak shape uncertainties have the most significant pulls in the fit. This impact is small and found to be up to 1 % for cross-sections and negligible for $R_c^{\pm}$.

The measured W+$D^{(*)}$ cross-section is compared to a set of theoretical predictions differing by the PDF sets used for each prediction PDF sets are configured (see Chapter 3.2) and use the signal MG_AMC@NLO+PY8 (NLO) samples for all predictions. Regularization of the cross-section is obtained by choosing a finite charm quark mass of $m_c = 1.55$GeV. A full CKM matrix is used to calculate the hard-scattering amplitudes. Alternative generator weights using the LHAPDF prescription [135] provide the uncertainties calculated for each PDF set. Parton shower modeling uncertainties are estimated from a comparison the baseline configuration a configuration using HERWIG [7.2] [136]. Calculations comparing the MG_AMC@NLO+PY8 (NLO) cross-sections with the calculations using POWHEG (Ref. [39]) provide NLO matching uncertainties. Off-shell W boson decays (including spin correlations) and non-diagonal CKM matrix elements effects are included in all calculations. Renormalization and factorization scales are set to one half of the transverse mass. All final-state partons and leptons are used for the calculation of the transverse mass. Both samples utilize the same ABMP16_3_NLO PDF set with the α_S value set to 0.118 and Monash PYTHIA8 [8.2] tune. production fractions uncertainties for direct charm is taken from the

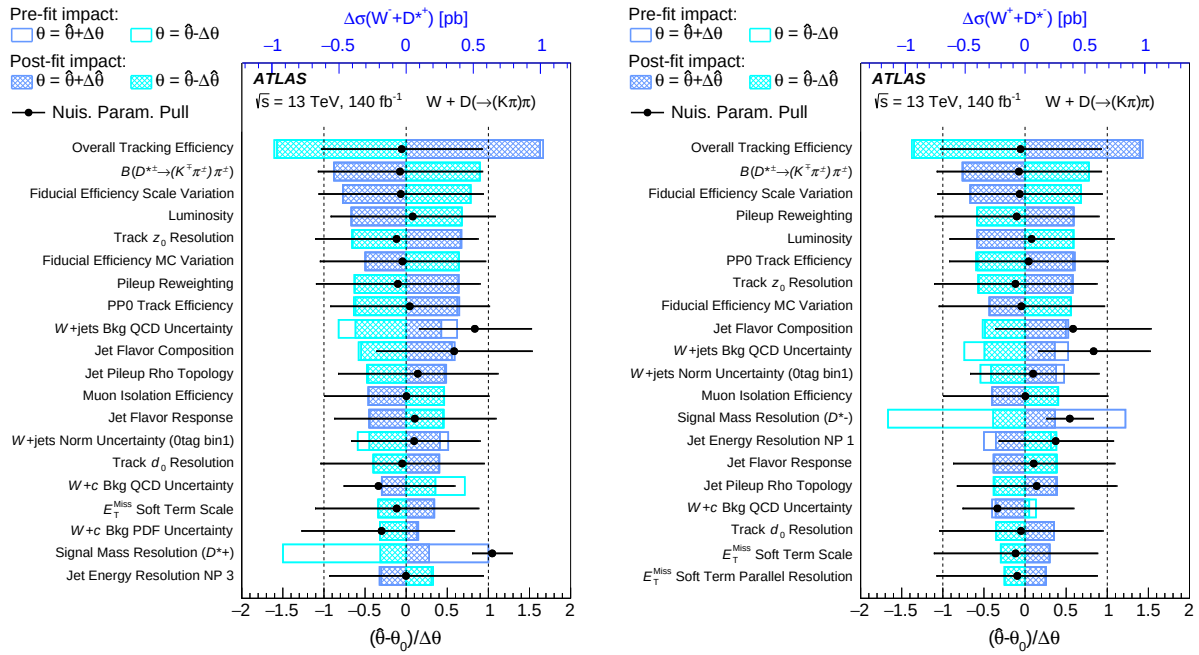


Figure 10.2: Impact of systematic uncertainties, for the 20 largest contributions, on the fitted cross-section from the $p_T(D^{(*)})$ fits, sorted in decreasing order. Impact on: (a) $\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^-+D^{*+})$, and (b) $\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^++D^{*-})$. The impact of pre-fit (post-fit) nuisance parameters $\vec{\theta}$ on the signal strength are shown with empty (colored) boxes. The post-fit central value ($\hat{\theta}$) and uncertainty are shown for each parameter with black dots.

results in Ref. [67].

Both charged channels results are consistent with one another as shown in Figure 10.3. The measured fiducial cross-sections are compared with the theoretical predictions mentioned in the previous paragraph. A PDF set highly sensitive to the strangeness of the proton – NNPDF3.1_strange [16] is also included in these plots. The total NLO theory calculations are the largest uncertainty for the theory predictions. The PDF uncertainties are comparable to the size of the measured cross-section uncertainties. The measured and set of predicted cross-sections are consistent within the total theory and PDF uncertainties combined.

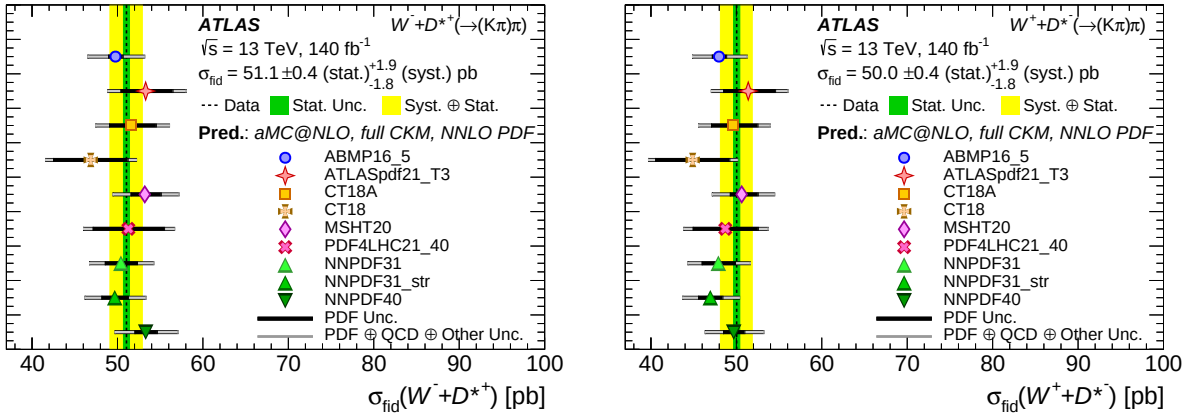
The cross-section ratio, $R_c^{\pm}$, is shown for the combined D^+ and D^{*+} channel measurements in Figure 10.4. This combined result is consistent with theoretical predictions for all PDF sets, although the prediction obtained using `NNPDF4.0nnlo` shows some tension with the measurement. Unlike the cross-section measurements, which are dominated by systematic uncertainties, the measurements of $R_c^{\pm}$ have comparable statistical and systematic uncertainties. PDF set uncertainties for $R_c^{\pm}$ fall into two categories. Those sets that impose the

restriction that the strange-sea be symmetric ($s = \bar{s}$), such as CT18 and AMBP16, predict $R_c^{\pm}$ with high precision while PDF fits that allow the s and $\bar{s}$ distributions to differ, such as NNPDF or MSHT, have larger uncertainties. These measurements are consistent with the predictions obtained with PDF sets that impose a symmetric s-$\bar{s}$ sea, suggesting that any s-$\bar{s}$ asymmetry is small in the Bjorken-x region probed by this measurement. Reference [19] presents a detailed study of the NLO and NNLO fiducial cross-sections for different charm-jet selections. That study uses the same lepton fiducial definition as this paper. While W+c-jet cross-section calculations cannot be compared with $\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W+D^{(*)})$ measurements, they provide insight into the behavior of $R_c^{\pm}$. The W+c-jet $R_c^{\pm}$ value calculated at NLO using an OS-SS selection is consistent within statistical uncertainties with that obtained for $W+D^{(*)}$ using MG_aMC@NLO+Py8 (NLO) and the same PDF set (`NNPDF3.0nlo`). The NNLO+EW(NLO) value of the W+c-jet $R_c^{\pm}$ is smaller than the NLO value by ∼1%, but the two are consistent within the quoted 1% statistical uncertainty. The effects of NNLO scale uncertainties on $R_c^{\pm}$ are below 0.3%. These results suggest that the PDF comparisons presented in Figure 10.4 are likely to look similar for an NNLO+EW(NLO) calculation.

The differential cross-sections are shown in Figure 10.5, together with the predicted cross-sections obtained with different choices of NNLO PDF set: ABMP16_5 [30], ATLASpdf21_T3 [13], CT18A, CT18 [31],MSHT20 [4], PDF4LHC21_40 [32], NNPDF31 [33], NNPDF31_str [16], NNPDF40 [34]. ABMP16_5, ATLASpdf21_T3, CT18A, and CT18 impose symmetric strange-sea PDFs. For each charge, the differential distributions are plotted in three separate panels. The top panel compares the measured differential cross-section with theoretical predictions obtained using the same PDF sets as in Figure 10.3. Systematic uncertainties in the predictions are correlated between bins and are dominated by uncertainties in the normalization. Differences between PDF sets can be seen more clearly in the middle and lower panels, which show the normalized differential cross-sections and the ratio of the predictions to the normalized cross-sections, respectively. Because the integral of the normalized cross-section across all bins is constrained to be unity, the measurements are highly correlated between bins: if the normalized cross-section in one bin increases, that in another bin must decrease.

The $p_T(D^{(*)})$ distribution and shape predicted for various PDF sets are all compatible with the measured distribution within their uncertainties. p_T-dependent systematic uncertainties in the fiducial efficiency serve to reduce the sensitivity of PDF choice. No strong constraint on PDF choice can be extracted from this measurement. It will nevertheless provide an important test of the modeling of this distribution in the MC predictions. The $|\eta(\ell)|$ measurement however provides a good sensitivity to PDF choice due to the overall smaller systematic uncertainties. All PDF sets predict a more narrow distribution of the $|\eta(\ell)|$ variable than that of the measured differential cross-sections. The largest discrepancy is shown in the highest p_T bin which shows some tension between the theory predictions and the measured value.

The similarity of the measurement and theoretical predictions can be calculated using a formula to calculate a χ^2 value which estimates their consistency. This formula uses both



(a) (b)

Figure 10.3: Measured fiducial cross-section times the single-lepton-flavor W branching ratio compared with different NNLO PDF predictions for (a) W^-+D^{*+}, and (b) W^++D^{*-}. The dotted vertical line shows the central value of the measurement, the green band shows the statistical uncertainty and the yellow band shows the combined statistical and systematic uncertainty. The PDF predictions are designated by markers. The inner error bars on the theoretical predictions show the 68% CL uncertainties obtained from the error sets provided with each PDF set, while the outer error bar represents the quadrature sum of the 68% CL PDF, scale, hadronization, and matching uncertainties. The PDF predictions are based on NLO calculations performed using MG_aMC@NLO+Py8 (NLO) and a full CKM matrix.

the experimental and theory covariance matrices through,

$$\chi^2 = \sum_{i,j} (x_i - \mu_i) \left(C^{-1}\right)_{ij} (x_j - \mu_j),$$

where $\vec{x}$ are the measured differential cross-sections in the 10 differential bins, and $\vec{\mu}$ are the predicted cross-sections in the same bin and depend on the choice of PDF set. The total covariance matrix C is the sum of the experimental (measurement error) and theory (theory predictions) contributions to the uncertainty. A p-value with 10 degrees of freedom is calculated from this χ^2 value. This calculation is performed only for the $|\eta(\ell)|$ results, since they show the most dependence on PDF choice.

Table 10.3 shows the p-values for the different choices of PDF sets calculated as described above. The result of the "Exp. Only" calculation, where no systematic uncertainties related to the theory predictions are included is shown in the first column. Each subsequent column shows the result of adding to the theory covariance matrix and the p-value calculation the named uncertainty. The last column includes all theory uncertainties: QCD scale,

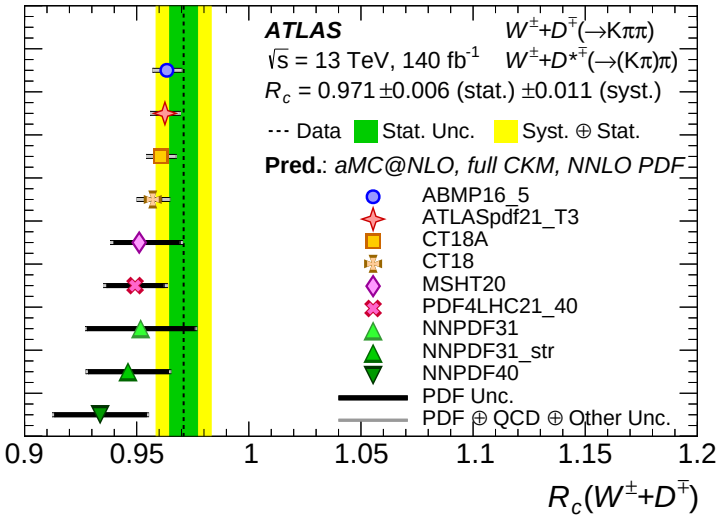


Figure 10.4: Measured fiducial cross-section ratio, $R_c^{\pm}$, compared with different PDF predictions. The data are a combination of the separate $W+D^+$ and $W+D^{*+}$ channel measurements. The dotted vertical line shows the central value of the measurement, the green band shows the statistical uncertainty and the yellow band shows the combined statistical and systematic uncertainty. The PDF predictions are designated by markers. The inner error bars on the theoretical predictions show the 68% CL uncertainties obtained from the error sets provided with each PDF set, while the outer error bar represents the quadrature sum of the 68% CL PDF, scale, hadronization, and matching uncertainties. The PDF predictions are based on NLO calculations performed using MG_aMC@NLO+Py8 (NLO) and a full CKM matrix.

"hadronization and matching", and PDF uncertainties. The combination of multiple systematics make up the hadronization and matching uncertainty. This is calculated by adding in quadrature the uncertainties due to: choice of tune, showering program, charm production fractions and the matching algorithm used. Fully correlated uncertainties between the $W^+ + D^{*-}$ and $W^- + D^{*+}$ channels is chosen for this purpose. p-values are below 10% for most of the PDFs when not considering theory uncertainties. Taking into account additional uncertainties (hadronization and QCD scale) raises the calculated p-values to 24%. The PDF uncertainty has a large effect on the shape of the differential $|\eta(\ell)|$ distribution and its inclusion greatly raises the p-values calculated for various PDF sets. These measurements will therefore provide a good constraint on PDF variations when included in a global PDF fit.

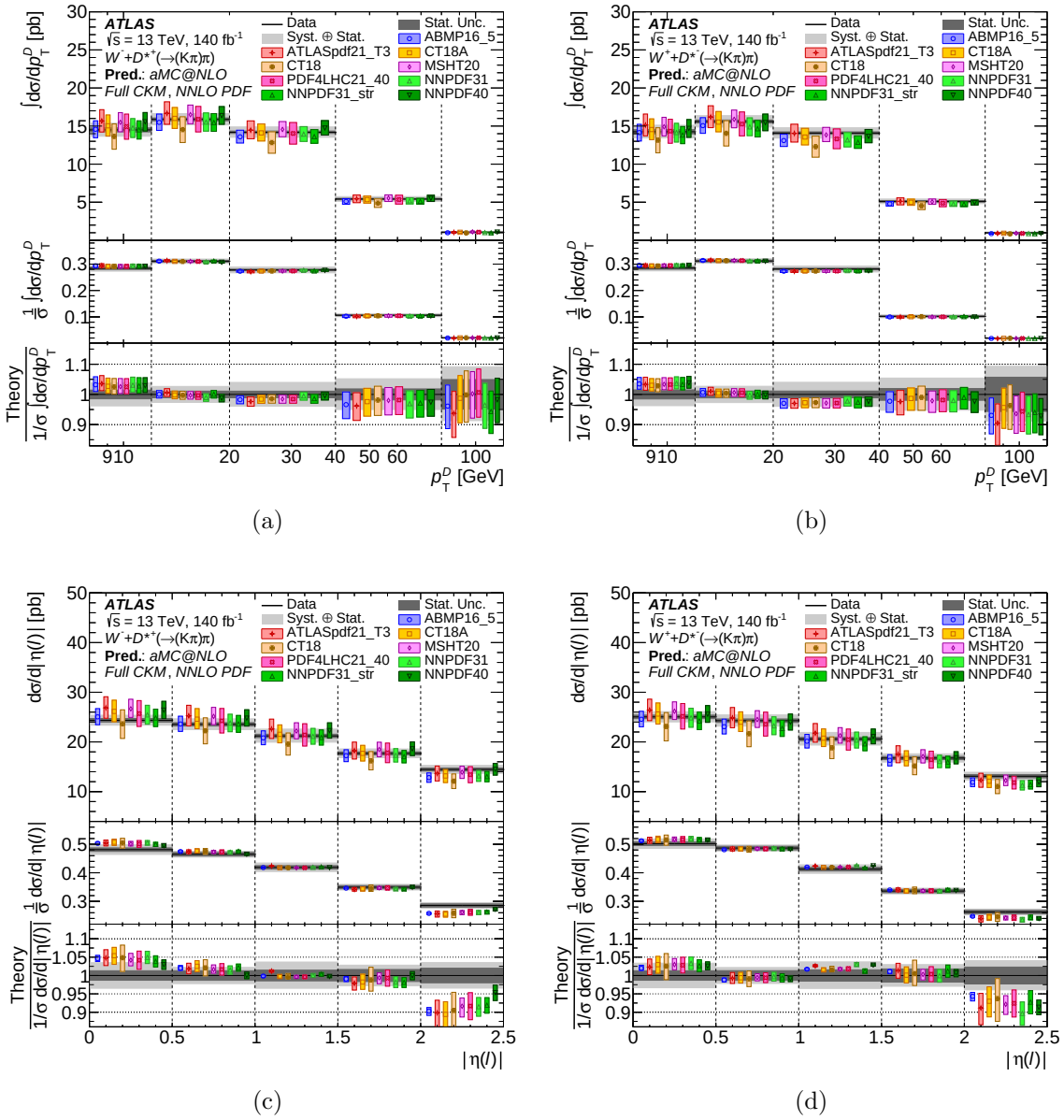


Figure 10.5: Measured differential fiducial cross-section times the single-lepton-flavor W branching ratio compared with different PDF predictions in the D^{*+} channel: (a) W^-+D^{*+} $p_T(D^{*+})$, (b) W^++D^{*-} $p_T(D^{*+})$, (c) W^-+D^{*+} $|\eta(\ell)|$, and (d) W^++D^{*-} $|\eta(\ell)|$. The displayed cross sections in p_T (D^+) plots are integrated over each differential bin. Error bars on the MC predictions are the quadrature sum of the QCD scale uncertainty, PDF uncertainties, hadronization uncertainties, and matching uncertainty. The PDF predictions are based on NLO calculations performed using MG_aMC@NLO+Py8 (NLO) and a full CKM matrix.

Table 10.3: The p-values for compatibility of the measurement and the predictions, calculated with the χ^2 formula using experimental and theory covariance matrices. The first column shows the p-values for the $|\eta(\ell)|$ (D^{*+}) differential cross-section using only experimental uncertainties. The next columns show p-values when progressively more theory systematic uncertainties are included. The "Had"" column also includes effects of NLO matching uncertainties. The PDF predictions are based on NLO calculations performed using MG_aMC@NLO+Py8 (NLO) and a full CKM matrix: ABMP16_5 [30], ATLASpdf21_T3 [13], CT18A, CT18 [31], MSHT20 [4], PDF4LHC21_40 [32], NNPDF31 [33], NNPDF31_str [16], NNPDF40 [34]. ABMP16_5, ATLASpdf21_T3, CT18A, and CT18 impose symmetric strange-sea PDFs.

Channel	D^{*+} $\|\eta(\ell)\|$			
p-value for PDF [%]	Exp. Only	$\oplus$ QCD Scale	$\oplus$ Had.	$\oplus$ PDF
ABMP16_5_nnlo	22.8	23.7	25.0	28.8
ATLASpdf21_T3	1.9	2.9	3.4	33.7
CT18ANNLO	6.5	6.9	7.8	47.3
CT18NNLO	9.4	19.2	19.7	52.8
MSHT20nnlo_as118	7.0	9.4	10.4	31.3
PDF4LHC21_40	14.2	14.2	15.2	51.4
NNPDF31_nnlo_as_0118_hessian	5.0	5.1	5.5	34.9
NNPDF31_nnlo_as_0118_strange	11.4	12.4	13.2	46.0
NNPDF40_nnlo_as_01180_hessian	4.5	6.1	6.4	36.0

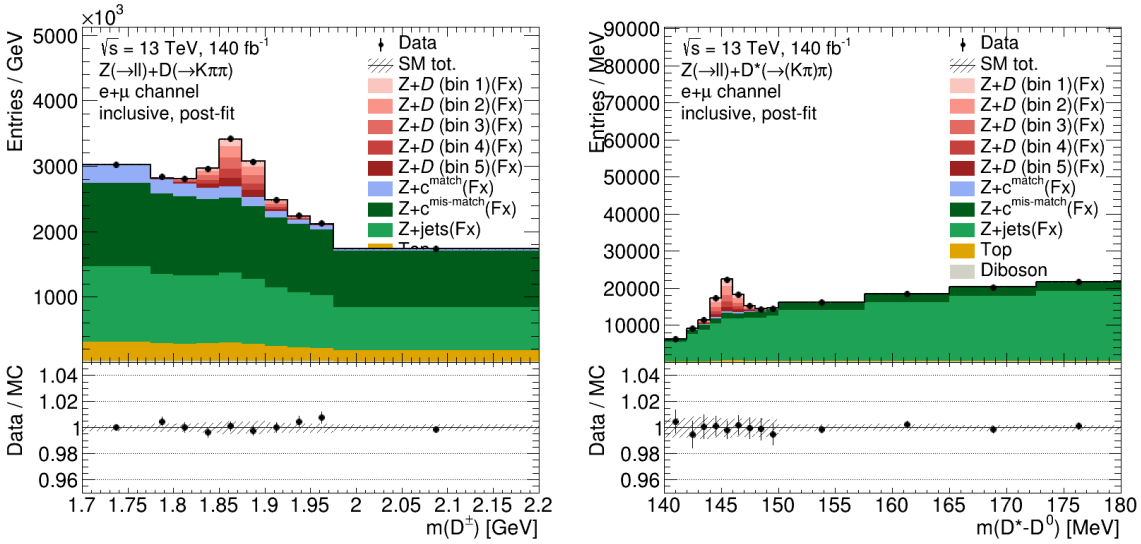
10.2 Z+$D^{(*)}$ Results

Post-fit comparisons between the data and MC distributions for the D^+ and D^{*+} channels are shown in Figure 10.6. All of the data points are within the resulting 1σ systematic uncertainty band. The Z+$D^{(*)}$ post-fit yields obtained with the likelihood fit are given in Table 10.4.

The resulting cross-sections $\sigma_{\text{fid}}(Z\text{+}D^{(*)}) \times B(Z \to \ell\ell)$ are presented in Table 10.5. The results presented here are obtained using the $p_T(D^{(*)})$ fit. For this analysis, a different set of reconstruction criteria where used for the D mesons that went into our signal region (see Chapter 5). The 3D impact parameter cut imposed for the W+D analysis removes a significant fraction (67%) of the candidates where the D meson was the product of a B meson decay. Removing such candidates is appropriate in the W+D analysis, which targets prompt charm production. The goal Z+D analysis, however, is to measure inclusive Z+D production, including those D mesons that result from B decays. The removal of the 3D impact parameter cut and the re-optimization of some additional cuts ensures good acceptance for both prompt

D and non-prompt D. Because a significant fraction of the D mesons result from B decays, there is no guarantee that the ratios of cross-sections of the D^+ to D^* channels will be consistent with production fractions, which were obtained using prompt charm hadrons.

Theoretical predictions of the $Z+D^{(*)}$ cross-section for two state-of-the-art MC generators are obtained using the signal SHERPA (2.2.11) and MG_AMC@NLO+PY8 (FxFx) samples with the configurations described in Chapter 3.2. Figure 10.7 shows the comparison of the measured cross-sections and the theoretical MC generator predictions. The total fiducial cross-section in the D^+ falls within the two theoretical predictions. In the D^* channel, the measured cross-section is larger than both theoretical predictions. One possible reason for the discrepancy is an underestimate of the contribution of B meson production to the $Z+D^{(*)}$ cross section.



(a) (b)

Figure 10.6: Post-fit $Z+D^{(*)}$ signal and background predictions compared with data: (a) $Z+D^+$ channel, and (b) $Z+D^{*+}$ channel. The "SM Tot." line represents the sum of all signal and background samples and the corresponding hatched band shows the full post-fit systematic uncertainty. The five bins associated with the signal samples are the truth bins of the $p_T(D^{(*)})$ differential distribution.

The impact of the nuisance parameters on the fitted values of the absolute fiducial cross-section in the differential $p_T(D^{(*)})$ fits is shown as a "ranking plot" in Figure 10.8. The 20 nuisance parameters with the largest contribution are ordered by decreasing impact on the corresponding observable. The post-fit central values and uncertainties of the corresponding parameters are given in the same plots. The γ parameters are the Poisson-constrained parameters accounting for the MC statistical uncertainties in the combined signal-plus-background

Table 10.4: Post-fit yields in the $Z+D^{(*)}$ SR from the $p_T(D^{(*)})$ differential fit. Uncertainties in individual SM components are the full post-fit systematic uncertainties.

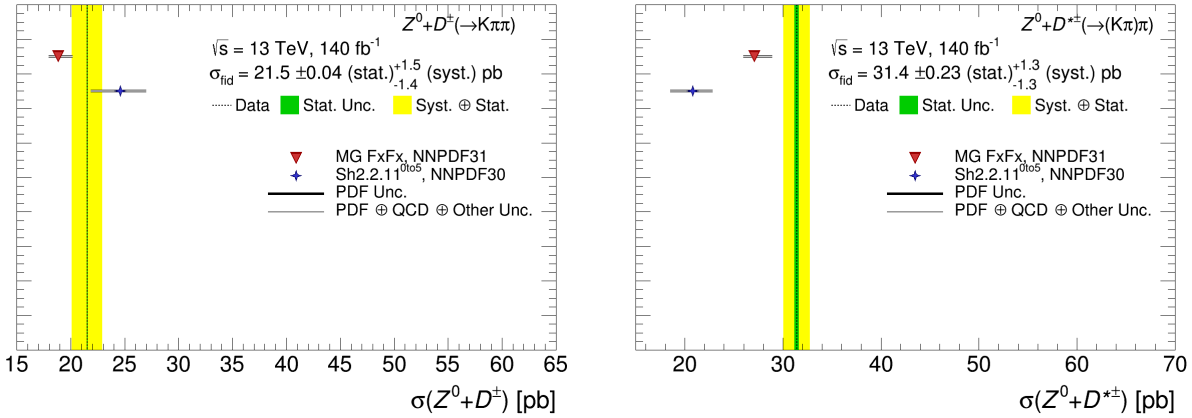
Sample	$Z+D^+$ SR ($p_T(D^+)$ fit)		$Z+D^{*+}$ SR ($p_T(D^*)$ fit)	
$Z+D^{(*)}$ (bin 1)	6808 $\pm$	620	4452 $\pm$	318
$Z+D^{(*)}$ (bin 2)	11 718 $\pm$	579	6124 $\pm$	314
$Z+D^{(*)}$ (bin 3)	11 599 $\pm$	491	4691 $\pm$	218
$Z+D^{(*)}$ (bin 4)	9192 $\pm$	480	3651 $\pm$	150
$Z+D^{(*)}$ (bin 5)	9933 $\pm$	658	3420 $\pm$	139
$Z+c^{\text{match}}$	55 577 $\pm$	4368	3821 $\pm$	305
$Z+c^{\text{mis-match}}$	513 767 $\pm$	27 942	77 491 $\pm$	3941
Z+jets	429 662 $\pm$	30 184	590 804 $\pm$	4568
$t\bar{t}$ + single top	107 939 $\pm$	13 344	9954 $\pm$	1261
Diboson	8624 $\pm$	1734	3981 $\pm$	802
Total SM	1 164 820 $\pm$	2359	708 389 $\pm$	1758
Data	1 164 829 $\pm$	1079	708 394 $\pm$	842

Table 10.5: Measured fiducial cross-sections times the single-lepton-flavor Z boson branching ratio.

Channel	$\sigma_{\text{fid}}(Z+D^{(*)}) \times B(Z \to \ell\ell)$ [pb]
$Z+D^+$	21.9 ± 0.04 (stat.) $^{+1.5}_{-1.4}$ (syst.)
$Z+D^{*+}$	31.1 ± 0.23 (stat.) $^{+1.3}_{-1.3}$ (syst.)

mass template and are centered around 1 by definition. The ranking plots for both $Z+D^{(*)}$ modes share some similarities: Fiducial Efficiency MC Variation systematics, Z+background template uncertainties and γ parameters rank highly. The fit has the degrees of freedom to constrain the size of the Z backgrounds and their systematic uncertainties which results in large pulls and constraints. The presence of multiple γ parameters mean the measurement is limited by the amount of data and MC events available for the fits. Increasing both the amount of data collected and MC events generated would lower the effect of these parameters and the overall uncertainties on the measurement. The ranking plots demonstrate that most nuisance parameters with large impact on the integrated fiducial cross-section do not deviate significantly from the initial values in the likelihood fit. The parameters that have the most significant pulls in the fit have an impact of the corresponding systematic uncertainties on the observables which is small (up to 2 % for cross-sections).

The normalized differential cross-sections are given in Tables 10.6 and 10.7. Shown in

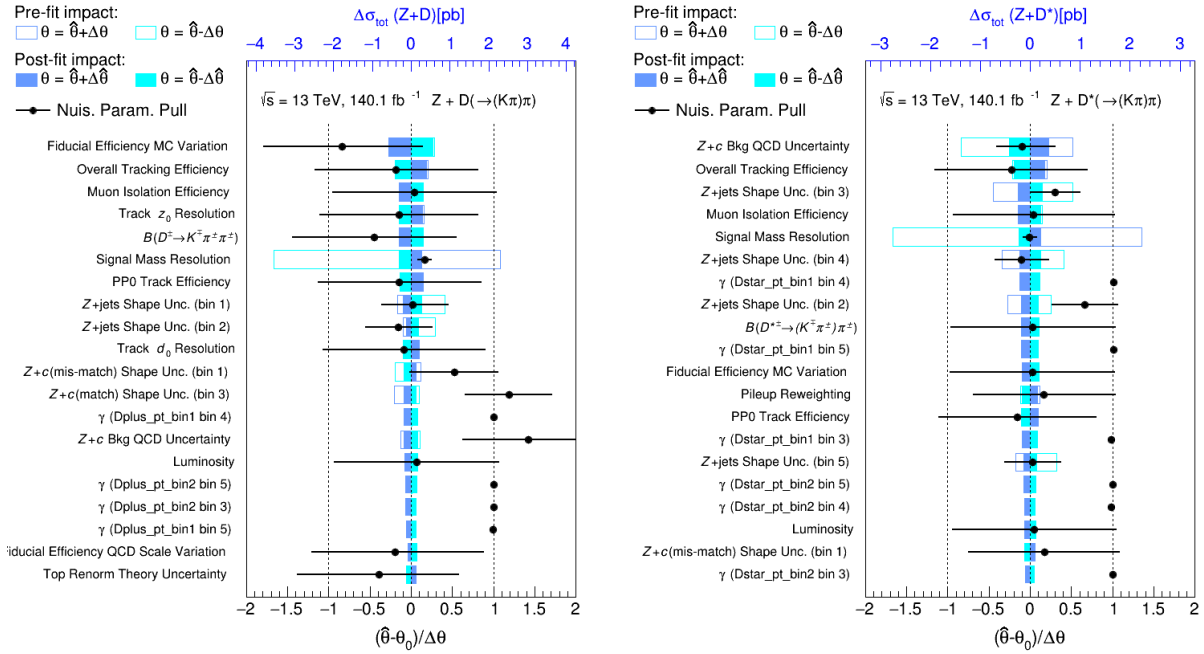


(a) (b)

Figure 10.7: Measured total fiducial cross-section times the double-lepton-flavor Z branching ratio compared with different MC generator predictions in the (a) D^+ channel, and the (b) D^* channel. Error bars on the MC predictions are the quadrature sum of the QCD scale uncertainty, PDF uncertainties, and matching uncertainty. The PDF predictions are based on NLO calculations.

Figure 10.9 are the measured differential cross-sections together with the predicted cross-sections obtained with the different choices of theory predictions. For each meson, the differential distributions are plotted in three separate panels. The top panel compares the measured differential cross-section with theoretical predictions. Systematic uncertainties in the predictions are correlated between bins and are dominated by uncertainties in the normalization. Differences between MC generators can be seen more clearly in the middle and lower panels, which show the normalized differential cross-sections and the ratio of the predictions to the normalized cross-sections, respectively. Because the integral of the normalized cross-section across all bins is constrained to be unity, the measurements are highly correlated between bins: if the normalized cross-section in one bin increases, that in another bin must decrease.

Variations in the shape of the $p_T(D^{(*)})$ distribution depend only weakly on the choice of MC generator. Both MC generators give theoretical predictions consistent with one another within their uncertainties. The measured $D^{(*)}$ p_T distribution is better modeled by MG_aMC@NLO+Py8 (FxFx) in the D^+ channel and SHERPA (2.2.11) in the D^* channel. The modeling in the D^+ channel shows some differences in the first and third bin, showing an overall harder p_T spectra in the data than that predicted by either MC generator. The modeling in the D^* channel shows a consistent result between the measured data and the two theory predictions. The measurements of the cross-section as a function of $p_T(D^{(*)})$ are an important test of the quality of MC modeling which show that a better modeling for the



(a) (b)

Figure 10.8: Impact of systematic uncertainties, for the 20 largest contributions, on the fitted cross-section from the $p_T(D^{(*)})$ fits, sorted in decreasing order. Impact on: (a) $\sigma_{\text{fid}}(Z{+}D^{+})$, and (b) $\sigma_{\text{fid}}(Z{+}D^{*+})$. The impact of pre-fit (post-fit) nuisance parameters $\vec{\theta}$ on the signal strength are shown with empty (colored) boxes. The post-fit central value ($\hat{\theta}$) and uncertainty are shown for each parameter with black dots.

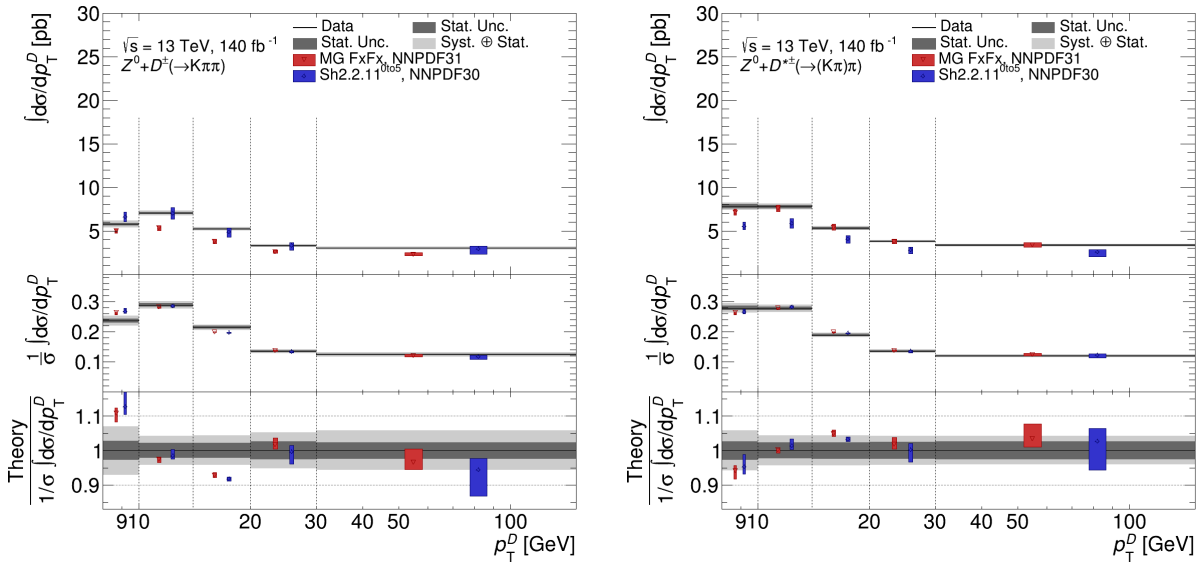
D^{+} p_T spectra can be obtained.

Table 10.6: Measured normalized differential cross-sections times the single-lepton-flavor Z boson branching ratio for the D^+ p_T measurement.

D^+ p_T bin	$\frac{1}{\sigma}\int_{bin}\frac{d\sigma}{dp_T} \times B(Z \to \ell\ell)$
8 - 10 GeV	0.237 ± 0.007 (stat.) $^{+0.015}_{-0.015}$ (syst.)
10 - 14 GeV	0.286 ± 0.006 (stat.) $^{+0.011}_{-0.010}$ (syst.)
14 - 20 GeV	0.214 ± 0.005 (stat.) $^{+0.008}_{-0.008}$ (syst.)
20 - 30 GeV	0.140 ± 0.004 (stat.) $^{+0.006}_{-0.006}$ (syst.)
30 - 150 GeV	0.123 ± 0.003 (stat.) $^{+0.006}_{-0.006}$ (syst.)

Table 10.7: Measured normalized differential cross-sections times the single-lepton-flavor Z boson branching ratio for the D^* p_T measurement.

D^* p_T bin	$\frac{1}{\sigma}\int_{bin}\frac{d\sigma}{dp_T} \times B(Z \to \ell\ell)$
8 - 10 GeV	0.280 ± 0.008 (stat.) $^{+0.014}_{-0.014}$ (syst.)
10 - 14 GeV	0.277 ± 0.006 (stat.) $^{+0.010}_{-0.010}$ (syst.)
14 - 20 GeV	0.189 ± 0.005 (stat.) $^{+0.007}_{-0.007}$ (syst.)
20 - 30 GeV	0.135 ± 0.003 (stat.) $^{+0.005}_{-0.004}$ (syst.)
30 - 150 GeV	0.120 ± 0.003 (stat.) $^{+0.004}_{-0.004}$ (syst.)



(a) (b)

Figure 10.9: Measured differential fiducial cross-section times the double-lepton-flavor Z branching ratio compared with different MC generator predictions in the D^+ channel: (a) Z+D^+ $p_T(D^+)$, and the D^* channel: (b) Z+D^{*+} $p_T(D^*)$. The displayed cross sections in p_T $(D^{(*)})$ plots are integrated over each differential bin. Error bars on the MC predictions are the quadrature sum of the QCD scale uncertainty, PDF uncertainties, and matching uncertainty. The PDF predictions are based on NLO calculations.

Chapter 11

Conclusion

Fiducial cross-sections for W boson production and Z boson production in association with a $D^{(*)}$ meson are measured as a function of $p_T(D^{(*)})$ and $|\eta(\ell)|$ using $140.1\,\text{fb}^{-1}$ of $\sqrt{s} = 13\,\text{TeV}$ pp collision data collected with the ATLAS detector at the Large Hadron Collider. A secondary-vertex fit is used to tag events containing a D^+ or a D^{*+} meson and a profile likelihood fit is used to extract the $V{+}D^{(*)}$ observables.

The $W{+}D^{(*)}$ single-lepton-species integrated cross-sections and cross-section ratios for the fiducial region $p_T(\ell) > 30\,\text{GeV}$, $|\eta(\ell)| < 2.5$, $p_T(D^{(*)}) > 8\,\text{GeV}$ and $|\eta(D^{(*)})| < 2.2$ are measured to be:

$$\begin{aligned}
\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^-{+}D^{*+}) &= 51.1 \pm 0.4\,(\text{stat.})\,^{+1.9}_{-1.8}\,(\text{syst.})\,\text{pb} \\
\sigma_{\text{fid}}^{\text{OS}-\text{SS}}(W^+{+}D^{*-}) &= 50.0 \pm 0.4\,(\text{stat.})\,^{+1.9}_{-1.8}\,(\text{syst.})\,\text{pb} \\
R_c^{\pm}(D^{(*)}) &= 0.971 \pm 0.006\,(\text{stat.}) \pm 0.011\,(\text{syst.})
\end{aligned}$$

The uncertainty in the measured absolute integrated and the fiducial differential cross-sections is about 3 %-5 % and is dominated by the systematic uncertainty. On the other hand, cross-section ratios and normalized differential cross-sections are measured with percent-level precision in the $W{+}D^{(*)}$ measurement and have comparable contributions from systematic and statistical uncertainties. The experimental precision of these measurements is comparable to the PDF uncertainties and smaller than the total theory uncertainty. Measured differential cross-sections of the $W{+}D^{(*)}$ selection as a function of $|\eta(\ell)|$ have a broader distribution than the central values of the predictions. These measurements are, however, consistent with the predictions if the uncertainties associated with the PDF sets are included, indicating that these measurements would provide useful constraints for global PDF fits. The measured values of $R_c^{\pm}$ are consistent with predictions obtained with a range of PDF sets, including those that constrain the s-$\overline{s}$ sea to be symmetric.

The $Z{+}D^{(*)}$ lepton-species integrated cross-sections for the fiducial region of $p_T(\ell) >$

27 GeV, $|\eta(\ell)| < 2.5$, $p_T(D^{(*)}) > 8$ GeV and $|\eta(D^{(*)})| < 2.2$ are measured to be:

$$\begin{aligned} \sigma_{\text{fid}}(Z{+}D^{+}) &= 21.9 \pm 0.04\,(\text{stat.})\,^{+1.5}_{-1.4}\,(\text{syst.})\,\text{pb} \\ \sigma_{\text{fid}}(Z{+}D^{*+}) &= 31.1 \pm 0.23\,(\text{stat.})\,^{+1.3}_{-1.3}\,(\text{syst.})\,\text{pb} \end{aligned}$$

In the $Z{+}D^{(*)}$ analysis the uncertainty in the measured absolute integrated and differential fiducial cross-sections is about 5 %-7 % and has a larger contribution from statistical uncertainties. The modeling of the $Z{+}D^{(*)}$ $p_T(D^{(*)})$ distribution show the MC generators predict greater amount of low p_T D mesons than is measured in data. The results of the $Z{+}D^{(*)}$ measurement can be used to better model the $p_T(D^{(*)})$ distribution in Monte Carlo theoretical predictions.

Bibliography

[1] ATLAS Collaboration. "Measurement of the production of a W boson in association with a charmed hadron in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 108 (3 2023), p. 032012. DOI: 10.1103/PhysRevD.108.032012. URL: https://link.aps.org/doi/10.1103/PhysRevD.108.032012.

[2] MissMJ and Cush. *Standard model of elementary particles: the 12 fundamental fermions and 5 fundamental bosons.* 2009. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[3] *Summary of several Standard Model total production cross section measurements.* URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/index.html#ATLAS_b_SMSummary_FiducialXsect.

[4] S. Bailey et al. "Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs". In: *Eur. Phys. J. C* 81.4 (2021), p. 341. DOI: 10.1140/epjc/s10052-021-09057-0. arXiv: 2012.04684 [hep-ph].

[5] HERMES Collaboration. "Measurement of Parton Distributions of Strange Quarks in the Nucleon from Charged-Kaon Production in Deep-Inelastic Scattering on the Deuteron". In: *Phys. Lett. B* 666 (2008), pp. 446–450. DOI: 10.1016/j.physletb.2008.07.090. arXiv: 0803.2993 [hep-ex].

[6] ZEUS Collaboration. "Charm production in charged current deep inelastic scattering at HERA". in: *JHEP* 05 (2019), p. 201. DOI: 10.1007/JHEP05(2019)201. arXiv: 1904.03261 [hep-ex].

[7] CCFR Collaboration. "Determination of the strange quark content of the nucleon from a next-to-leading order QCD analysis of neutrino charm production". In: *Z. Phys. C* 65 (1995), pp. 189–198. DOI: 10.1007/BF01571875. arXiv: hep-ex/9406007.

[8] CHORUS Collaboration. "Leading order analysis of neutrino induced dimuon events in the CHORUS experiment". In: *Nucl. Phys. B* 798 (2008), pp. 1–16. DOI: 10.1016/j.nuclphysb.2008.02.013. arXiv: 0804.1869 [hep-ex].

[9] NuTeV Collaboration. "Precise Measurement of Dimuon Production Cross-Sections in ν_μ Fe and $\bar{\nu}_\mu$ Fe Deep Inelastic Scattering at the Tevatron." In: *Phys. Rev. D* 64 (2001), p. 112006. DOI: 10.1103/PhysRevD.64.112006. arXiv: hep-ex/0102049.

[10] NuTeV Collaboration. "Measurement of the Nucleon Strange-Antistrange Asymmetry at Next-to-Leading Order in QCD from NuTeV Dimuon Data". In: *Phys. Rev. Lett.* 99 (2007), p. 192001. DOI: 10.1103/PhysRevLett.99.192001.

[11] NOMAD Collaboration. "A Precision Measurement of Charm Dimuon Production in Neutrino Interactions from the NOMAD Experiment". In: *Nucl. Phys. B* 876 (2013), pp. 339–375. DOI: 10.1016/j.nuclphysb.2013.08.021. arXiv: 1308.4750 [hep-ex].

[12] Lyndon Evans and Philip Bryant. "LHC Machine". In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[13] 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 (2021), p. 438. DOI: 10.1140/epjc/s10052-022-10217-z. arXiv: 2112.11266 [hep-ex].

[14] 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].

[15] ATLAS Collaboration. "Measurement of the transverse momentum and ϕ^*_η distributions of Drell-Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 76 (2016), p. 291. DOI: 10.1140/epjc/s10052-016-4070-4. arXiv: 1512.02192 [hep-ex].

[16] Ferran Faura et al. "The strangest proton?" In: *Eur. Phys. J. C* 80.12 (2020), p. 1168. DOI: 10.1140/epjc/s10052-020-08749-3. arXiv: 2009.00014 [hep-ph].

[17] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[18] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[19] Michał Czakon et al. "NNLO QCD predictions for W+c-jet production at the LHC". in: *JHEP* 06 (2021), p. 100. DOI: 10.1007/JHEP06(2021)100. arXiv: 2011.01011 [hep-ph].

[20] U. Baur et al. "The charm content of W + 1 jet events as a probe of the strange quark distribution function". In: *Phys. Lett. B* 318 (1993), pp. 544–548. DOI: 10.1016/0370-2693(93)91553-Y. arXiv: hep-ph/9308370.

[21] Walter T. Giele, Stephane Keller, and Eric Laenen. "QCD corrections to W boson plus heavy quark production at the Tevatron". In: *Phys. Lett. B* 372 (1996), pp. 141–149. DOI: 10.1016/0370-2693(96)00078-0. arXiv: hep-ph/9511449.

[22] CDF Collaboration. "Observation of the Production of a W Boson in Association with a Single Charm Quark". In: *Phys. Rev. Lett.* 110.7 (2013), p. 071801. DOI: 10.1103/PhysRevLett.110.071801. arXiv: 1209.1921 [hep-ex].

[23] ATLAS Collaboration. "Measurement of the production cross section for W-bosons in association with jets in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector". In: *Phys. Lett. B* 698 (2011), p. 325. DOI: 10.1016/j.physletb.2011.03.012. arXiv: 1012.5382 [hep-ex].

[24] ATLAS Collaboration. "Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 05 (2014), p. 068. DOI: 10.1007/JHEP05(2014)068. arXiv: 1402.6263 [hep-ex].

[25] CMS Collaboration. "Measurement of associated W + charm production in pp collisions at $\sqrt{s} = 7\,\text{TeV}$". In: *JHEP* 02 (2014), p. 013. DOI: 10.1007/JHEP02(2014)013. arXiv: 1310.1138 [hep-ex].

[26] CMS Collaboration. "Measurements of the associated production of a W boson and a charm quark in proton–proton collisions at $\sqrt{s} = 8\,\text{TeV}$". In: *Eur. Phys. J. C* 82 (2021), p. 1094. DOI: 10.1140/epjc/s10052-022-10897-7. arXiv: 2112.00895 [hep-ex].

[27] CMS Collaboration. "Measurement of associated production of a W boson and a charm quark in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 79 (2019), p. 269. DOI: 10.1140/epjc/s10052-019-6752-1. arXiv: 1811.10021 [hep-ex].

[28] LHCb Collaboration. "Study of W boson production in association with beauty and charm". In: *Phys. Rev. D* 92.5 (2015), p. 052001. DOI: 10.1103/PhysRevD.92.052001. arXiv: 1505.04051 [hep-ex].

[29] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an]. Erratum: in: *Eur. Phys. J. C* 73 (2013), p. 2501. DOI: 10.1140/epjc/s10052-013-2501-z.

[30] S. Alekhin et al. "Parton distribution functions, α_s, and heavy-quark masses for LHC Run II". in: *Phys. Rev. D* 96.1 (2017), p. 014011. DOI: 10.1103/PhysRevD.96.014011. arXiv: 1701.05838 [hep-ph].

[31] Tie-Jiun Hou et al. "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC". in: *Phys. Rev. D* 103.1 (2021), p. 014013. DOI: 10.1103/PhysRevD.103.014013. arXiv: 1912.10053 [hep-ph].

[32] Richard D. Ball et al. "The PDF4LHC21 combination of global PDF fits for the LHC Run III". in: *J. Phys. G* 49.8 (2022), p. 080501. DOI: 10.1088/1361-6471/ac7216. arXiv: 2203.05506 [hep-ph].

[33] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[34] Richard D. Ball et al. "The path to proton structure at 1% accuracy". In: *Eur. Phys. J. C* 82.5 (2022), p. 428. DOI: 10.1140/epjc/s10052-022-10328-7. arXiv: 2109.02653 [hep-ph].

[35] Michał Czakon et al. "A detailed investigation of W+c-jet at the LHC". in: (Dec. 2022). arXiv: 2212.00467 [hep-ph].

[36] Andrea Banfi, Gavin P. Salam, and Giulia Zanderighi. "Infrared safe definition of jet flavor". In: *Eur. Phys. J. C* 47 (2006), pp. 113–124. DOI: 10.1140/epjc/s2006-02552-4. arXiv: hep-ph/0601139.

[37] Micha Czakon et al. "B-hadron production in NNLO QCD: application to LHC $t\bar{t}$ events with leptonic decays". In: *JHEP* 10 (2021), p. 216. DOI: 10.1007/JHEP10(2021)216. arXiv: 2102.08267 [hep-ph].

[38] 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-ph].

[39] G. Bevilacqua et al. "W + charm production with massive c quarks in PowHel". In: *JHEP* 04 (2022), p. 056. DOI: 10.1007/JHEP04(2022)056. arXiv: 2106.11261 [hep-ph].

[40] Zoltan Trocsanyi, M. V. Garzelli, and Adam Kardos. "NLO event samples for the LHC". in: *PoS* EPS-HEP2011 (2011), p. 282. DOI: 10.22323/1.134.0282. arXiv: 1111.1446 [hep-ph].

[41] The NNPDF Collaboration et al. *A Determination of the Charm Content of the Proton*. 2016. arXiv: 1605.06515 [hep-ph].

[42] Tie-Jiun Hou et al. "CT14 intrinsic charm parton distribution functions from CTEQ-TEA global analysis". In: *Journal of High Energy Physics* 2018.2 (2018). DOI: 10.1007/jhep02(2018)059. URL: https://doi.org/10.1007%2Fjhep02%282018%29059.

[43] John M. Campbell et al. "Associated production of a Z Boson and a single heavy quark jet". In: *Phys. Rev. D* 69 (2004), p. 074021. DOI: 10.1103/PhysRevD.69.074021. arXiv: hep-ph/0312024.

[44] S.J. Brodsky et al. "The intrinsic charm of the proton". In: *Physics Letters B* 93.4 (1980), pp. 451–455. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(80)90364-0. URL: https://www.sciencedirect.com/science/article/pii/0370269380903640.

[45] S. J. Brodsky, C. Peterson, and N. Sakai. "Intrinsic heavy-quark states". In: *Phys. Rev. D* 23 (11 1981), pp. 2745–2757. DOI: 10.1103/PhysRevD.23.2745. URL: https://link.aps.org/doi/10.1103/PhysRevD.23.2745.

[46] CMS Collaboration. "Measurement of differential production cross-sections for a Z boson in association with b-jets in 7 TeV proton-proton collisions with the ATLAS detector". In: *Journal of High Energy Physics* 2014.10 (2014). DOI: 10.1007/jhep10(2014)141. URL: https://doi.org/10.1007%2Fjhep10%282014%29141.

[47] CMS Collaboration. "Measurement of associated Z + charm production in proton-proton collisions at $\sqrt{\{s\}} = 8$ TeV". in: *The European Physical Journal C* 78.4 (2018). DOI: 10.1140/epjc/s10052-018-5752-x. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-018-5752-x.

[48] CMS Collaboration. "Measurement of differential cross sections for Z bosons produced in association with charm jets in pp collisions $\sqrt{\{s\}} = 13$ TeV". in: *Journal of High Energy Physics* 2021.4 (2021). DOI: 10.1007/jhep04(2021)109. URL: https://doi.org/10.1007%2Fjhep04%282021%29109.

[49] LHCb Collaboration. "Study of Z Bosons Produced in Association with Charm in the Forward Region". In: *Phys. Rev. Lett.* 128 (8 2022), p. 082001. DOI: 10.1103/PhysRevLett.128.082001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.082001.

[50] Ewa Lopienska. "The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022". In: (2022). General Photo. URL: https://cds.cern.ch/record/2800984.

[51] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[52] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3*. Tech. rep. Geneva: CERN, 2023. arXiv: 2305.16623. URL: https://cds.cern.ch/record/2859916.

[53] ATLAS Collaboration. *ATLAS Insertable B-Layer: Technical Design Report*. ATLAS-TDR-19; CERN-LHCC-2010-013. 2010. URL: https://cds.cern.ch/record/1291633. Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009. 2012. URL: https://cds.cern.ch/record/1451888.

[54] ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *JINST* 15 (2020), P10004. DOI: 10.1088/1748-0221/15/10/P10004. arXiv: 2007.12539 [hep-ex].

[55] URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[56] ATLAS Collaboration. "ATLAS data quality operations and performance for 2015–2018 data-taking". In: *JINST* 15 (2020), P04003. DOI: 10.1088/1748-0221/15/04/P04003. arXiv: 1911.04632 [physics.ins-det].

[57] ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ using the ATLAS detector at the LHC". in: (2022). arXiv: 2212.09379 [hep-ex].

[58] 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.

[59] ATLAS Collaboration. "Performance of electron and photon triggers in ATLAS during LHC Run 2". In: *Eur. Phys. J. C* 80 (2020), p. 47. DOI: 10.1140/epjc/s10052-019-7500-2. arXiv: 1909.00761 [hep-ex].

[60] 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 [hep-ex].

[61] ATLAS Collaboration. "The ATLAS Simulation Infrastructure". In: *Eur. Phys. J. C* 70 (2010), p. 823. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568 [physics.ins-det].

[62] GEANT4 Collaboration, S. Agostinelli, et al. "GEANT4 – a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[63] T. Sjöstrand, S. Mrenna, and P. Skands. "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-ph].

[64] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys. B* 867 (2013), p. 244. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[65] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model.* ATL-PHYS-PUB-2016-017. 2016. URL: https://cds.cern.ch/record/2206965.

[66] ATLAS Collaboration. "Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS Detector at the LHC". in: *Phys. Rev. Lett.* 117 (2016), p. 182002. DOI: 10.1103/PhysRevLett.117.182002. arXiv: 1606.02625 [hep-ex].

[67] Mykhailo Lisovyi, Andrii Verbytskyi, and Oleksandr Zenaiev. "Combined analysis of charm-quark fragmentation-fraction measurements". In: *Eur. Phys. J. C* 76.7 (2016), p. 397. DOI: 10.1140/epjc/s10052-016-4246-y. arXiv: 1509.01061 [hep-ex].

[68] ATLAS Collaboration. *Reweighting heavy-flavor production fractions to reduce flavor modelling uncertainties for ATLAS*. ATL-PHYS-PUB-2022-035. 2022. URL: https://cds.cern.ch/record/2816367.

[69] Enrico Bothmann et al. "Event generation with Sherpa 2.2". In: *SciPost Phys.* 7.3 (2019), p. 034. DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].

[70] Tanju Gleisberg and Stefan Höche. "Comix, a new matrix element generator". In: *JHEP* 12 (2008), p. 039. DOI: 10.1088/1126-6708/2008/12/039. arXiv: 0808.3674 [hep-ph].

[71] Federico Buccioni et al. "OpenLoops 2". In: *Eur. Phys. J. C* 79.10 (2019), p. 866. DOI: 10.1140/epjc/s10052-019-7306-2. arXiv: 1907.13071 [hep-ph].

[72] Fabio Cascioli, Philipp Maierhöfer, and Stefano Pozzorini. "Scattering Amplitudes with Open Loops". In: *Phys. Rev. Lett.* 108 (2012), p. 111601. DOI: 10.1103/PhysRevLett.108.111601. arXiv: 1111.5206 [hep-ph].

[73] Ansgar Denner, Stefan Dittmaier, and Lars Hofer. "COLLIER: A fortran-based complex one-loop library in extended regularizations". In: *Comput. Phys. Commun.* 212 (2017), pp. 220–238. DOI: 10.1016/j.cpc.2016.10.013. arXiv: 1604.06792 [hep-ph].

[74] Richard D. Ball et al. "Parton distributions for the LHC run II". in: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[75] Steffen Schumann and Frank Krauss. "A parton shower algorithm based on Catani–Seymour dipole factorisation". In: *JHEP* 03 (2008), p. 038. DOI: 10.1088/1126-6708/2008/03/038. arXiv: 0709.1027 [hep-ph].

[76] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. "A modified cluster-hadronization model". In: *Eur. Phys. J. C* 36 (2004), pp. 381–395. DOI: 10.1140/epjc/s2004-01960-8. arXiv: hep-ph/0311085.

[77] Stefan Höche et al. "A critical appraisal of NLO+PS matching methods". In: *JHEP* 09 (2012), p. 049. DOI: 10.1007/JHEP09(2012)049. arXiv: 1111.1220 [hep-ph].

[78] S. Catani et al. "QCD Matrix Elements + Parton Showers". In: *JHEP* 11 (2001), p. 063. DOI: 10.1088/1126-6708/2001/11/063. arXiv: hep-ph/0109231.

[79] Stefan Höche et al. "QCD matrix elements and truncated showers". In: *JHEP* 05 (2009), p. 053. DOI: 10.1088/1126-6708/2009/05/053. arXiv: 0903.1219 [hep-ph].

[80] Stefan Höche et al. "QCD matrix elements + parton showers. The NLO case". In: *JHEP* 04 (2013), p. 027. DOI: 10.1007/JHEP04(2013)027. arXiv: 1207.5030 [hep-ph].

[81] Enrico Bothmann, Marek Schönherr, and Steffen Schumann. "Reweighting QCD matrix-element and parton-shower calculations". In: *Eur. Phys. J. C* 76.11 (2016), p. 590. DOI: 10.1140/epjc/s10052-016-4430-0. arXiv: 1606.08753 [hep-ph].

[82] ATLAS Collaboration. "Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment". In: *JHEP* 08 (2021), p. 089. DOI: 10.1007/JHEP08(2022)089. arXiv: 2112.09588 [hep-ex].

[83] Leif Lönnblad. "Correcting the Colour-Dipole Cascade Model with Fixed Order Matrix Elements". In: *JHEP* 05 (2002), p. 046. DOI: 10.1088/1126-6708/2002/05/046. arXiv: hep-ph/0112284.

[84] Leif Lönnblad and Stefan Prestel. "Matching tree-level matrix elements with interleaved showers". In: *JHEP* 03 (2012), p. 019. DOI: 10.1007/JHEP03(2012)019. arXiv: 1109.4829 [hep-ph].

[85] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014. URL: https://cds.cern.ch/record/1966419.

[86] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152. DOI: 10.1016/S0168-9002(01)00089-4.

[87] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". in: *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph].

[88] G. Bonvicini et al. "Dalitz plot analysis of the D+ —> K- pi+ pi+ decay". In: *Phys. Rev. D* 78 (2008), p. 052001. DOI: 10.1103/PhysRevD.78.052001. arXiv: 0802.4214 [hep-ex].

[89] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction". In: *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

[90] Paolo Nason. "A new method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[91] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[92] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". in: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[93] ATLAS Collaboration. *Studies on top-quark Monte Carlo modelling for Top2016*. ATL-PHYS-PUB-2016-020. 2016. URL: https://cds.cern.ch/record/2216168.

[94] Stefano Frixione et al. "Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations". In: *JHEP* 04 (2007), p. 081. DOI: 10.1088/1126-6708/2007/04/081. arXiv: hep-ph/0702198.

[95] Pierre Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *JHEP* 03 (2013), p. 015. DOI: 10.1007/JHEP03(2013)015. arXiv: 1212.3460 [hep-ph].

[96] M. Beneke et al. "Hadronic top-quark pair production with NNLL threshold resummation". In: *Nucl. Phys. B* 855 (2012), pp. 695–741. DOI: 10.1016/j.nuclphysb.2011.10.021. arXiv: 1109.1536 [hep-ph].

[97] Matteo Cacciari et al. "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation". In: *Phys. Lett. B* 710 (2012), pp. 612–622. DOI: 10.1016/j.physletb.2012.03.013. arXiv: 1111.5869 [hep-ph].

[98] Peter Bärnreuther, Michal Czakon, and Alexander Mitov. "Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \to t\bar{t} + X$". In: *Phys. Rev. Lett.* 109 (2012), p. 132001. DOI: 10.1103/PhysRevLett.109.132001. arXiv: 1204.5201 [hep-ph].

[99] Michal Czakon and Alexander Mitov. "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels". In: *JHEP* 12 (2012), p. 054. DOI: 10.1007/JHEP12(2012)054. arXiv: 1207.0236 [hep-ph].

[100] Michal Czakon and Alexander Mitov. "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction". In: *JHEP* 01 (2013), p. 080. DOI: 10.1007/JHEP01(2013)080. arXiv: 1210.6832 [hep-ph].

[101] Michal Czakon, Paul Fiedler, and Alexander Mitov. "Total Top Quark Pair Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$". In: *Phys. Rev. Lett.* 110 (2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004. arXiv: 1303.6254 [hep-ph].

[102] Michal Czakon and Alexander Mitov. "Top++: A program for the calculation of the top-pair cross-section at hadron colliders". In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].

[103] M. Bähr et al. "Herwig++ physics and manual". In: *Eur. Phys. J. C* 58 (2008), p. 639. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[104] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[105] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[106] Stefano Frixione et al. "Single-top hadroproduction in association with a W boson". In: *JHEP* 07 (2008), p. 029. DOI: 10.1088/1126-6708/2008/07/029. arXiv: 0805.3067 [hep-ph].

[107] M. Aliev et al. "HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR". in: *Comput. Phys. Commun.* 182 (2011), pp. 1034–1046. DOI: 10.1016/j.cpc.2010.12.040. arXiv: 1007.1327 [hep-ph].

[108] P. Kant et al. "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions". In: *Comput. Phys. Commun.* 191 (2015), pp. 74–89. DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403 [hep-ph].

[109] ATLAS Collaboration. *Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO*. ATL-PHYS-PUB-2017-007. 2017. URL: https://cds.cern.ch/record/2261938.

[110] Ansgar Denner et al. "Electroweak radiative corrections to e+ e- —> nu anti-nu H". in: *Nucl. Phys. B* 660 (2003), pp. 289–321. DOI: 10.1016/S0550-3213(03)00269-4. arXiv: hep-ph/0302198.

[111] ATLAS Collaboration. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data". In: *JINST* 14 (2019), P12006. DOI: 10.1088/1748-0221/14/12/P12006. arXiv: 1908.00005 [hep-ex].

[112] ATLAS Collaboration. "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 81 (2021), p. 578. DOI: 10.1140/epjc/s10052-021-09233-2. arXiv: 2012.00578 [hep-ex].

[113] ATLAS Collaboration. "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 76 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

[114] 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].

[115] Matteo Cacciari, Gavin P. Salam, and Gregory 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].

[116] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet user manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[117] ATLAS Collaboration. "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Eur. Phys. J. C* 81 (2020), p. 689. DOI: 10.1140/epjc/s10052-021-09402-3. arXiv: 2007.02645 [hep-ex].

[118] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. ATLAS-CONF-2014-018. 2014. URL: https://cds.cern.ch/record/1700870.

[119] ATLAS Collaboration. "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 79 (2019), p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex].

[120] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017. URL: https://cds.cern.ch/record/2273281.

[121] ATLAS Collaboration. "ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset". In: (2022). arXiv: 2211.16345 [physics.data-an].

[122] ATLAS Collaboration. "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 78 (2018), p. 903. DOI: 10.1140/epjc/s10052-018-6288-9. arXiv: 1802.08168 [hep-ex].

[123] ATLAS Collaboration. *E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions*. ATLAS-CONF-2018-023. 2018. URL: https://cds.cern.ch/record/2625233.

[124] ATLAS Collaboration. *Early Inner Detector Tracking Performance in the 2015 Data at $\sqrt{s} = 13$ TeV*. ATL-PHYS-PUB-2015-051. 2015. URL: https://cds.cern.ch/record/2110140.

[125] Particle Data Group, R. L. Workman, et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[126] ATLAS Collaboration. "Tools for estimating fake/non-prompt lepton backgrounds with the ATLAS detector at the LHC". in: (2022). arXiv: 2211.16178 [hep-ex].

[127] Lorenzo Moneta et al. "The RooStats Project". In: (2011). arXiv: 1009.1003 [physics.data-an].

[128] Wouter Verkerke and David Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116 [physics.data-an].

[129] Roger J. Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[130] Mark J. Oreglia. "A Study of the Reactions $\psi' \to \gamma\gamma\psi$." PhD thesis. Stanford University, 1980.

[131] Tomasz Skwarnicki. "A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances". PhD thesis. Cracow, INP, 1986.

[132] ATLAS Collaboration. "Alignment of the ATLAS Inner Detector in Run-2". In: *Eur. Phys. J. C* 80 (2020), p. 1194. DOI: 10.1140/epjc/s10052-020-08700-6. arXiv: 2007.07624 [hep-ex].

[133] ATLAS Collaboration. "Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13\,\text{TeV}$ collected with the ATLAS detector". In: *Eur. Phys. J. C* 82 (2021), p. 95. DOI: 10.1140/epjc/s10052-021-09843-w. arXiv: 2109.10627 [hep-ex].

[134] ATLAS Collaboration. "Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 fb^{-1} of ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV". in: (Jan. 2023). arXiv: 2301.06319 [hep-ex].

[135] Andy Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *Eur. Phys. J. C* 75 (2015), p. 132. DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].

[136] Johannes Bellm et al. "Herwig 7.2 release note". In: *Eur. Phys. J. C* 80.5 (2020), p. 452. DOI: 10.1140/epjc/s10052-020-8011-x. arXiv: 1912.06509 [hep-ph].

Appendix A

Validation of the Pixel Readout RD53B ASIC Digital Design

A.1 The RD53B Readout Circuit

RD53B is a pixel readout chip framework that can be instantiated into different size physical chips. The design work and much of the verification are largely independent of the final instantiated size. RD53B consists of a *pixel matrix* and a *chip bottom*. The pixel matrix is built up of identical 8 by 8 pixel *cores* stepped and repeated in columns and rows. A core is physically 400 μm by 400 μm. The selected numbers of core columns and rows determine the chip size. The chip bottom contains all the system functionality and should be viewed as a fixed element that does not depend on matrix size. A physical chip, therefore, cannot be *narrower* than 20 mm (50 cores), because that is the width of the unique wire bonding pad frame in the chip bottom, but it can be wider. The *height* (number of core rows) is not constrained by the chip bottom, but is limited to a maximum of 50 by power and bias distribution as well as readout timing. This high level organization concept is shown in Figure A.1.

The core contains 64 pixel front ends organized in 16 identical so-called *analog islands* with 4 fronts ends each, which are embedded in a flat digital synthesized "sea" as shown in Figure A.2.

The chip bottom contains all system functionality and the wire bond pads. RD53B is a system-on-chip including power management, sophisticated digital communication, sensing and monitoring.

A.2 The Command Decoder

RD53B is fully controlled with a 160 Mbps differential serial input stream with a custom, DC-balanced encoding described in Sec. A.2.2. The differential receiver circuit is described

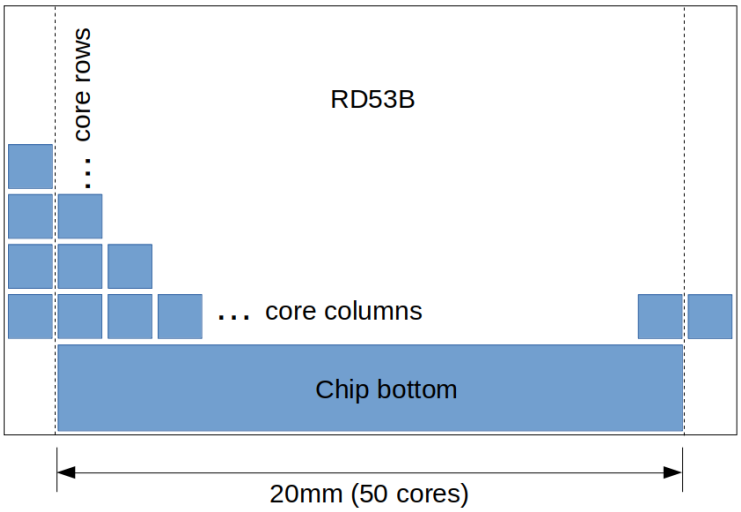


Figure A.1: Conceptual depiction of RD53B framework, with a matrix composed of 50 or more columns by up to 50 rows of identical cores, and a fixed chip bottom. The dashed lines indicate the minimum width of 50 cores.

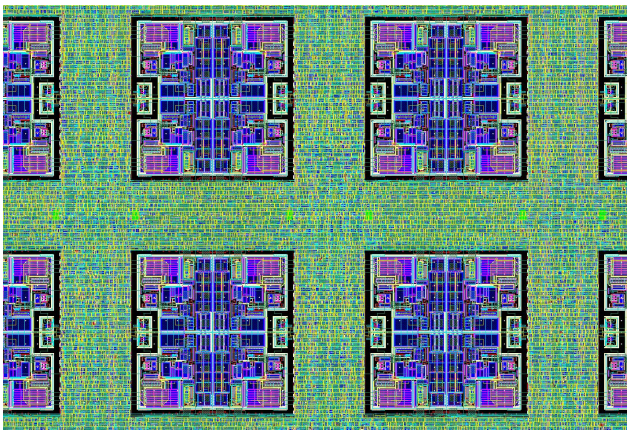


Figure A.2: Layout view of analog islands within synthesized logic. Four complete islands can be seen in the center of the figure. One core contains four by four analog islands.

in Sec. A.2.1. The received signal, without any processing, can be optionally repeated on the general purpose differential outputs. The command input also contains an activity detector that will cause a reset when the rate of transitions falls to very low value. A Clock and Data Recovery circuit (CDR) recovers the input bitstream and also produces the internal clocks for the chip, based on the transitions on input stream. A dedicated command (PLL_LOCK) equivalent to a clock pattern is provided to ease locking the internal phase locked loop.

A.2.1 Receiver Circuit

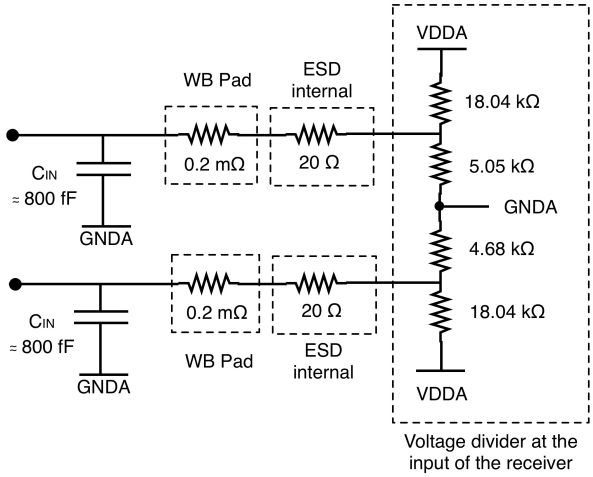


Figure A.3: Equivalent circuit for differential receiver input.

The CMD receiver is implemented as a differential amplifier with a rail-to-rail input stage. The inputs are connected to an on-chip resistor bias network in the kΩ range, which allows the receiver to be ac-coupled to the serial input stream. The resulting input common-mode voltage is $0.21\times$VDD_PLL (260 mV for nominal 1.2 V supply). The bias network also adds a small offset voltage to the differential input signal to keep the receiver in a static input state in case of a broken signal connection. A termination resistance is not implemented in the CMD receiver to allow multiple RD53B chips to be connected to the same CMD line (a multi-drop configuration with one termination at the end). The differential receivers for the data aggregation inputs use the same rail-to-rail input stage. The input impedance of the receiver, together with the ESD protection and wire bond pads, have been simulated with extracted parasitics. The capacitive and resistive contributions are shown Figure A.3.

A.2.2 Command Protocol

The input stream is a continuous sequence of commands. All commands are built in 16-bit frames made out of two 8-bit symbols. As the bitrate is 160 Mbps each frame spans four periods of the 40 MHz bunch crossing clock. Commands that are one frame long (four BX clocks) and are called short commands (Sec. refsec:rd53b:short-commands) of which Triggers are an example. Frames are interpreted one at a time and short commands are executed immediately, while long (multi-frame) commands (Sec. A.2.2) are executed after their last frame is received. Long commands have the property that they can be interrupted by short commands without the need of restarting the interrupted command. This gives the ability

to send Trigger commands (which are short) whenever needed, and to send long commands during data taking without worrying if trigger might be coming.

The command input is intended to be shared by multiple chips (multi-drop). Commands can be *broadcast*, in which case all chips sharing the command input will execute them, or *addressed*, in which case only the chip with the selected address will execute it and all other chips receiving it will ignore it. A chip can have one of 16 possible chip ID values (set by 4 wire bonds to ground overriding internal pull-up resistors). The first frame of addressed commands consists of an 8-bit symbol identifying which of the 7 commands it is, and a data 8-bit symbol specifying a chip ID. Addressed commands can also be sent in broadcast mode by specifying a chip ID value greater than 15. A chip that receives a command not broadcast or addressed to it will still process it (so as not to produce "unexpected data frame" errors), but will not execute it. The PLL_LOCK, Sync, and trigger command are always broadcast, while all others are addressed.

Each 16-bit frame is exactly DC balanced. DC balance is needed for A/C coupling, reliable transmission, and clock recovery. The symbols used also provide error detection [1]. There is a unique sync frame (used to perform frame alignment as explained in Sec. A.2.3), plus 3 kinds of TTC (Trigger, Timing and Control) frames: trigger, command, or data. TTC frames contain two 8-bit symbols which are themselves DC-balanced. Furthermore, symbols that begin or end with three or more 1's or 0's are not used, resulting in a maximum run length of 4, except for the sync frame which has a run length of 6. The valid symbols and commands are given in Tables A.1, A.3, and A.2. There is one sync frame, 7 non-trigger commands, 15 trigger symbols allowing the encoding of 15 trigger patterns (Tables A.1, A.3), and 32 data symbols allowing the encoding of 10 bits of content per data frame or 5 bits of chip ID per command frame (Table A.2). All valid symbols are allowed to be used as trigger tags in the trigger frame; thus there are 54 possible tags. A single bit flip always results in an invalid symbol (formally, all symbols are separated by a Hamming distance of 2).

RD53B interprets the protocol in three phases (which will be transparent to the user): Initialization A.2.3, Data Transmission A.2.4 and Decoding A.2.5. The decoding timing and exception handling are covered in Sec. A.2.6

Short Commands

PLL_LOCK (broadcast only):
This command allows a clock pattern to be sent to the chip without any action being executed by the command decoder. The clock pattern is needed to efficiently lock the Phase Locked Loop (PLL) to the correct frequency at start of operation. Once locked, the PLL no longer needs a perfect clock pattern and regular commands and

[1] All these properties could have been obtained with 8b/10b encoding, but the 10-bit frame length of 8b/10b would have required 200 Mbps link speed in order to maintain an integer number of bunch crossings per frame, as needed for synchronous triggering. The 160 Mbps bitrate of the RD53B custom protocol makes for better transmission on low mass cables and can be directly driven from GBT e-links.

Command	Encoding		(T)ag, (A)ddress or (D)ata 5-bit content			
Sync	1000_0001	0111_1110				
PLLlock	1010_1010	1010_1010				
Trigger	tttt_tttt	Tag[0..53]				
Read_trigger	0110_1001	ID<4:0>	00,T<7:5>	T<4:0>		
Clear	0101_1010	ID<4:0>				
Global Pulse	0101_1100	ID<4:0>				
Cal	0110_0011	ID<4:0>	D<19:15>	D<14:10>	D<9:5>	D<4:0>
WrReg(0)	0110_0110	ID<4:0>	0,A<8:5>	A<4:0>	D<15:11>	D<10:0>,0000
WrReg(1)	0110_0110	ID<4:0>	1,xxxx	xxxxx	N×(D<9:5>	D<4:0>)
RdReg	0110_0101	ID<4:0>	0,A<8:5>	A<4:0>		

Table A.1: List of protocol commands/frames and address or data fields associated with each. Unused padding bits are indicated by "0". Double vertical lines denote frame boundaries. tttt_tttt is one of 15 trigger commands (Table A.3). The before-encoded bit content of chip ID, Address or Data is shown. These are all encoded as 8-bit data symbols (Table A.2).

sync frames can be sent. This command can also be used as an idle when there is nothing to be sent during normal operation. This is equivalent to a No Operation (NOOP) command in many processors, but we do not use that terminology here. It repeats the same 8-bit symbol twice to produce a clock pattern (Table A.1).

Sync (broadcast only):
The Sync is the only command where the two 8-bit symbols used are not themselves DC balanced (both together the 16 bits are DC balanced). This is what makes it unique and allows it to be recognized for frame alignment.

Clear:
Clears the entire data path. All pending triggers and stored hits will be erased.

Global Pulse:
The global pulse command sends a single pulse with a duration of $2N$ bunch crossings, where N is the value of the 8-bit register GlobalPulseWidth. The value $N = 0$ is treated like $N = 1$. A pulse shorter than two bunch crossings is not possible. The global pulse can be routed to different places of the chip and has many uses. It can provide reset signals, control the ring oscillators, the ADC, etc.

Trigger (broadcast only):
Because one 16-bit frame spans 4 LHC bunch crossings, the trigger command must specify a 4-bit map indicating which of the 4 bunch crossings are actually triggered; hence 15 trigger patterns. The triggering is synchronous, and therefore trigger frames must be sent at specific times. The second symbol in a trigger frame can be any legal

Symbol Name	Encoding	Data Value	Symbol Name	Encoding	Data Value
Data_00	0110_1010	5'b00000	Data_16	1010_0110	5'b10000
Data_01	0110_1100	5'b00001	Data_17	1010_1001	5'b10001
Data_02	0111_0001	5'b00010	*Data_18*	*0101_1001*	5'b10010
Data_03	0111_0010	5'b00011	Data_19	1010_1100	5'b10011
Data_04	0111_0100	5'b00100	Data_20	1011_0001	5'b10100
Data_05	1000_1011	5'b00101	Data_21	1011_0010	5'b10101
Data_06	1000_1101	5'b00110	Data_22	1011_0100	5'b10110
Data_07	1000_1110	5'b00111	Data_23	1100_0011	5'b10111
Data_08	1001_0011	5'b01000	Data_24	1100_0101	5'b11000
Data_09	1001_0101	5'b01001	Data_25	1100_0110	5'b11001
Data_10	1001_0110	5'b01010	Data_26	1100_1001	5'b11010
Data_11	1001_1001	5'b01011	Data_27	1100_1010	5'b11011
Data_12	1001_1010	5'b01100	Data_28	1100_1100	5'b11100
Data_13	1001_1100	5'b01101	Data_29	1101_0001	5'b11101
Data_14	1010_0011	5'b01110	Data_30	1101_0010	5'b11110
Data_15	1010_0101	5'b01111	Data_31	1101_0100	5'b11111

Table A.2: List of command symbols used to encode data values.

symbol and is interpreted as one of 54 possible 6-bit tag bases to identify the trigger(s) in later readout. The mapping from symbol to tag base number is given in Table A.4. The trigger tag will be returned with the data corresponding to that trigger.

Symbol Name	Encoding	Trig. Pattern	Symbol Name	Encoding	Trig. Pattern
			Trigger_08	0011_1010	T000
Trigger_01	0010_1011	000T	Trigger_09	0011_1100	T00T
Trigger_02	0010_1101	00T0	Trigger_10	0100_1011	T0T0
Trigger_03	0010_1110	00TT	Trigger_11	0100_1101	T0TT
Trigger_04	0011_0011	0T00	Trigger_12	0100_1110	TT00
Trigger_05	0011_0101	0T0T	Trigger_13	0101_0011	TT0T
Trigger_06	0011_0110	0TT0	Trigger_14	0101_0101	TTT0
Trigger_07	0011_1001	0TTT	Trigger_15	0101_0110	TTTT

Table A.3: List of trigger symbols used to encode the 15 possible trigger patterns spanning four bunch crossings. Note there is no 0000 pattern as that is the absence of an trigger. The Trigger_01 (000T) means that the first bunch crossing of the trigger window is meant to be readout, and the extended tag returned will have 00 following the supplied tag base.

Tag values (decimal)	**Meaning**
0-215	extended tags from trigger command
216-231	Self triggers
216-219 (*)	Single bit-flip detected in tag symbol of a trig. command
220-223 (*)	unrecognized tag symbol
224-255	spares - not used in RD53B ATLAS

Table A.4: Possible extended tag values and their meaning. (*) The bit flip and unrecognized symbol values overlap the self-trigger tags in RD53B-ATLAS. This bug is corrected in future versions.

Long Commands

Cal (Calibration Injection):
The same command is used for both analog and digital injection. Whether injection will be analog or digital is decided by global configuration register CalibrationConfig, but the Cal command produces the same output regardless.

WrReg(0) (Write Register, single):
The WrReg command has two modes: single write and multiple writes to register 0. The command frame is the same and the distinction between single and multiple is made by the first bit of the payload (0=single, 1=multiple). The WrReg(0) or single has 9 bits of Address and 16 bits of Data. Up to 512 16-bit wide registers can be addressed, but not all 512 possible register addresses are used. If an attempt is made to write to an unused address, the command will do nothing and no warning will be generated. This command does not produce any output from the chip.

WrReg(1) (Write Register, multiple):
The WrReg command has two modes: single write and multiple writes to register 0. The command frame is the same and the distinction between single and multiple is made by the first bit of the payload (0=single, 1=multiple). The WrReg(1) or multiple must have the address value set to 0. It can only be used to initiate multiple writing to register 0. Register 0 is a virtual register called PIX_PORTAL, used to write and read pixel configuration (Sec. A.2.8).

Following a WrReg(1) command, one may send data frames to the chip (as many as desired) without any preceding command. All these data frames will be written to the PIX_PORTAL (register 0). This must be done in conjunction with auto-increment (see Sec. A.2.8). This permits very efficient transfer of data to the PIX_PORTAL. The write multiple mode remains in effect until a new long command is received (short commands will be executed and not end the WrReg(1) command mode). Note that

the chips not addressed by the WrReg(1) command will still recognize the multiple write mode and will therefore not issue "unexpected data frame" warnings, but they will not write the data to their register 0. The placement of the 10 bits from each data frame into the 16 bits of register 0 is described in Sec. A.2.8

RdReg (Read Register):
This command has 9 bits of address and no data. It initiates the readout of the addressed register. Address 0 is special: it is the the pixel register as described in Sec. A.2.8. The 16-bit register value is returned in the data stream. Not all 512 possible register addresses are used. If readback of an unused address is requested, the data value returned will be 0, the address returned will be the requested (non-existent) one, without any warning generated.

RdTrig (Read Trigger):
This command has an 8-bit extended tag value. In two-trigger mode, it selects a previously received tag for readout. It is not useful in single trigger mode.

A.2.3 Command Protocol Initialization

Until the PLL is locked and produces a stable chip clock, the command decoder will be in its reset state. During this period, PLL_LOCK frames should be sent to the chip. The transitions in the string of PLL_LOCK frames will allow the clock recovery circuit to lock to the correct 160 MHz frequency. The user does not know when the PLL has locked, but simply sends PLL_LOCK frames for a long enough time that the lock cycle is surely completed. For debugging, the PLL lock condition can be observed in the recovered CMD output of the general purpose LVDS, which is a default output. At this point the protocol initialization begins. Before any command decoding, the input bitstream is processed by the Channel Synchronizer circuit (FigureA.4), and the initialization correctly sets up this circuit.

The sync pattern (Table A.1) can not be produced through any combination of TTC frames and therefore can be searched for to lock the correct frame boundaries (the search procedure is explained in the next paragraph). Sync frames must be sent at the start of operation so that the framing can be locked (this different from PLL lock!). It is mandatory to send one sync frame in every 32 frames or so in order to maintain lock or allow the command decoder to re-lock if lock was lost. If no sync frames are received in a long time frame lock will be declared lost and the command decoder will stop interpreting commands until a new lock is acquired. Typically at the start of operation (power up) there are no commands or triggers to immediately send, and so sending a large number of sync frames to ensure initial lock is not a problem. The channel synchronizer lock is available in the chip status CMOS output.

Using the 160 MHz recovered clock, the channel synchronizer will search for sync symbols and count each valid appearance of this pattern in 16 separate channels (one channel for each

possible frame alignment). When the count for one of the channels, i, reaches a threshold N_{lock}, *sync lock* is declared as acquired, channel i is adopted as the correct channel, and the count of the remaining 15 channels is reset. The value N_{lock} has default value of 16 and can be changed in configuration register ChSyncConf. At the start of transmission the command decoder will not interpret any commands until it has received N_{lock} Sync commands. Thus one should begin transmission by sending at least N_{lock} Sync commands. The 40 MHz bunch crossing clock is generated as the bit pattern 1100110011001100 aligned to channel i. Thus there are 4 bunch crossings with a fixed phase relationship to the sync frame, which can be labeled $BXa..BXd$. The counting of sync sequences continues in all the channels, but every new sync sequence detected on the lock channel i resets the count for all the other channels. If the count for a channel that is not the lock channel ever reaches a threshold of $N_{lock}/2$, lock is declared lost, and a new sync lock is acquired on the first channel that reaches the locking threshold N_{lock}. This allows for continuous channel monitoring and automatic sync lock as long as enough sync symbols are transmitted. Additionally, if zero sync frames are received in the lock channel within 64 frames (regardless of other channels), lock will be declared lost and no further commands will be decoded until a new lock is acquired. This value is hard-wired and cannot be changed. This is useful to prevent prolonged, random input due to an upstream exception from corrupting the chip operation, but makes it mandatory to regularly send Sync symbols.

A.2.4 Command Protocol Transmission

During transmission a correct sequence of commands is sent to control the chip. Trigger frames are sent at specific times, and the "space between trigger frames" is filled with commands (including the required Syncs). Long commands are decoded regardless of intervening short commands. The PLL_LOCK command can be used as an idle frame, as it has the most transitions and will therefore best maintain PLL operation. Sync commands can also be used as idles, since they must be sent periodically anyway, but they have the fewest transitions, so are not ideal for maintaining PLL lock. The best approach is therefore to always send Syncs every 32 frames and PLL_LOCK commands in between if and when there is nothing else to send.

A.2.5 Command Protocol Decoding

The data bits recovered from the locked channel are fed to the Command Decoder as shown in Figure A.4. In the absence of a sync lock, nothing is fed to the command decoder, so until a lock happens no commands will be interpreted. The locked condition guarantees that the bits fed to the command decoder are correctly aligned with the 40 MHz bunch crossing clock. Protocol consistency is ensured by checking that the decoded frames are valid and also that they match what is expected (analogous to checking both spelling and grammar). The 16 bits are fed to the command decoder with a parallel bus. In case of correct detection,

the indicated action is performed according to the command type and Chip ID. All symbols are always checked and decoded, even if they follow a Chip ID that does not match the wire bonded ID. However, the Command Decoder will act on the rest of the chip only if the command is a trigger, if decoded Chip ID matches the wire bonded ID, or if the decoded broadcast bit is 1 (the PLL_LOCK command is not addressed, but has no internal action-no operation). The detection of an invalid symbol is handled differently depending on the frame and expectation (current state). The handling of exceptions is shown in Table A.5.

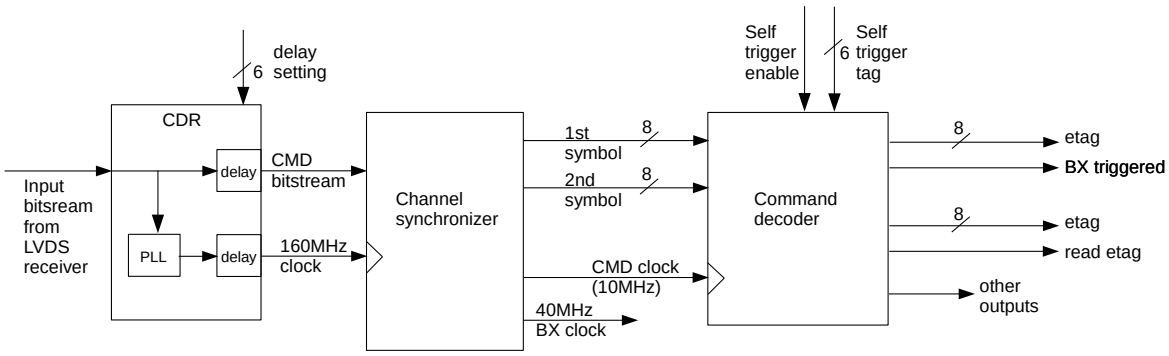


Figure A.4: Clock and command recovery and decoding path from chip input to internal signals, showing trigger pluses and tags in particular. Other outputs of the command decoder, such as global register address and write signal, not shown. 16-bit Command patterns are successively loaded into the Command Decoder with the correct frame alignment as determined by the Channel Synchronizer.

A.2.6 Command Protocol Timing

The decoded commands are executed 25 ns after the end of the last frame of the command data. "Executed" means that the outputs of the Command Decoder block in Figure A.4 change state, which happens on a rising edge of the beam clock. In many cases the execution is instantaneous (outputs change state and that is it), but the Trigger, Cal and Global Pulse commands have a delay and duration. The trigger command sends 1 to 4 pulses in 4 consecutive beam clock cycles, and thus is completely finished before a new command can be completely received (since 1 frame is 4 beam clock cycles). The Cal and Global pulse commands can occupy their respective output lines (CAL_edge, CAL_aux, and Global_pulse) for many clock cycles. A new Cal or Global pulse command should not be sent before the prior such command is complete (up to the DAQ to ensure this), but any other command can be sent and will be executed normally.

Frame received	**Frame Expected**	**Error/Action**1
invalid, data	data	Aborted command
data, invalid	data	Aborted command
invalid, invalid	data	Aborted command
invalid, data	not data	Lost trigger
invalid, invalid	not data	Corrupted frame
invalid, sync	any	Corrupted sync
sync, invalid	any	Corrupted sync
invalid, command	any	Execute with warning
command, invalid	any	Execute with warning
trigger, bit-flip (*)	any	Execute w/tag base 54
trigger, invalid (*)	any	Execute w/tag base 55
command, command	data	Ignored command

Table A.5: Command Decoder response to invalid or unexpected symbols. (*) bit-flip refers to an 8-bit pattern produced from flipping a single bit in a valid symbol, while invalid references to any other invalid 8-bit pattern.

A.2.7 Global Configuration

The global configuration is stored in 16-bit registers which are accessed like a RAM with the write and read register commands of Table A.1. Each register has a default value.

A.2.8 Pixel Configuration

Each pixel has 8 bits of local configuration. From the point of view of the write and read register commands, each pixel is seen as one half of one configuration data register. All pixels are paired.

The 8 pixel bits are divided into 5 TDAC bits (threshold tuning bits) and 3 enable bits (also known as mask bits). These two types of bits can be written together or independently (always for two pixels at a time). Thus one can choose to write all 8 bits at once, only the 5 TDAC bits, or only the 3 enable bits. The single write register command (WrReg(0)) of Table A.1 always writes al 8 bits of both pixels, where the 16 bit data frame is subdivided as follows:

Single Write:

left-pixel(TDAC[15:11]HitBus[10]InjEn[9]Enable[8]),
right-pixel(TDAC[7:3]HitBus[2]InjEn[1]Enable[0])

The multiple write register command (WrReg(1)) instead writes the mask bits or the TDAC bits depending on the Mask or TDAC bit of global configuration register PIX_MODE. The

mapping from 10-bit data frame to two pixel TDAC or mask bits is as follows:

PIX_MODE[1] = 0: unused[9:8], right-pixel-mask[7:5], unused[4:3], left-pixel-mask[2:0]
PIX_MODE[1] = 1: right-pixel-TDAC[9:5], left-pixel-TDAC[4:0]

Internally, the writing and reading of configuration values from the pixels uses an addressed bus to every 2×1 pixel pair. All reading and writing is done two pixels at a time in a given column of 4-pixel regions. However, multiple core columns can be written in parallel, while readback can only take place from one pixel-pair column at a time. There are thus two write modes, single pixel-pair and broadcast, while read is always single pixel-pair.

The write and read operations are controlled by three global registers, the REGION_COL, REGION_ROW, and PIX_MODE configuration registers. The pixel data is written into or retrieved from global register 0 (PIX_PORTAL) with the normal write and read register commands (see Sec. A.2.2). This is a virtual register acting as a portal to whatever pixel pair is pointed to by the column and row configuration registers (called PIX_PORTAL). The row register has a special feature called auto increment (Auto Row), which reduces the number of commands needed to fully configure the chip. This mode is enabled by a configuration bit and increments the row register value after every write or read operation to PIX_PORTAL.

The typical pixel matrix configuration write sequence, using the write single register command, is given in Table A.6. Note that this takes 77200 (73008) commands for ATLAS (CMS) chips to accomplish. These numbers should be multiplied times 4 to obtain number of frames, and each frame takes 100 ns to transmit. If one is only configuring a chip, it will therefore take about 30 ms. For the case of configuring during data taking (called trickle configuration), much of the command bandwidth will be taken up by trigger commands, and configuration will therefore take longer. The worst case is two-level trigger operation with 4 MHz L0 + 1 MHz L1 trigger operation. This will use up 60% of the command bandwidth. We should also remember that 6% of the command bandwidth must be used to send periodic Sync commands. with only 34% of the command bandwidth available, 77200 Write Register commands will take 88 ms instead of 30 ms. With some DAQ overheads we assume 100 ms. So for a 4-chip module trickle configuration in the worst case will take 400 ms. (If more chips share the same command line it will take proportionally longer). Writing a uniform (all pixels the same) configuration is 50 (54) times faster for ATLAS (CMS), because each Write Pixel command can write to all core columns Table A.7. Alternatively, using the multiple instead of single Write Register command means one frame instead of 40 frames per write, which will reduce the above 88 ms to 22 ms (100 ms for a 4-chip module in worst case of trickle configuration). The readback of the pixel configuration for the whole matrix can proceed exactly as shown in Table A.6, substituting the Read Register command instead of Write Register. This can be carried out in broadcast mode to any number of chips in parallel, so will always take 50 ms (half as much as writing a single chip because the read register command is two frames instead of four).

Writing or reading an individual, arbitrary pixel pair follows steps 1-3 of Table A.6. For calibration operations it is often required to write only the mask bits many times to shift a pattern through the matrix, leaving the TDAC bits alone. This can be done with broadcasted commands (same mask for all chips even though the TDACs are different), and it must use the write multiple command as the write single command always writes all the 8 configuration bits per pixel. Writing masks to a single pixel at-a-time will take 77200 write operations as in Table A.6, but one frame per pixel write instead of four (still four frames per write for steps 1 and 2), resulting in 78400 frames, which takes 8.3ms if all the command bandwidth (minus 6% for Syncs) is used, or 23 ms in the worst case of trickle calibration. Writing one row at a time will take 1/50 of this per mask, regardless of the number of chips, as it is done in column broadcast mode using broadcasted commands. It can even be faster if not all rows need a new mask each time.

Step	Command	Address	Explanation
1	Write_Register	column and mode config	set columns 0-1 and auto row mode
2	Write_Register	row config	set row 0
3	Write_Register	0	config first 2 pixels
4	Write_Register	0	config for next row 2 pixels
386	Write_Register	0	config for last row 2 pixels in cols 0-1
387	Write_Register	column and mode config	set columns 2-3 and auto row mode
388	Write_Register	row config	set row 0
389	Write_Register	0	config for next row 2 pixels
77200	Write_Register	0	config last 2 pixels in chip

Table A.6: Sequence to write an arbitrary pixel configuration to ATLAS size chip using write register single commands. Each column pair takes 386 commands, times 200 column pairs leads to 77200 commands. For readback replace Write_Register 0 with Read_Register 0 commands.

When used in a radiation environment it is possible to write to an non-existing pixel address at the end of a configuration operation. This will make pixel configuration less sensitive to accidental SEU/SET caused overwriting a pixel register.

A.3 Digital Validation

The complexity of the RD53B chip and the challenging specifications of operation are such that a traditional design and verification approach, based only on directed tests and targeted architectural simulations, is no longer sufficient. Digital architecture optimization and functional verification are among the greatest challenges of the RD53B demonstrator

Step	Command	Address	Explanation
1	Write_Reg.	column and mode config	set broadcast, cols. 0-1, and auto row mode
2	Write_Reg.	row config	set row 0
3	Write_Reg.	0	config all pixels, first row in cols 0-1
4	Write_Reg.	0	config all pixels, second row
386	Write_Reg.	0	config all pixels, last row
387	Write_Reg.	column and mode config	set broadcast, cols. 0-1, and auto row mode
388	Write_Reg.	row config	set row 0
389	Write_Reg.	0	config all pixels, first row in cols 2-3
1544	Write_Reg.	0	config last 2 pixels in cols 6-7

Table A.7: Sequence to write a default (all pixels the same) configuration for ATLAS size chip. Only the first core column (columns 0-7) are written because all core columns will be "CC-ed" in parallel.

design. To this purpose, a simulation and verification platform called VEPIX53 (Verification Environment for RD53 PIXel chips) was developed addressing both these issues. The verification platform supports the definition of ad-hoc directed tests, which are still useful to address specific features and blocks, in a trade-off between comprehensive verification and time constraints. VEPIX53 is based on consolidated high level tools and methodologies coming from the industry domain: the hardware description and verification language SystemVerilog (SV), which has been itself defined as a Verilog extension; and the Universal Verification Methodology (UVM) library. The former provides both enhancements for the description of the Design Under Test (DUT) at different levels of abstraction and advanced verification features; the latter is based on a documented set of standard classes for the building blocks of the environment. VEPIX53 provides a set of dynamic components that are reused in order to support the design flow at different steps, from initial architectural modeling to the verification of the final design, providing: the generation of different types of input pixel hits including Monte Carlo physics data, the possibility of simulating designs under test (DUTs) at different levels of abstraction and automated verification features e.g. pixel chip output prediction, conformity checks, statistics and coverage collection and configurable reporting.

A.3.1 VEPIX53 Framework

The VEPIX53 framework, represented in Figure A.5, consists of a top level testbench which contains the DUT (wrapped in a top level harness module) and different UVM verification components (UVCs); these are instantiated and configured according to the test scenario, which is specified by the different tests that are run (defined in the test library).

UVCs communicate between each other using transaction class objects. Most of them are associated to pixel chip interfaces and can be configured either as active or passive, depending on the interfaces being related to the input or the output of the chip, respectively. An interface UVC generally contains sequencer and driver components that are in charge of generating the input stimuli to the DUT: this takes place in a layered fashion through sequences of transactions, specified in the tests, that are translated into physical signals. There are also monitor and subscriber components that work the other way around, i.e. they create transactions according to the physical signals in the interface for checks, coverage collection and forwarding to other parts of the testbench. When the interface UVC is configured as active, all the mentioned components are instantiated inside it; in case of a passive components, instead, the stimuli generation is missing. The interface UVCs are shared and versioned in a generic repository (not RD53B-specific), as they are reused for different purposes in multiple testbenches. Along with the interface UVCs there are module UVCs, which are specific to a pixel chip and contain predictors and checkers associated to single building blocks.

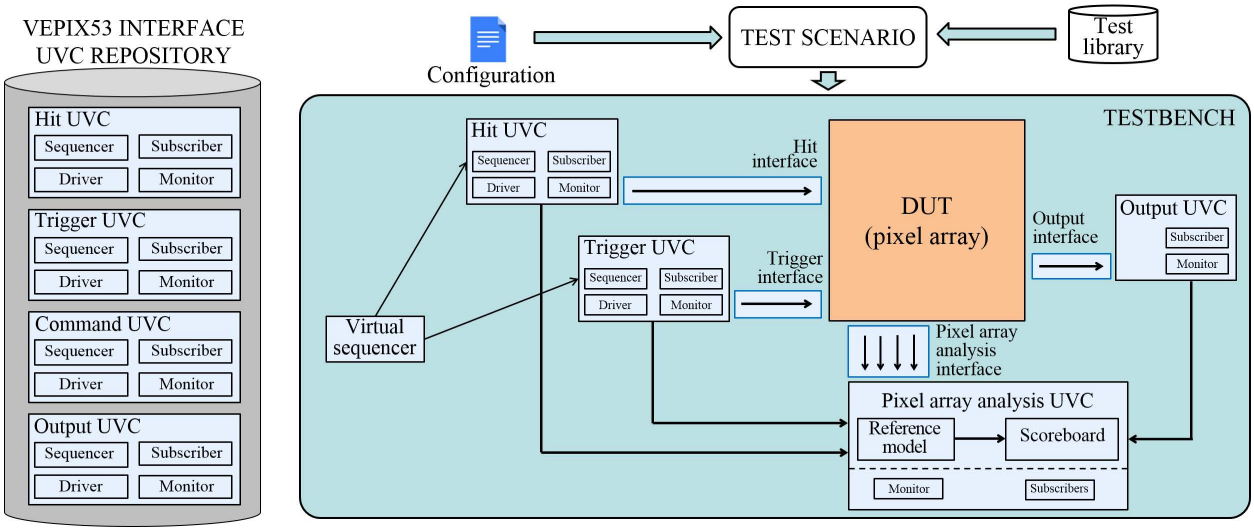


Figure A.5: Block diagram of the VEPIX53 testbench configured for architecture simulation.

Here follows a description of the VEPIX53 UVCs:

- the *hit UVC*, associated to the hit interface, has the main function of generating the charge signals associated to particles crossing the detector, and injecting them into the pixel matrix. The input hits can be either generated in a constrained-random fashion, according to a set of pre-defined classes of clustered hits, or read from physics data in ROOT format produced by Monte Carlo pixel detector simulations. It is also possible to mix the externally sourced hit data with the internally generated ones;

- the *trigger UVC*, associated to the trigger interface, is in charge of generating the external trigger signal of the pixel array according to configurable trigger rate and latency;
- the virtual sequencer controls the coordinated generation of hit and trigger transactions;
- the *output UVC*, associated to the pixel array output interface, takes care of producing data transactions by monitoring the data at the output of the pixel array;
- the *pixel array analysis UVC* is the module UVC associated to the whole pixel array. It contains a reference model that predicts the pixel array output according to the monitored hit and trigger transactions (it is, in practice, a transaction level description of the pixel array used as a golden reference for the DUT); a scoreboard that checks for conformity between predicted and actual output; additional components, used for performance assessment, that monitor internal signals of the pixel array and keep track of its status. The main features of the pixel array analysis UVC are generic and can be seamlessly used up to full chip top-level post-layout simulations. Monitoring of the internal signals for detailed performance studies, requires instead to modify signal probing performed at top-level, as the hierarchical paths are modified by synthesis;
- the *command UVC* and the *Aurora UVC* are additionally defined for the functional verification of the pixel chip: the command UVC is in charge of generating the input command stream of the chip in agreement with a dedicated serial protocol and the Aurora UVC monitors data transactions at the pixel chip output, encoded with the Xilinx Aurora protocol.

A.3.2 Validation of the Command Decoder

The VEPIX53 testbench for verifying the RD53B demonstrator reuses and extends the environment implemented for the architecture study, which was dedicated to the pixel array logic only. When taking into account the full chip, as shown in the digital functional block diagram of Figure A.6, there are additional building blocks to be verified: the array includes also pixel configuration, calibration injection pulse generation circuitry, and an additional HitOR channel, while the chip periphery contains control and readout logic and global configuration registers. This section focuses only on the verification and validation of the command decoder block.

The command interface is associated to the serial 160 Mbps input stream of the chip, which follows a dedicated, DC-balanced custom protocol that encodes clock, trigger and other commands on a single link and features built-in framing and error detection. A corresponding command UVC has been implemented. Its main function is to generate, drive and monitor all the possible command types, among which are synchronization pulse, calibration pulse and

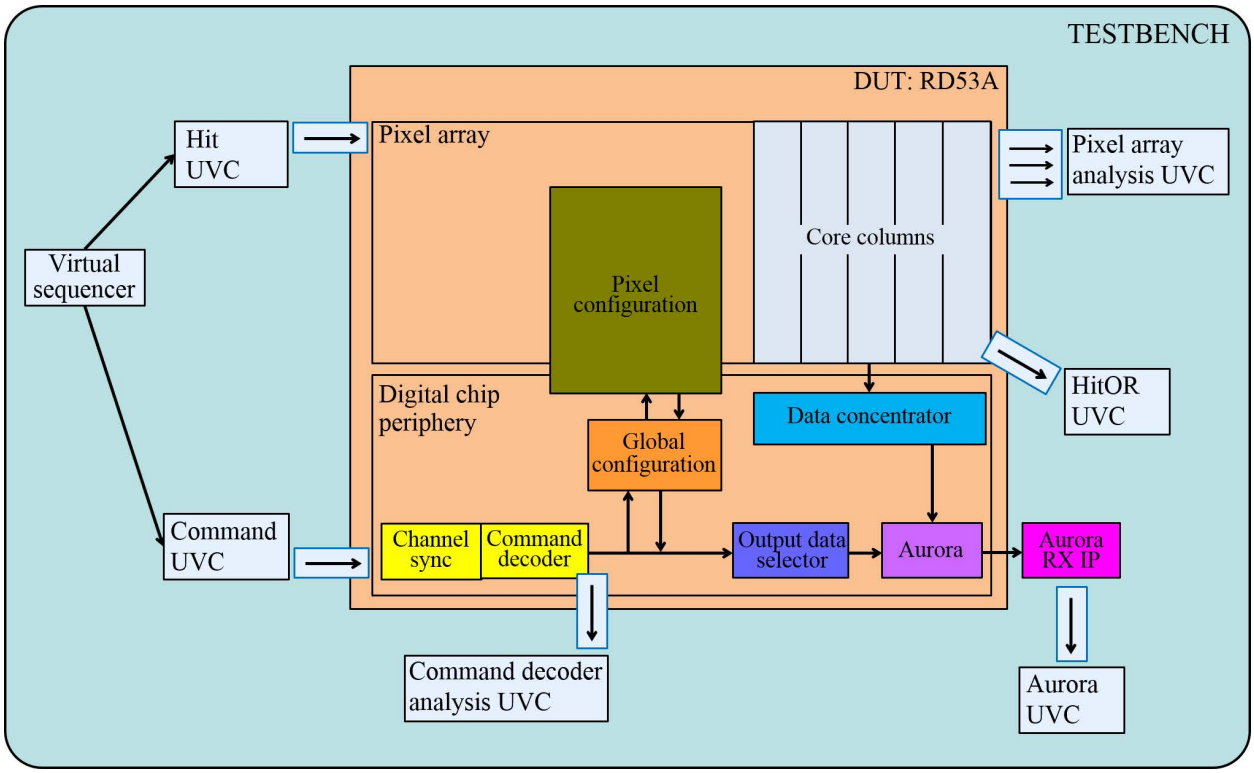


Figure A.6: Block diagram of the VEPIX53 testbench configured for RD53B functional verification.

read/write configuration registers; moreover, it extends the functionality of the pre-existing trigger UVC as the driver encodes incoming trigger transactions into high priority input commands to the chip. The chip output encodes pixel data, configuration data and messages on 1 to 4 serial links (programmable) at 1.28 Gbps nominal bandwidth using the Xilinx Aurora 64b/66b protocol. The associated Aurora UVC monitors the chip output decoded by a Xilinx IP receiver module and builds corresponding hit transactions or monitoring data (configuration or messages) transactions. The additional HitOR output of the chip is monitored as well with a dedicated UVM component. From the module UVC point of view, a simple monitor and scoreboard have been implemented for automatic verification of the RD53B command decoder module and its error correction functions.

Both constrained-random and directed tests were written with respect to a detailed verification plan document which lists in a prioritized fashion all the main functions and test cases to be verified for the command decoder, from the single building block to the full chip level. Directed tests were written to to test very specific functions of the command decoder: reading and writing to a register, triggering and reading out the pixel matrix, testing the digital calibration injection circuit. These tests were also written to test the handling of an

incorrect data stream fed into the block: single bit flips, multiple bit flips, incorrect data frames. Finally, constrained-random tests were written to stress the digital design of the circuit to validate proper handling of a large thorough-put of valid and invalid data frames, and subsequent recovery of the chip. An example of a constrained-random test written for the command decoder digital block is shown in Figure A.7 using Cadence's SimVision software program. The results of these tests were communicated to the digital design engineers which could then use as input to fix, update and upgrade the design of the command decoder.

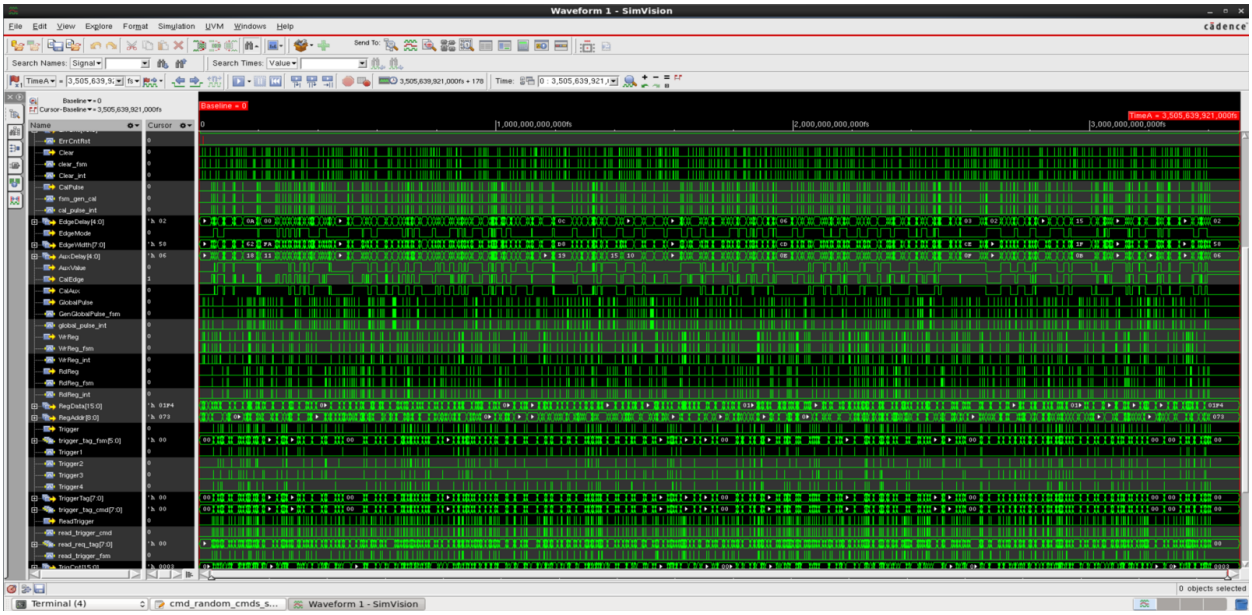


Figure A.7: Example of a constrained-random test written and executed to validate the design on the Command Decoder digital block.

The tests were first run at RTL, both for single pixel core columns and full pixel array, to find possible bugs in the logic. Then full chip test regressions containing a set of the most relevant functional tests (e.g. processing of random hits and triggers, digital injections, read back of configuration data) were run at gate level in order to verify the chip including timing back-annotation for three different simulation corners. For the final debug of the chip several design iterations needed to be carried out, with fixes applied both to functional bugs and timing issues not covered by static timing analysis (e.g. reset strategy, signals crossing asynchronous clock domains, etc.). Test regressions were repeated at both RTL and gate level at each iteration until receiving an all-pass result.