

Search for dark Higgs boson production in  $pp$  events at  $\sqrt{s} = 13$  TeV with the  
ATLAS detector at the LHC

by

Benjamin Wright  
B.Sc., Queen's University, 2019

A Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Physics and Astronomy

© Benjamin Wright, 2021  
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by  
photocopying or other means, without the permission of the author.

Search for dark Higgs boson production in  $pp$  events at  $\sqrt{s} = 13$  TeV with the  
ATLAS detector at the LHC

by

Benjamin Wright  
B.Sc., Queen's University, 2019

Supervisory Committee

---

Dr. J. Albert, Supervisor  
(Department of Physics and Astronomy)

---

Dr. R. Sobie, Departmental Member  
(Department of Physics and Astronomy)

## Supervisory Committee

---

Dr. J. Albert, Supervisor  
(Department of Physics and Astronomy)

---

Dr. R. Sobie, Departmental Member  
(Department of Physics and Astronomy)

## ABSTRACT

This thesis presents the design of a signal region, as well as two control regions to constrain the  $t\bar{t}$  background, in the search for the theorized dark sector process  $q\bar{q} \rightarrow Z'^* \rightarrow Z's$  with  $s \rightarrow WW \rightarrow q\bar{q}\ell\nu$  and  $Z' \rightarrow \chi\chi$ (invisible). In this process, dark matter is produced as a new heavy vector boson,  $Z'$ , together with a scalar “dark Higgs” boson,  $s$ , that decays to a pair of  $W$  bosons with a semileptonic final state. In this model, a new Higgs mechanism associated with  $s$  generates the masses of the dark sector particles, and the mixing of  $s$  with the Standard Model Higgs boson allows it to decay to Standard Model particles and be observable over a large range of model parameter space at the Large Hadron Collider. This search uses  $139 \text{ fb}^{-1}$  of proton-proton collision data collected by the ATLAS experiment at a center-of-mass energy of 13 TeV. Using the designed signal and control regions, in this search we will be able to exclude  $Z'$  masses up to a range between approximately 1200 to 2200 GeV, and  $s$  masses from approximately 150 GeV up to a range between approximately 290 to 360 GeV.

# Contents

|   |             |
|---|-------------|
| <b>Supervisory Committee</b>                                | <b>ii</b>   |
| <b>Abstract</b>   | <b>iii</b>  |
| <b>Table of Contents</b>                                    | <b>iv</b>   |
| <b>List of Tables</b>                                       | <b>vii</b>  |
| <b>List of Figures</b>                                      | <b>viii</b> |
| <b>Acknowledgements</b>                                     | <b>xii</b>  |
| <b>Declaration of Contributions</b>                         | <b>xiii</b> |
| <b>1 Introduction</b>                                       | <b>1</b>    |
| <b>2 Theory</b>   | <b>4</b>    |
| 2.1 The Standard Model . . . . .                            | 4           |
| 2.1.1 The Particles . . . . .                               | 4           |
| 2.1.2 Quantum Electrodynamics (QED) . . . . .               | 6           |
| 2.1.3 Weak Interactions . . . . .                           | 6           |
| 2.1.4 Quantum Chromodynamics (QCD) . . . . .                | 7           |
| 2.1.5 The Higgs Boson . . . . .                             | 8           |
| 2.2 Beyond the Standard Model . . . . .                     | 9           |
| 2.2.1 Dark Matter . . . . .                                 | 10          |
| 2.2.2 Dark Higgs Boson Model . . . . .                      | 11          |
| <b>3 The Large Hadron Collider and the ATLAS Experiment</b> | <b>14</b>   |
| 3.1 The Large Hadron Collider . . . . .                     | 14          |
| 3.2 The ATLAS Experiment . . . . .                          | 16          |

|          |  |           |
|----------|--|-----------|
| 3.2.1    | Inner Tracking Detector . . . . .                      | 18        |
| 3.2.2    | Calorimeters . . . . .                                 | 19        |
| 3.2.3    | Muon Spectrometers . . . . .                           | 22        |
| 3.2.4    | Trigger and Data Acquisition . . . . .                 | 23        |
| <b>4</b> | <b>Analysis Preparation</b>                            | <b>24</b> |
| 4.1      | Signal Model . . . . .                                 | 24        |
| 4.2      | Monte Carlo Production and Data . . . . .              | 26        |
| 4.2.1    | Signal MC . . . . .                                    | 26        |
| 4.2.2    | Background MC . . . . .                                | 27        |
| 4.2.3    | Data . . . . .   | 28        |
| 4.3      | Object Definition . . . . .                            | 29        |
| 4.3.1    | Muons . . . . .  | 29        |
| 4.3.2    | Electrons . . . . .                                    | 30        |
| 4.3.3    | Small-radius ( $R = 0.4$ ) Jets . . . . .              | 31        |
| 4.3.4    | Track-Assisted-Reclustered Jets . . . . .              | 32        |
| 4.3.5    | Missing Transverse Momentum ( $E_T^{miss}$ ) . . . . . | 34        |
| 4.4      | Data and MC Preparation . . . . .                      | 35        |
| <b>5</b> | <b>Analysis</b>  | <b>37</b> |
| 5.1      | Cut and Count Analysis . . . . .                       | 37        |
| 5.2      | Analysis Strategy . . . . .                            | 38        |
| 5.3      | Merged Signal Region . . . . .                         | 39        |
| 5.3.1    | Signal and Background Characterization . . . . .       | 39        |
| 5.3.2    | TAR Jet Lepton Disentanglement . . . . .               | 39        |
| 5.3.3    | Merged Signal Region Optimization . . . . .            | 42        |
| 5.4      | Resolved Signal Region . . . . .                       | 47        |
| 5.5      | $t\bar{t}$ Control Regions . . . . .                   | 47        |
| 5.6      | $W$ +jets Control Region . . . . .                     | 50        |
| 5.7      | Trigger Strategy . . . . .                             | 51        |
| <b>6</b> | <b>Results</b>   | <b>55</b> |
| 6.1      | Statistical Framework . . . . .                        | 55        |
| 6.2      | Minimum $m_s$ . . . . .                                | 57        |
| 6.3      | Systematic Uncertainties . . . . .                     | 60        |
| 6.3.1    | $R = 0.4$ Jets . . . . .                               | 60        |

|          |  |           |
|----------|--|-----------|
| 6.3.2    | Track Uncertainties . . . . .                            | 60        |
| 6.3.3    | Muon and Electron . . . . .                              | 60        |
| 6.3.4    | $E_T^{miss}$ . . . . .                                   | 60        |
| 6.3.5    | Luminosity . . . . .                                     | 61        |
| 6.4      | Fit Results . . . . .                                    | 61        |
| 6.4.1    | Background Only Fit . . . . .                            | 61        |
| 6.4.2    | Exclusion Fit . . . . .                                  | 67        |
| <b>7</b> | <b>Conclusions</b>                                       | <b>70</b> |
| <b>A</b> | <b>Additional Information</b>                            | <b>72</b> |
| A.1      | Preselection level plots . . . . .                       | 72        |
| A.2      | Data-MC Comparisons in Control Regions . . . . .         | 76        |
| A.2.1    | $t\bar{t}$ Control Regions . . . . .                     | 76        |
| A.2.2    | $W$ +jets Control Regions . . . . .                      | 80        |
| A.3      | CR-SR Distribution Comparisons . . . . .                 | 84        |
| A.4      | $N - 1$ Plots - Highly Weighted Event Included . . . . . | 87        |
|          | <b>Bibliography</b>                                      | <b>89</b> |

# List of Tables

|           |  |    |
|-----------|--|----|
| Table 4.1 | Muon object definition criteria . . . . .  | 30 |
| Table 4.2 | Electron object definition criteria . . . . .  | 31 |
| Table 4.3 | $R = 0.4$ jets object definition criteria . . . . .  | 32 |
| Table 4.4 | TAR jet reconstruction parameters . . . . .  | 33 |
| Table 5.1 | Optimized selection criteria for the “merged” category signal region.  | 45 |
| Table 5.2 | Overview of “resolved” SR cuts. . . . .  | 47 |
| Table 5.3 | Summary of differences in selections on $N(b\text{-jets})$ and $\mathcal{S}$ between<br>signal regions and $t\bar{t}$ control regions. . . . .   | 50 |
| Table 5.4 | Summary of differences in selections on $\Delta R(W_{\text{Cand}}, \ell)$ and $\mathcal{S}$ be-<br>tween signal regions and $W$ +jets control regions. . . . .                             | 51 |
| Table 5.5 | ATLAS unprescaled $E_T^{\text{miss}}$ triggers used in Run-2 data-taking. . . . .  | 52 |
| Table 5.6 | ATLAS single muon triggers used in Run-2 data-taking. . . . .  | 52 |
| Table 6.1 | Pre and post-fit yields for each background category in the merged<br>signal and control regions for a background-only fit. . . . .  | 62 |
| Table 6.2 | Pre and post-fit yields for each background category in the re-<br>solved signal and control regions for a background-only fit. . . . .  | 62 |
| Table 6.3 | Effect of systematic uncertainties in the merged signal and control<br>regions for a background-only fit. Uncertainties with a magnitude<br><1% of total yield are excluded. . . . .       | 65 |
| Table 6.4 | Effect of systematic uncertainties in the resolved signal and con-<br>trol regions for a background-only fit. Uncertainties with a mag-<br>nitude <1% of total yield are excluded. . . . . | 66 |

# List of Figures

|            |   |    |
|------------|---|----|
| Figure 2.1 | The elementary particles of the Standard Model [7]. . . . .   | 5  |
| Figure 2.2 | The fundamental QED vertex. . . . .   | 6  |
| Figure 2.3 | Some fundamental weak interaction vertices. . . . .   | 7  |
| Figure 2.4 | Some fundamental QCD interaction vertices. . . . .  | 8  |
| Figure 2.5 | Shapes of the Higgs potential for different cases of $\mu^2$ . [16] .   | 9  |
| Figure 2.6 | The Bullet Cluster showing separation of x-ray density (pink) and gravitational density (blue) [20]. . . . .  | 11 |
| Figure 3.1 | Overview of the CERN accelerator complex, including the LHC injector system and the main surrounding experiments. [22] . . . . .  | 15 |
| Figure 3.2 | A cut-out overview of the ATLAS detector and main components. Diagram from [23]. . . . .  | 17 |
| Figure 3.3 | A cut-out view of the ATLAS inner detector. Diagram from [23]. . . . .  | 18 |
| Figure 3.4 | A cut-out view of the ATLAS calorimeter system. Diagram from [23]. . . . .  | 20 |
| Figure 3.5 | A schematic demonstrating the components and accordion-like geometry of the ATLAS LAr EM calorimeter. Diagram from [24]. . . . .  | 21 |
| Figure 3.6 | A cut-out view of the ATLAS muon spectrometer system. Diagram from [23]. . . . .  | 22 |
| Figure 4.1 | Branching fractions for a SM Higgs boson as a function of the Higgs boson mass [27]. . . . .  | 25 |
| Figure 4.2 | Dark Higgs boson decay . . . . .  | 25 |
| Figure 4.3 | Signal point locations in $m_s$ vs. $m_{Z'}$ parameter space. Blue points represent the original signal grid, while green points are added in the second signal grid. . . . . | 27 |

|            |  |    |
|------------|--|----|
| Figure 4.4 | A visual summary of the TAR Jet algorithm. [47] . . . . .  | 34 |
| Figure 5.1 | Schematic representation of the differing reconstruction technique between merged and resolved regions. . . . .  | 39 |
| Figure 5.2 | Normalized $R = 1.0$ TAR jet mass distributions with and without lepton disentanglement applied for four representative signal points. <b>Upper row:</b> electron channel, <b>lower row:</b> muon channel. . . . .   | 41 |
| Figure 5.3 | TAR jet mass distributions for two representative signal points, compared with constituent anti- $k_t$ $R = 0.4$ vs. anti- $k_t$ $R = 0.2$ jets. Left: sample signal point at $(m_s, m_{z'}) = (310, 500)$ GeV. Right: sample signal point at $(m_s, m_{z'}) = (160, 1700)$ GeV. . | 42 |
| Figure 5.4 | “ $N - 1$ ” plots in the “merged” signal region showing the distribution of variables with all cuts applied except that of the variable shown. Grey bands represent MC statistical uncertainty on each bin. . . . .  | 46 |
| Figure 5.5 | $t\bar{t}$ yield in the “merged” $t\bar{t}$ control region for various $E_T^{miss}$ significance cut placements. . . . .   | 49 |
| Figure 5.6 | $E_T^{miss}$ trigger efficiency in the electron channel. . . . .   | 53 |
| Figure 5.7 | $E_T^{miss}$ trigger efficiency in the muon channel. . . . .   | 53 |
| Figure 5.8 | $E_T^{miss}$ trigger efficiency plotted against $E_T^{miss}$ with muons invisible. . . . .   | 53 |
| Figure 5.9 | Efficiency of the logical OR of $E_T^{miss}$ and single muon triggers. . . . .   | 54 |
| Figure 6.1 | Distributions of $m_s^{\min}$ in the “merged” and “resolved” signal regions. Grey bands represent MC statistical uncertainty on each bin. . . . .  | 59 |
| Figure 6.2 | Pre and post-fit yields in the merged signal and control regions for a background-only fit. . . . .  | 63 |
| Figure 6.3 | Pre and post-fit yields in the resolved signal and control regions for a background-only fit. . . . .  | 64 |
| Figure 6.4 | Expected exclusion limits for the dark Higgs signal model obtained using the “merged” signal and control regions. . . .  | 68 |
| Figure 6.5 | Expected exclusion limits for the dark Higgs signal model obtained using the “resolved” signal and control regions. . .  | 68 |

|             |  |    |
|-------------|--|----|
| Figure 6.6  | Expected exclusion limits for the dark Higgs signal model obtained using the “merged” and “resolved” signal and control regions. . . . .     | 69 |
| Figure 6.7  | Expected and observed exclusion limits for the dark Higgs signal model from the fully hadronic $s \rightarrow WW$ channel from [26]. . . . . | 69 |
| Figure A.1  | Preselection level distributions. Grey bands represent MC statistical uncertainty on each bin. . . . .                                       | 72 |
| Figure A.2  | Preselection level distributions. Grey bands represent MC statistical uncertainty on each bin. . . . .                                       | 73 |
| Figure A.3  | Preselection level distributions. Grey bands represent MC statistical uncertainty on each bin. . . . .                                       | 74 |
| Figure A.4  | Preselection level distributions. Grey bands represent MC statistical uncertainty on each bin. . . . .                                       | 75 |
| Figure A.5  | Data-MC Comparisons in the “merged” $t\bar{t}$ control region. Grey bands represent MC statistical uncertainty on each bin. . . .            | 76 |
| Figure A.6  | Data-MC Comparisons in the “merged” $t\bar{t}$ control region. Grey bands represent MC statistical uncertainty on each bin. . . .            | 77 |
| Figure A.7  | Data-MC Comparisons in the “resolved” $t\bar{t}$ control region. Grey bands represent MC statistical uncertainty on each bin. . . .          | 78 |
| Figure A.8  | Data-MC Comparisons in the “resolved” $t\bar{t}$ control region. Grey bands represent MC statistical uncertainty on each bin. . . .          | 79 |
| Figure A.9  | Data-MC comparisons in the “merged” $W$ +jets control region. Grey bands represent MC statistical uncertainty on each bin. . . .             | 80 |
| Figure A.10 | Data-MC comparisons in the “merged” $W$ +jets control region. Grey bands represent MC statistical uncertainty on each bin. . . .             | 81 |
| Figure A.11 | Data-MC comparisons in the “resolved” $W$ +jets control region. Grey bands represent MC statistical uncertainty on each bin. . . . .         | 82 |
| Figure A.12 | Data-MC comparisons in the “resolved” $W$ +jets control region. Grey bands represent MC statistical uncertainty on each bin. . . . .         | 83 |

|             |   |    |
|-------------|---|----|
| Figure A.13 | Comparisons of kinematic variables between the “merged” signal region and $t\bar{t}$ control region. <b>Left column:</b> signal region, <b>Right column:</b> control region. Grey bands represent MC statistical uncertainty on each bin. . . . .   | 84 |
| Figure A.14 | Comparisons of kinematic variables between the “merged” signal region and $t\bar{t}$ control region. <b>Left column:</b> signal region, <b>Right column:</b> control region. Grey bands represent MC statistical uncertainty on each bin. . . . .   | 85 |
| Figure A.15 | Comparisons of kinematic variables between the “resolved” signal region and $t\bar{t}$ control region. <b>Left column:</b> signal region, <b>Right column:</b> control region. Grey bands represent MC statistical uncertainty on each bin. . . . . | 85 |
| Figure A.16 | Comparisons of kinematic variables between the “resolved” signal region and $t\bar{t}$ control region. <b>Left column:</b> signal region, <b>Right column:</b> control region. Grey bands represent MC statistical uncertainty on each bin. . . . . | 86 |
| Figure A.17 | $N - 1$ plots in the “merged” signal region with the inclusion of a highly negatively weighted background event. Grey bands represent MC statistical uncertainty on each bin. . . . .   | 87 |
| Figure A.18 | $N - 1$ plots in the “merged” signal region with the inclusion of a highly negatively weighted background event. Grey bands represent MC statistical uncertainty on each bin. . . . .   | 88 |

## ACKNOWLEDGEMENTS

Unprecedented. That word has been used more than any other to describe the circumstances surrounding the last two years. The majority of my time at the University of Victoria has been overshadowed by the COVID-19 pandemic. Undertaking a major research project through such “unprecedented” times has required an unparalleled amount of support. The list of people who have helped me through this thesis is far too long to include in a single page, but I would like to give thanks to a select few who have had an especially large impact:

**Healthcare workers**, in Canada and worldwide. Your tireless work and boundless compassion have been an inspiration, and we wouldn’t be here without you.

**Dr. Justin Albert**, for seamlessly navigating the challenges of supervising me and my scatterbrain through zoom calls and emails.

**Danika MacDonell**, for welcoming me into the analysis team, and for your patience and kindness showing me the ins and outs of being an ATLAS analyst.

**The Mono-s analysis team**, Philipp, Dr. Hamano, and Dr. Kowalewski for your help and feedback throughout this project.

**My family**, Mom, Dad, Emily, and Claire, for always pushing me to succeed, but showing me the same love and encouragement no matter what. You are the world’s very best support network.

**My partner**, Jaedyn, for always being by my side. From you, that means the world.

*La lutte elle-même vers les sommets suffit à remplir un cœur d’homme. Il faut imaginer Sisyphe heureux.*

Albert Camus

# Declaration of Contributions

## Analysis Contributions

- Tested and implemented TAR jet lepton-disentanglement algorithm. (Track-Assisted-Reclustered [TAR] jets are defined in subsection 4.3.4, and lepton-disentanglement is described in subsection 5.3.2.)
- Developed and tested selection criteria for a signal region. (The signal region optimization is described in subsection 5.3.3.)
- Developed and  $t\bar{t}$  control region selection criteria. (The  $t\bar{t}$  control regions are described in Section 5.5.)
- Created data-MC comparisons for  $W$ +jets and  $t\bar{t}$  control regions (shown in Appendix A.2).
- Developed and implemented single-muon OR  $E_T^{miss}$  trigger strategy (described in Section 5.7).
- Worked on production of samples (ntuples) used by the analysis.
- Worked on implementation of HistFitter package for exclusion fits.
- Contributed to internal documentation for full Run-2 analysis paper.
- Co-developed and maintained plotting repository used for analysis plots.

# Chapter 1

## Introduction

Developed throughout the latter half of the 20th century, the **Standard Model of Particle Physics (SM)** [1–4] is the most successful theory describing the fundamental particles and their interactions. It describes matter as consisting of **quarks** (which make up protons and neutrons) and **leptons** (such as electrons), which interact via three fundamental forces (the electromagnetic force, the strong force, and the weak force) via the exchange of mediating **bosons**. The **Higgs boson**, the final SM particle to be discovered, generates the mass of the other particles via the “Higgs mechanism”. The predictions of the Standard Model have been extremely rigorously tested experimentally, and in almost all confirmed cases, have held true. There are, however, some areas in which the Standard Model falls short. For example, it does not include a description of the fourth fundamental force, gravity, nor does it explain the existence of dark matter, an invisible form of matter which is measured via astrophysical observations.

One extension to the SM proposed by theorists is the “dark Higgs” model [5]. This proposes the existence of three new particles:

- a dark matter fermion  $\chi$ ,
- a heavy mediator boson  $Z'$ , and
- a **dark Higgs boson**  $s$ .

In this model, the dark Higgs boson generates the masses of the  $Z'$  and  $\chi$  particles, and each of the three particles in this new “dark sector” can interact with their two counterparts. The proposed dark sector interacts with the SM via the coupling of the

$Z'$  to quarks and the mixing of the  $s$  with the SM Higgs boson, which allows the  $s$  to interact with SM particles.

In order to probe the limits of the SM, and to search for evidence of new phenomena such as the dark Higgs model, particle accelerators such as the **Large Hadron Collider (LHC)** collide protons at high energies. A **Toroidal LHC Apparatus (ATLAS)** experiment, one of the four main particle detectors at the LHC, detects the products of the proton-proton collisions, and measures their positions, energies, and momenta as they move through the detector. The collected data is then analyzed event-by-event to search for theorized signatures of new physics.

One technique to perform an analysis, which will be discussed in this thesis, is the “cut and count” method. We use this method to search for evidence of the dark Higgs signal model. We use Monte Carlo (MC) simulations to generate SM background and dark Higgs signal physics processes as they would occur from proton-proton collisions in the LHC, and also to simulate the behaviour of the resulting particles in the ATLAS detector. These simulations form the basis for us to predict what the ATLAS detector will measure with or without the existence of the signal model.

Such signal processes, however, would be rare, and searching for them in event displays among billions of proton-proton collisions would be impossible. Instead, using the cut and count method, we identify measurable properties of events that allow us to differentiate between events likely to be categorized as signal or background. We then cut our dataset down using these distinguishing properties to define a “signal region”, which we expect based on simulations to have a high concentration of signal events in the case where the dark Higgs model exists. We can then compare the expected distribution of events in this region from MC simulated data with measured ATLAS data and reach a conclusion about the likelihood that the signal model correctly describes nature.

This thesis focuses on the design of a signal region in the ATLAS analysis searching for the dark Higgs model, for the case where the  $s$  produced decays to a pair of  $W$  bosons, one of which decays to a pair of quarks and the other of which decays to a pair of leptons. It also discusses the design of a “control region”, which is used to constrain the SM expectation of one of the leading sources of background events.

Chapter 2 provides an overview of the Standard Model of particle physics, as well as the motivation for searching beyond its limits, and a description of the theoretical framework for the dark Higgs model. A description of the LHC and the ATLAS detector system are presented in Chapter 3. Chapter 4 describes analysis preparation

work, including MC simulation and analysis object definition. The signal and control region design and definition are described in Chapter 5, while Chapter 6 presents the statistical framework used to evaluate the analysis sensitivity and the results. Finally, Chapter 7 provides a summary and conclusions.

# Chapter 2

## Theory

The **Standard Model of Particle Physics (SM)** [1–4] is a quantized relativistic field theory that describes all known elementary particles, as well as their interactions via three of the four known fundamental forces. It has been tested rigorously since its development in the 1960s and 1970s, and in every confirmed result of laboratory experiments, its predictions have held true. A final piece of the puzzle fell into place with the experimental verification of the existence of the Higgs Boson in 2012 [6].

Despite its successes, however, the SM is considered an incomplete theory. It does not describe the interaction of matter via the fourth fundamental force, gravity, nor does it account for the existence of *dark matter* in our universe, nor the asymmetry between the observed amounts of matter and anti-matter. These shortcomings provide motivation to extend the SM by searching beyond it for new phenomena.

This chapter will provide a short overview of the SM theory, before describing some of the deficiencies that motivate an extension of it.

## 2.1 The Standard Model

### 2.1.1 The Particles

The elementary particles of the Standard Model are shown in Figure 2.1. They can be categorized into two groups: *fermions* and *bosons*. *Fermions* carry half-integer spin, and constitute the matter that surrounds us. For each fermion there also exists a corresponding anti-particle with an opposite electric charge. In this document anti-particles will be denoted either by the charge (e.g.  $e^+$  vs.  $e^-$ ) or by a bar overhead e.g. ( $t$  vs.  $\bar{t}$ ). The fermions can be further divided into two groups: *leptons* and

*quarks*. *Leptons*, the most familiar of which is the electron, interact via the weak force and, if electrically charged, the electromagnetic force. They come in three flavour generations, each of which has a neutral particle (*neutrino*) and a charged fermion with electric charge  $-1$ . Unlike leptons, which regularly exist freely, *quarks* exist mostly in bound states called *hadrons*, the most well-known of which are the proton and neutron, which are respectively composed of  $(u, u, d)$  and  $(u, d, d)$  quarks. Like leptons, quarks come in three generations, each of which has a pair of particles with electric charges  $+2/3$  and  $-1/3$ . They interact via all three forces of the standard model: strong, weak, and electromagnetic.

|                |                                 |                                  |                                  |                                      |                                  |
|----------------|---------------------------------|----------------------------------|----------------------------------|--------------------------------------|----------------------------------|
| mass           | $\approx 2.2 \text{ MeV}/c^2$   | $\approx 1.28 \text{ GeV}/c^2$   | $\approx 173.1 \text{ GeV}/c^2$  | 0                                    | $\approx 124.97 \text{ GeV}/c^2$ |
| charge         | $2/3$                           | $2/3$                            | $2/3$                            | 0                                    | 0                                |
| spin           | $1/2$                           | $1/2$                            | $1/2$                            | 1                                    | 0                                |
|                | <b>u</b><br>up                  | <b>c</b><br>charm                | <b>t</b><br>top                  | <b>g</b><br>gluon                    | <b>H</b><br>higgs                |
| <b>QUARKS</b>  | $\approx 4.7 \text{ MeV}/c^2$   | $\approx 96 \text{ MeV}/c^2$     | $\approx 4.18 \text{ GeV}/c^2$   | 0                                    |                                  |
|                | $-1/3$                          | $-1/3$                           | $-1/3$                           | 0                                    |                                  |
|                | $1/2$                           | $1/2$                            | $1/2$                            | 1                                    |                                  |
|                | <b>d</b><br>down                | <b>s</b><br>strange              | <b>b</b><br>bottom               | $\gamma$<br>photon                   |                                  |
|                | $\approx 0.511 \text{ MeV}/c^2$ | $\approx 105.66 \text{ MeV}/c^2$ | $\approx 1.7768 \text{ GeV}/c^2$ | $\approx 91.19 \text{ GeV}/c^2$      |                                  |
|                | $-1$                            | $-1$                             | $-1$                             | 0                                    |                                  |
|                | $1/2$                           | $1/2$                            | $1/2$                            | 1                                    |                                  |
|                | <b>e</b><br>electron            | $\mu$<br>muon                    | $\tau$<br>tau                    | <b>Z</b><br>Z boson                  |                                  |
| <b>LEPTONS</b> | $< 1.0 \text{ eV}/c^2$          | $< 0.17 \text{ MeV}/c^2$         | $< 18.2 \text{ MeV}/c^2$         | $\approx 80.39 \text{ GeV}/c^2$      |                                  |
|                | 0                               | 0                                | 0                                | $\pm 1$                              |                                  |
|                | $1/2$                           | $1/2$                            | $1/2$                            | 1                                    |                                  |
|                | $\nu_e$<br>electron neutrino    | $\nu_\mu$<br>muon neutrino       | $\nu_\tau$<br>tau neutrino       | <b>W</b><br>W boson                  |                                  |
|                |                                 |                                  |                                  | <b>GAUGE BOSONS</b><br>VECTOR BOSONS | <b>SCALAR BOSONS</b>             |

Figure 2.1: The elementary particles of the Standard Model [7].

*Bosons* carry integer spin, and mediate the forces via which particles interact. Massless *gluons* and *photons* as well as massive *W* and *Z* bosons are spin-1 vector bosons, while the Higgs boson is a spin-0 scalar boson. The massive *W* and *Z* vector bosons are the carriers of the weak force, the photon carries the electromagnetic force, and the gluons carry the strong force binding quarks. Along with interacting with fermions via exchange, bosons are also able to interact among themselves. *W* bosons are able to directly interact with both *Z* bosons and photons, and can additionally self-interact. Gluons can also self-interact, but *Z* bosons and photons cannot. The Higgs boson interacts with all massive particles, including with itself,

and it is via interaction with the Higgs field that massive bosons obtain their mass.

### 2.1.2 Quantum Electrodynamics (QED)

*Quantum Electrodynamics* (QED) [1, 2, 8] is a quantum field theory of electrodynamics and the electromagnetic force. It describes the interaction of electrically charged particles via the exchange of photons. Mathematically, QED is an abelian (commutative) gauge theory with the gauge group  $U(1)$ . The fundamental interactions of the theory are: the emission or absorption of a single photon by a charged particle and the creation or annihilation of a pair of charged particles. These interactions can each be represented by various orientations of the Feynman vertex shown in Figure 2.2.

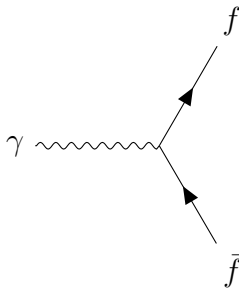


Figure 2.2: The fundamental QED vertex.

### 2.1.3 Weak Interactions

The weak force [3, 9–11], in contrast to the electromagnetic force, is mediated by the exchange of massive vector bosons  $W^+$ ,  $W^-$ , and  $Z$ . Owing to the charge of the  $W$  bosons, and to the weak interactions of the neutral fermions in addition to those of charged fermions, there are many more possible fundamental interactions. Figure 2.3 shows the vertices corresponding to three such interactions: (a) fermions interacting with the  $Z$  boson via a similar vertex to that of Figure 2.2, (b) a charged lepton and neutrino interacting with a  $W$  boson, and (c) a quark-antiquark pair interacting with a  $W$  boson.

At high energy scales, the aforementioned electromagnetic force and the weak force unify to become the electroweak interaction. Glashow, Salam, and Weinberg’s Electroweak theory unites the two forces into a single  $U(1) \otimes SU(2)$  gauge theory.

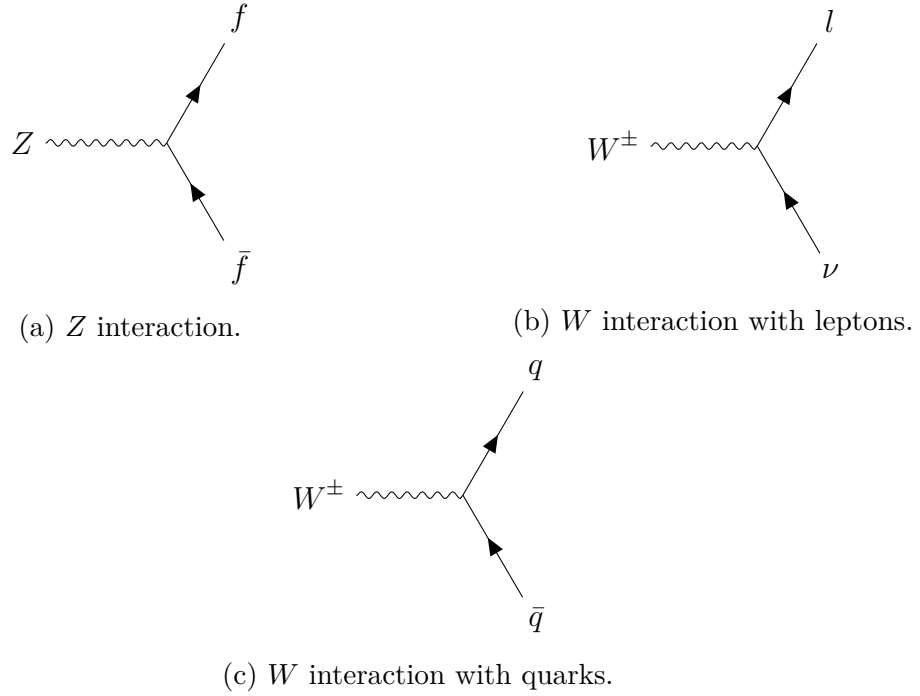


Figure 2.3: Some fundamental weak interaction vertices.

### 2.1.4 Quantum Chromodynamics (QCD)

*Quantum Chromodynamics* (QCD) [3, 12] is the theory describing the strong interaction between gluons and quarks. It is again a gauge theory, this time with the  $SU(3)$  symmetry group. There are three colour charges associated with this group, each one of which can be carried by both quarks and gluons. Each of the 8 gluons carries a unique colour charge and anti colour charge pair, while quarks carry a single colour charge. Because they carry colour charge, gluons are able to both interact with quarks and self-interact. This gives rise to several possible interaction vertices in the theory, some of which are shown in Figure 2.4.



(a) Gluon interaction with quarks.

(b) Gluon self-interaction.

Figure 2.4: Some fundamental QCD interaction vertices.

### 2.1.5 The Higgs Boson

Electroweak theory alone does not contain a mechanism to provide particles with mass. This contrasts with the observed reality that all fermions as well as  $W$  and  $Z$  bosons are in fact massive. If the Lagrangian of these theories were to contain mass terms, the theories would lose their gauge invariance and the SM would not be renormalizable. Instead, these particles acquire their mass through the *Higgs Mechanism* [13–15].

A Higgs field  $\Phi$  is introduced, with a Lagrangian which can be written as:

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger D^\mu \Phi + V(\Phi) \quad (2.1)$$

where  $D_\mu$  is the gauge covariant derivative of the electroweak theory. In order for the Lagrangian to remain gauge invariant, the Higgs potential  $V(\Phi)$  must take the form:

$$V(\Phi) = \mu^2 \Phi + \lambda \Phi^4 \quad (2.2)$$

where  $\lambda$  and  $\mu^2$  are free parameters. The value of the parameter  $\lambda$  is forced to be greater than 0 by requiring that the potential have a stable minimum. This leads the potential to have two different possible shapes, shown in Figure 2.5, given the sign of  $\mu^2$ :

1.  $\mu^2 > 0$ : The trivial case of a parabolic potential with a minimum at  $\Phi = 0$  arises.

2.  $\mu^2 < 0$ : A potential with minima at:

$$|\Phi| = v = \sqrt{\frac{\mu^2}{\lambda}} \quad (2.3)$$

arises, where  $v$  is known as the vacuum expectation value. The minima are occupied by infinitely degenerate states. This is the case that gives rise to the Higgs mechanism.

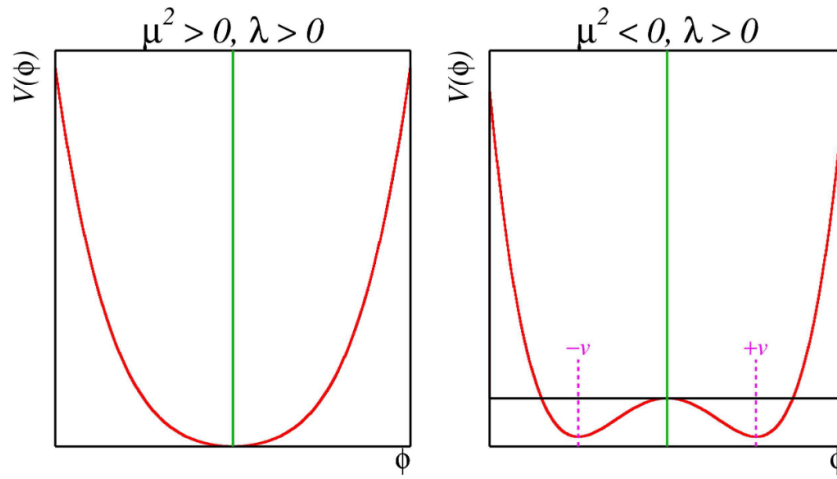


Figure 2.5: Shapes of the Higgs potential for different cases of  $\mu^2$ . [16]

In the second case, the ground states occupying the minimum are not equivalent under gauge transformation, which breaks the electroweak gauge symmetry. The masses of particles are then determined by the strength of their coupling to the Higgs field.

## 2.2 Beyond the Standard Model

The Standard Model as described above has proven extremely robust through precise experimental testing. There are, however, many remaining questions in physics that cannot be answered within its confines. Some of those not further explored in this thesis include: whether or not there is a quantum theory of gravity that can tie it to the SM, the exact value of the neutrino masses and whether they are Majorana particles (their own antiparticles), and what the cause is of the matter-antimatter

asymmetry in the universe. The fact that these questions are unanswered in the standard model tells us we must search deeper, and test beyond its present limits.

### 2.2.1 Dark Matter

Another currently open question is the nature of dark matter. Astrophysical observations including the dynamics of galaxy clusters [17] and rotational curves of galaxies [18] are not explained under Einstein's theory of gravitation by visible matter alone. Several theories have arisen over time to explain this discrepancy, including that Einstein's theory is not correct at galactic scales, or that these galaxies contain ordinary matter that is somehow unobservable to us.

Most evidence, however, points to the existence of a new type of matter that interacts gravitationally but not electromagnetically with ordinary matter. Perhaps the most clear evidence for this theory comes from observations of the Bullet Cluster [19], a pair of colliding clusters of galaxies. In such a scenario, with some normal matter and some dark matter in each cluster, the matter would collide and interact, slowing it down, while the dark matter would pass through largely undisturbed. In the case of the Bullet Cluster, this separation is observed when comparing the distribution of mass in the galaxies measured by gravitational lensing and the distribution of normal matter in the form of galaxy plasma, measured as emitted x-ray radiation. Figure 2.6 shows the observed separation, with gravitational density in blue and x-ray density shown in pink.



Figure 2.6: The Bullet Cluster showing separation of x-ray density (pink) and gravitational density (blue) [20].

### 2.2.2 Dark Higgs Boson Model

This work involves a search for dark matter produced in association with a new hypothetical scalar “dark Higgs” boson  $s$  [5].<sup>1</sup> In this model the dark matter particle  $\chi$  is a Majorana fermion, with mass generated by a Higgs mechanism in the dark sector. This hypothesized dark sector Higgs mechanism also generates the new  $s$  particle. If the mass of  $s$  is less than that of  $\chi$ , then the relic density is set by the rate of the process  $\chi\chi \rightarrow ss$ , followed by the decay of  $s$  to standard model particles. The addition of a further particle, such as a massive  $Z'$  boson, allows this model to be probed at a collider.

The model proposes that the dark matter particle  $\chi$  obtains its mass from the vacuum expectation value (vev)  $w$  of a new Higgs field  $S$ . A new  $U(1)'$  gauge group is proposed, under which  $S$  carries a charge  $q_s$ . As a result, the vev  $w$  of  $S$  breaks the gauge symmetry, and through this mechanism the mass of the corresponding  $Z'$  boson is generated. Additionally, the dark matter particle  $\chi$  couples to the  $Z'$  boson, allowing all particles in the new dark sector to interact. This gives rise to the renormalizable interaction Lagrangian, which can be written in terms of four

---

<sup>1</sup>In this work,  $s$  will refer to the dark Higgs boson and not the strange quark, unless otherwise specified.

independent parameters  $m_\chi$ ,  $m_{Z'}$ ,  $m_s$ , and  $g_\chi$ :

$$\mathcal{L}_\chi = -\frac{1}{2}g_\chi Z'^\mu \bar{\chi} \gamma^5 \gamma_\mu \chi - g_\chi \frac{m_\chi}{m_{Z'}} s \bar{\chi} \chi + 2g_\chi Z'^\mu Z'_\mu (g_\chi s^2 + m_{Z'} s), \quad (2.4)$$

where  $m_\chi$  is the mass of the dark matter particle,  $m_{Z'}$  is the mass of the  $Z'$  boson,  $m_s$  is the mass of the dark Higgs, and  $g_\chi$  is the dark matter coupling constant, which controls the rate of  $Z' \rightarrow \chi\chi$  decays and  $Z'^* \rightarrow Z's$  processes in the signal model. Additionally, the coupling of the  $Z'$  boson to standard model quarks is described by the Lagrangian:

$$\mathcal{L} = -g_q Z'^\mu \bar{q} \gamma_\mu q, \quad (2.5)$$

where  $g_q$  is the coupling constant between  $Z'$  and quarks, which controls the rate of  $Z'$  production in the ATLAS detector, and therefore the rate of signal events.

A final free parameter,  $\theta$ , is the non-zero mixing angle between the SM Higgs boson and the dark Higgs boson. The dark Higgs obtains its couplings to standard model particles through this mixing, and therefore shares the same standard model decay branching fractions as the SM Higgs boson.

For the analysis described in this work the free parameters are set as:

- $g_\chi = 1$ ,
- $m_\chi = 200$  GeV,
- $m_{Z'}$  is allowed to vary,
- $m_s$  is allowed to vary,
- $g_q = 0.25$ , and
- $\theta = 0.01$ .

The values of  $g_\chi$  and  $g_q$  were chosen to facilitate comparison with other LHC searches with similar models, which traditionally use the values selected. The value of  $\theta$  was chosen to match that of [5]. Its precise value is not relevant to this search, but it is sufficiently large that the dark Higgs decays promptly to standard model states and does not create a displaced vertex. The varied values of  $m_s$  and  $m_{Z'}$  form the parameter space covered by this search. When analyses of all  $s \rightarrow WW$  final states are complete, they will be statistically combined. It is anticipated that one or

more fixed parameters will be varied at the time of combination to probe the signal model more completely.

## Chapter 3

# The Large Hadron Collider and the ATLAS Experiment

The **Large Hadron Collider (LHC)** is the world's largest and most energetic proton-proton collider. It is located within a 27 km circular ring beneath Switzerland and France, below towns, farmland, and part of the Jura mountains next to the **Conseil Européen pour Recherche Nucléaire (CERN)** in suburban Geneva. It began operation in 2008, and since then has collided over  $10^{15}$  particles. The collider houses four major experiments: ALICE, ATLAS, CMS, and LHCb, along with several smaller projects. ALICE (A Large Ion Collider Experiment) studies heavy ion collisions, while LHCb (Large Hadron Collider beauty) specializes in studying the physics of the bottom quark. **ATLAS (A Toroidal LHC ApparatuS)** and CMS (Compact Muon Solenoid) are both general purpose experiments designed to study a wide range of interactions resulting from high-energy proton-proton collisions. This work uses data collected by the ATLAS experiment, and a description of the LHC accelerator and ATLAS detector follow in Sections 3.1 and 3.2.

### 3.1 The Large Hadron Collider

Built between 1998 and 2008, the LHC [21] began colliding protons in 2010 at a center-of-mass energy of 7 TeV, and most recently from 2015 to 2018 collisions occurred at 13 TeV. Successfully colliding particles at this energy is an immense technical challenge which is achieved by the many technologies of the LHC and its accelerator complex.

In order to reach a beam energy of 6.5 TeV, protons are slowly stepped up in energy

through a chain of accelerators before reaching the LHC. They begin their journey as hydrogen molecules, which are ionized into bare protons before being accelerated to an energy of 50 MeV by the Linac2 linear accelerator. Following this, the Proton Synchrotron Booster (PSB) accelerates them to 1.4 GeV, passing them to the Proton Synchrotron (PS) and the Super Proton Synchrotron to be accelerated to 25 and then 450 GeV. Finally, beams are split into clockwise and counterclockwise directions and injected into the LHC where they are accelerated to their final 6.5 TeV energy. The CERN accelerator complex, including this injection system as well as surrounding experiments is shown in Figure 3.1.

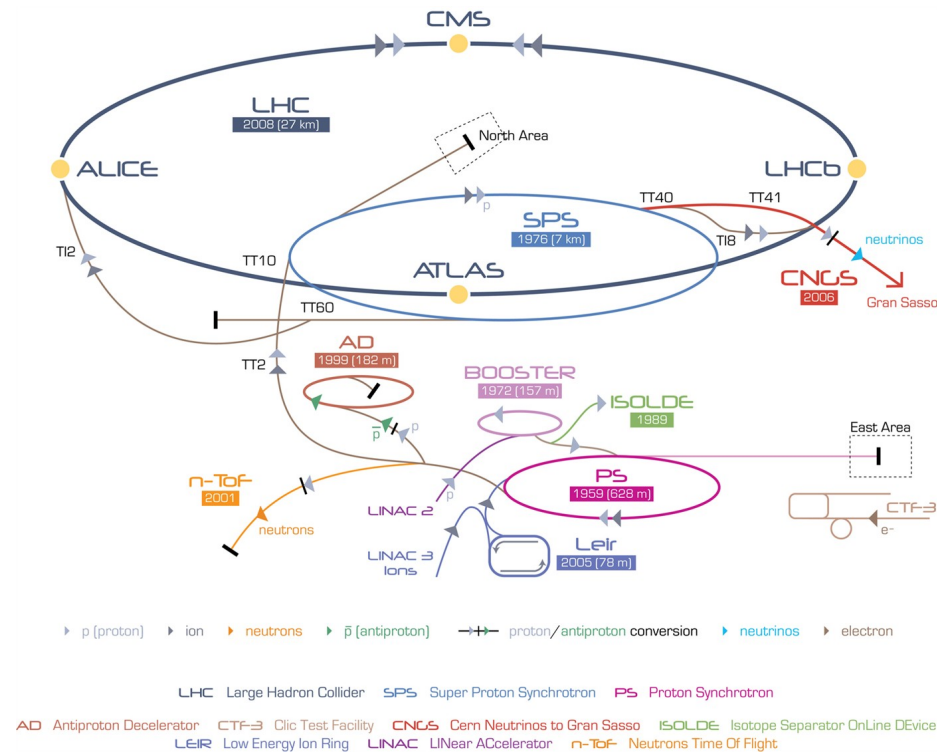


Figure 3.1: Overview of the CERN accelerator complex, including the LHC injector system and the main surrounding experiments. [22]

The acceleration of the protons is achieved by radio-frequency cavities. These cavities contain a resonant electromagnetic field oscillating at 400 MHz, which is applied to particles passing through. The LHC contains 8 cavities per beam, with each providing a maximum of 2 MV of potential, so each proton can receive up to 16 MeV of energy per lap. As a result it takes millions of laps over a period of around 20 minutes for a proton injected at 450 GeV to reach its collision energy of 6.5 TeV. These RF cavities also help to keep each beam in bunches of  $1.15 \times 10^{11}$  protons

spaced at intervals of just 25 ns.

The crown jewels of LHC technology are its magnets. The LHC contains 1,232 superconducting NbTi dipole magnets kept at 1.9 K, each spanning 14.3 m and weighing 35 tonnes, with each creating an 8.3 Tesla magnetic field. This field lies perpendicular to the beam path, bending the beams along their desired route. The bending dipole magnets are complemented by 392 quadrupole magnets that focus the beams to a small aperture, and many higher-order multipole magnets which provide small beam corrections.

Together with the collisional energy of the accelerated particles, the other most important measure of a particle accelerator is the luminosity it achieves. **Luminosity** ( $\mathcal{L}$ ) is used to determine the rate ( $R$ ) at which a given interaction occurs using:

$$R = \mathcal{L}\sigma, \quad (3.1)$$

where  $\sigma$  is the cross-section of the desired interaction. As a result, when searching for rare processes, obtaining a high luminosity is crucial.

The integrated luminosity,  $L$ , gives the total number of interactions over a period of time, and is defined as:

$$L = \int \mathcal{L} dt. \quad (3.2)$$

At the LHC, the luminosity is controlled by the number  $n_b$  of bunches circulating, the frequency of revolution  $f_r$ , the numbers  $N_{1,2}$  of particles in each colliding bunch, and the cross-sectional area of the colliding beams. This results in the equation:

$$\mathcal{L} = \frac{f_r n_b N_1 N_2}{4\pi\sigma_x\sigma_y} R_\phi, \quad (3.3)$$

where  $R_\phi$  is a geometrical loss factor caused by the beams crossing at an angle, and  $\sigma_{x,y}$  are the horizontal and vertical widths of the beams (which, to a good approximation, have Gaussian cross-sectional profiles). The LHC currently has a nominal peak luminosity of  $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ .

## 3.2 The ATLAS Experiment

The ATLAS experiment [23] uses the LHC in combination with the ATLAS detector to test a broad range of SM and BSM predictions. Together with the CMS experiment,

its greatest achievement to date was the detection and discovery of the SM Higgs boson in 2012 [6].

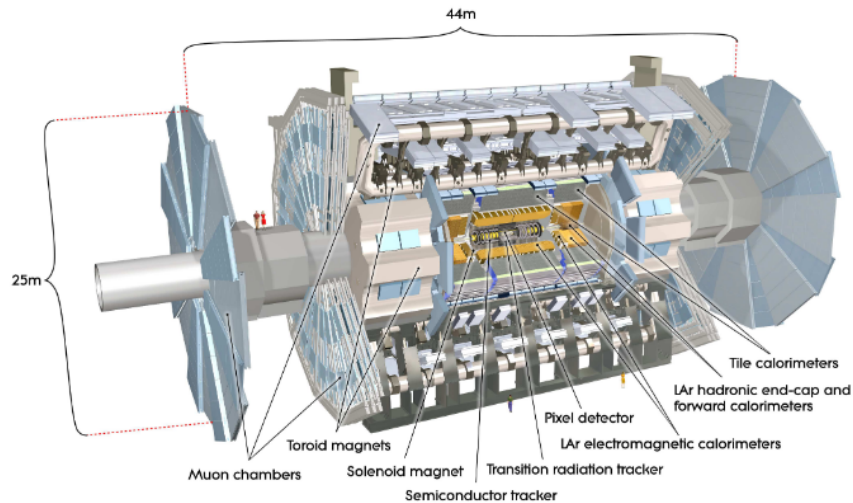


Figure 3.2: A cut-out overview of the ATLAS detector and main components. Diagram from [23].

The ATLAS detector, an overview of which is shown in Figure 3.2, is comprised of several layers of sub-detectors, which work in tandem to detect many diverse particles. From inner-most to outer-most the layers are:

- the **inner detector**, which in turn contains the pixel detector, the semiconductor tracker, and the transition radiation tracker, and which has excellent angular and positional resolution to track charged particles,
- the **electromagnetic calorimeters**, designed primarily to measure the energy and position of electrons and photons,
- the **hadronic calorimeters**, designed to measure the energy and position of hadrons,
- the **muon spectrometer**, to track and measure muons.

When combined with the electronics required to read out and trigger on measurements, the entire detector weighs over 7000 t and measures 44 m long by 25 m wide and high. Its barrelled shape with end caps covers nearly the full solid angle. The following sections will describe the layout and operations of the ATLAS detector in more detail.

### 3.2.1 Inner Tracking Detector

The ATLAS inner detector tracks the direction and momentum of charged particles immediately after the particles leave the interaction point. A 2 T superconducting solenoid surrounds the inner detector, bending charged particles via the Lorentz force, allowing their momentum to be determined from the curvature of their path. Within the inner detector, there are three separate sub-detectors: the **Pixel Detector**, the **Semiconductor Tracker (SCT)**, and the **Transition Radiation Tracker (TRT)**. A cutout view of the detector is shown in Figure 3.3.

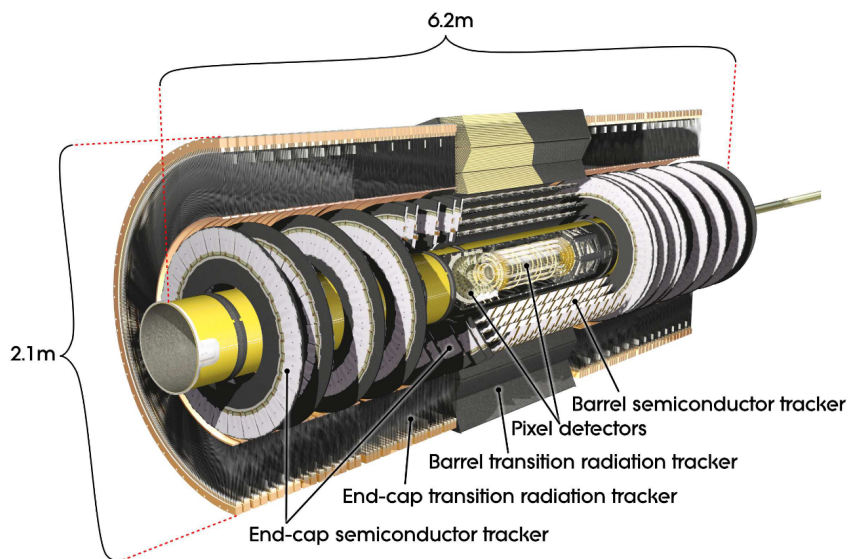


Figure 3.3: A cut-out view of the ATLAS inner detector. Diagram from [23].

The Pixel Detector consists of an array of 1744 pixel sensors, each containing 47232 silicon pixels, with most of these pixels each measuring  $50 \times 400 \mu\text{m}^2$ . They are arranged into three cylindrical layers in the barrel and three disk-shaped layers on each end. The high granularity of these detectors ensures strong angular resolution, providing precise measurements of vertex location and track momentum. Notably, this detector must be very radiation-hard in order to maintain performance over the detector lifespan, as the pixel detector is the ATLAS subdetector that receives the highest radiation dose from the intense particle flux.

Though it would be desirable, it would not be feasible to extend the high-resolution pixel detector because of the high cost and signal readout volume. As a result, in the SCT the small rectangular pixels are extended into silicon strip detectors, which

are laid in pairs at an angle of 40 mrad to each other to allow measurements in two dimensions. The SCT is made up of four concentric cylindrical layers in the barrel and two disk-shaped layers on each end-cap.

The TRT is the final component of the component of the inner detector. It consists of approximately 300000 drift tubes, primarily filled with Xenon gas, with the wall of each tube kept at -1.5 kV relative to a central wire at ground. When particles pass through the tubes, the gas is ionized and the resulting electrons drift to the center and are collected on the central wire. The electrical signal pulse is then amplified within electronics at the end of each wire. Additionally, the gaps between TRT straws are filled with polymer fibres (in the barrel) or foils (in the end caps). As a result, highly relativistic particles create transition radiation at the material interfaces, which aids with electron identification.

### **3.2.2 Calorimeters**

The ATLAS calorimeter system measures the energies and positions of both charged and neutral particles after they leave the inner detector. In addition to measuring the energy of each particle, the calorimeters must contain enough material to stop the vast majority of particles (with the exception of muons and undetected neutrinos) from reaching the muon spectrometers. To achieve this, high-density passive layers interact with passing particles and cause them to decay into showers, dispersing their energy. These layers are interleaved with active layers that detect the decaying showers to measure particle energy.

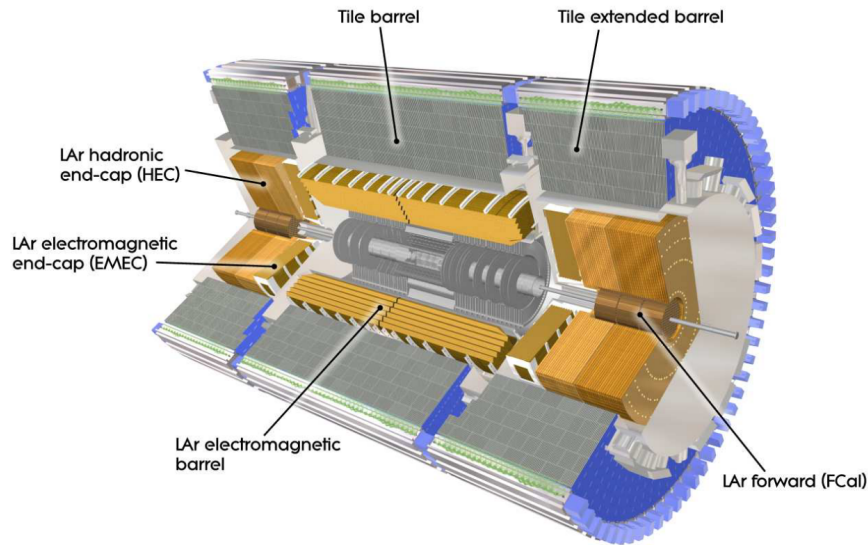


Figure 3.4: A cut-out view of the ATLAS calorimeter system. Diagram from [23].

The ATLAS calorimeters can be categorized into electromagnetic (EM) calorimeters which measure the energy of electrons and photons, and hadronic calorimeters which primarily measure the energy of hadrons. ATLAS uses a combination of two different calorimeter technologies to form these detectors. In the end-cap and forward regions, Liquid Argon (LAr) is used as the active layer in both the hadronic and EM calorimeters; while in the barrel region, plastic scintillating tiles are employed in the hadronic calorimeter. Figure 3.4 depicts the layout of the ATLAS calorimeter system.

### Electromagnetic calorimeters

The electromagnetic LAr calorimeter lies just outside the solenoid surrounding the inner detector. Layers of lead are used as the passive material. In the lead, electrons and photons interact with the closely-spaced atoms, initiating a cascading shower of decays. The charged electrons in the shower pass through the active layers and ionize the LAr inside. The electrons and ions then drift across the 2 kV difference to opposite electrodes, where they are amplified into an electrical signal.

Like the inner detector, the LAr EM calorimeter can be divided into an **Electromagnetic Barrel (EMB)** calorimeter and two **Electromagnetic End-Cap (EMEC)** calorimeters. In the barrel region, the layers are placed together in a folded accordion-like geometry to create uniformity and allow easy electronic read-out. An overview of this accordion-style layout is shown in Figure 3.5. The barrel

and end-cap regions are 53 cm and 63 cm thick respectively, and cover a minimum of 22 and 24 radiation lengths ( $X_0$ ), keeping shower leakage to a minimum.

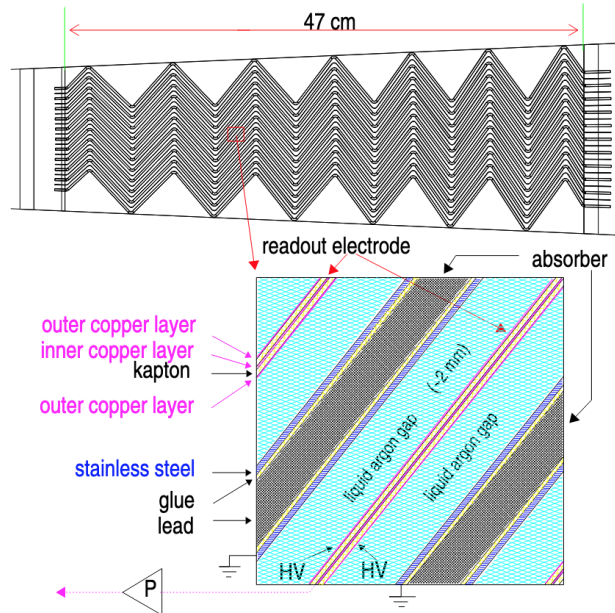


Figure 3.5: A schematic demonstrating the components and accordion-like geometry of the ATLAS LAr EM calorimeter. Diagram from [24].

## Hadronic Calorimeters

In the end cap regions, the **Hadronic End-Cap (HEC)** calorimeter shares a similar design to the EM LAr calorimeters, but with copper as the passive absorbing material between active layers. In the barrel region, the hadronic calorimetry is performed by the **Tile Calorimeter** with plastic scintillators as the active detectors. Particle showers passing through the plastic tiles produce scintillating light, which is directed to photomultipliers to be converted to an electrical signal. In the tile calorimeter, steel is the passive material laid between active layers. The depth of the combination of electromagnetic and hadronic calorimeters is important to prevent particles from punching through to the background-sensitive muon spectrometer. The total thickness, including supporting materials, is approximately 11 interaction lengths, which keeps punch-through below the levels of irreducible backgrounds in the muon spectrometer, and also ensures an accurate measurement of  $E_T^{miss}$  as very little energy escapes measurement.

The final calorimeter in the ATLAS detector is the **Forward Calorimeter (FCal)**,

located nearest the beam path. It has three layers of passive absorbing metals: an inner copper layer optimized for electromagnetic measurements and two outer tungsten layers to measure hadronic interactions. It too employs liquid argon as the active material, this time in a matrix of longitudinal channels parallel to the beam line.

### 3.2.3 Muon Spectrometers

After the calorimeters have stopped the vast majority of particles exiting the interaction point, muons and neutrinos remain undetected. The ATLAS experiment does not directly detect neutrinos, and instead reconstructs them from missing transverse momentum, however muons are measured by the muon spectrometer. A diagram of the muon system is shown in Figure 3.6. The centerpiece of this system is an arrangement of three air-core toroidal magnets, a barrel toroid and two end-cap toroids. Each toroid consists of eight coils in loops in a radially-symmetric pattern around the beam line. Together they form a 4 T magnetic field to bend muons.

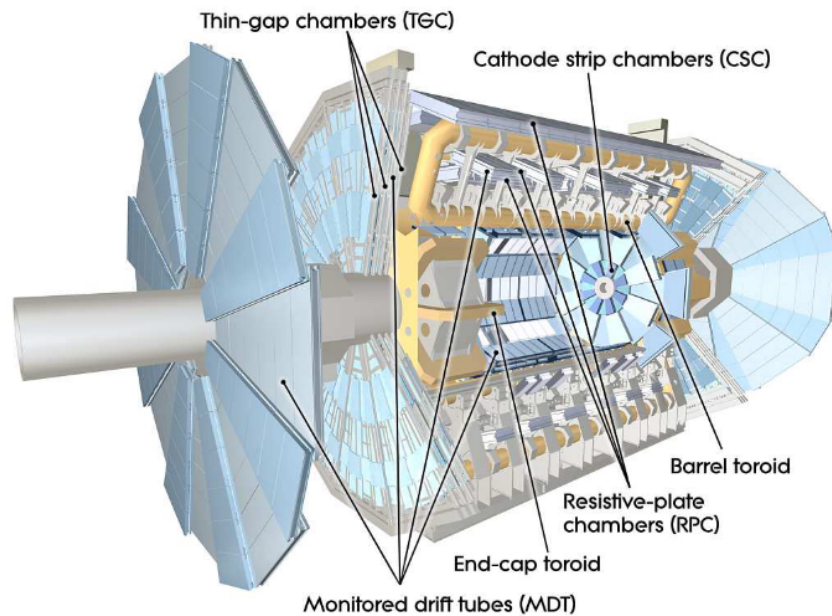


Figure 3.6: A cut-out view of the ATLAS muon spectrometer system. Diagram from [23].

Four different detector types are used to measure the muons. **Monitored Drift Tubes (MDTs)** are pressurized gas drift tubes that collect ionized electrons to measure track coordinates in the bending direction of the magnetic field. They are sup-

plemented by finer-grained and more radiation-hard **Cathode Strip Chambers (CSCs)** at  $|\eta| > 2$ . **Resistive Plate Chambers (RPCs)** and **Thin Gap Chambers (TGCs)** measure the muon track coordinates in the direction orthogonal to the magnetic bending direction, and are also used by the trigger system to provide bunch-crossing identification and  $p_T$  thresholds.

### 3.2.4 Trigger and Data Acquisition

The ATLAS detector can detect as many as 1.7 billion proton-proton collisions per second, but to save and process all of the information from each collision would require a prohibitive amount of readout electronics, computing power, and storage space. Instead, the trigger system quickly selects events based on key observables to slim down the data. On the detector itself, the **Level-1 (L1) trigger** uses custom electronics within the detector, and works with reduced granularity information from the muon chambers and calorimeters to select potentially interesting events with high- $p_T$  objects or high  $E_T^{miss}$ . The detector readout systems can handle a maximum acceptance rate of 100 kHz from the L1 trigger. The **Level 2 (L2) trigger** is outside the detector and uses traditional computing resources. It takes regions of interest where the L1 trigger finds interesting objects and reconstructs them more fully to further reduce the event rate below 3.5 kHz. Together with the L2 trigger, the **event filter** forms the second half of the **High-Level Trigger (HLT)**. At this stage, events are fully reconstructed before being further filtered to a rate of approximately 200 Hz to be stored and analyzed offline.

# Chapter 4

## Analysis Preparation

### 4.1 Signal Model

This work concerns a search for dark matter produced in association with a dark Higgs boson as described in subsection 2.2.2. The dark Higgs boson,  $s$ , then decays to standard model particles with the same branching ratios as a standard model Higgs boson of variable mass, as shown in Figure 4.1. At low  $s$  mass, this is dominated by a decay to a pair of  $b$  quarks, and is currently being studied via a reinterpretation of the ATLAS search for dark matter produced in association with a standard model Higgs boson decaying to  $b\bar{b}$  in [25]. At  $s$  masses above 160 GeV, however, the decay to a pair of  $W$  bosons, shown in Figure 4.2, becomes kinematically available on-shell. Even slightly below this threshold,  $s$  decays to  $WW$  are the dominant process. The pair of  $W$  bosons decay rapidly, and are not detected as final state particles by the ATLAS detector. Instead, they each decay either to hadronic ( $q\bar{q}$ ) or leptonic ( $l\nu$ ) final states. A search in the fully hadronic channel, with both  $W$  decaying to  $q\bar{q}$ , is complete (see [26]), and this work focuses on the semileptonic decay channel with one  $W \rightarrow q\bar{q}$  and one  $W \rightarrow l\nu$ . When both analyses are complete, the results will be statistically combined.

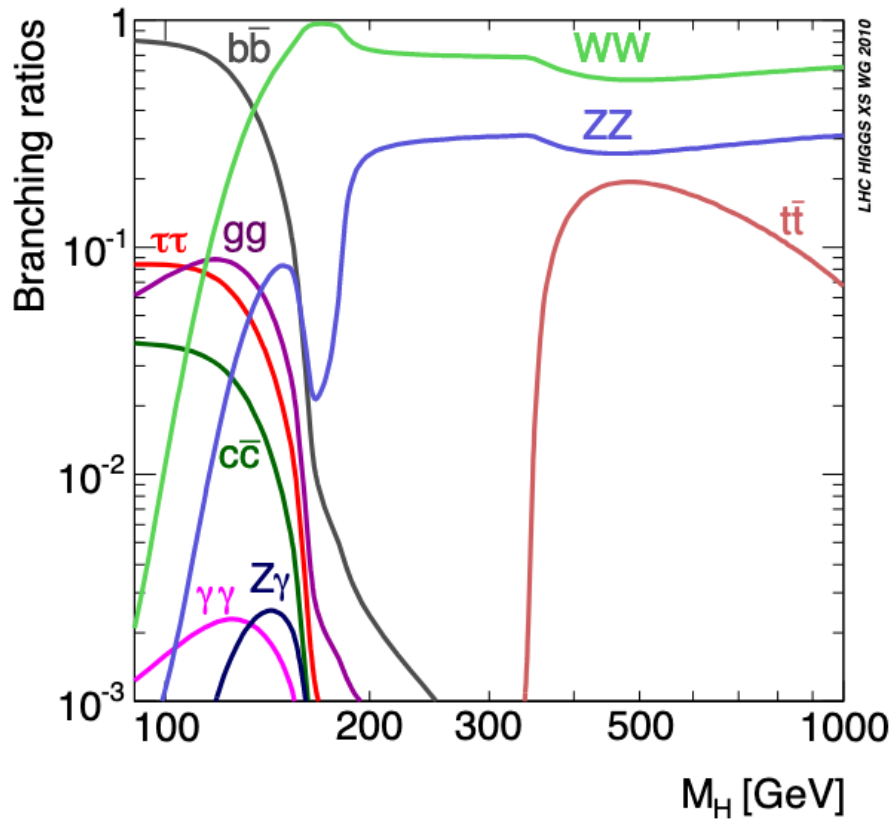


Figure 4.1: Branching fractions for a SM Higgs boson as a function of the Higgs boson mass [27].

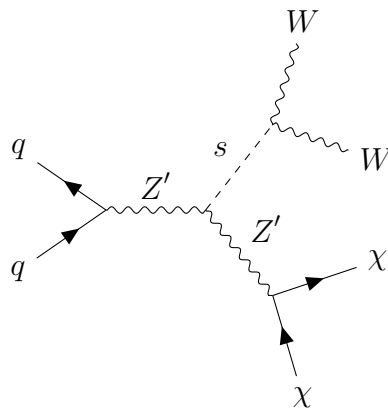


Figure 4.2: Dark Higgs boson decay

In the ATLAS detector, this signal signature will be characterized by a single lepton in the final state from the  $W \rightarrow l\nu$  decay, along with a pair of jets from the

$W \rightarrow q\bar{q}$  decay and a strong missing transverse momentum signature from the dark matter particles and neutrino.

## 4.2 Monte Carlo Production and Data

In order to search for a signal such as the one described above, it is necessary to form a prediction which can be compared with data from the ATLAS detector. In order to achieve this, simulated data is produced using Monte Carlo (MC) generators according to standard model predictions. This represents “background” data, or what we would expect to see if the signal process did not exist. Added to this, signal MC samples are produced according to the signal model. Then, after isolating for an area of phase space that would be rich in signal events, the ATLAS data can be compared with the simulated data with or without signal events to make statistical conclusions about the likelihood of the existence. The following sections will detail the MC and data samples produced for this analysis.

### 4.2.1 Signal MC

Signal samples are produced using the model and parameter choices outlined in subsection 2.2.2. They are produced by calculating the hard process cross-section at leading order (LO) using MADGRAPH5\_aMC@NLO 2.7.2 [28] interfaced with PYTHIA8 8.230 [29] for parton shower modelling. The CKKW-L [30] procedure with a matching scale of  $\min(\frac{m_s}{4}, 40 \text{ GeV})$  is used to prevent overlap between the cross-section matrix element and parton shower. Two separate signal grids were produced, with the second covering an expanded parameter space. In the first, MADGRAPH5\_aMC@NLO is used with the NNPDF30 LO PDF set with  $\alpha_s = 0.13$  [31], while in the second MADGRAPH5\_aMC@NLO is used with the NNPDF30 NLO PDF set with  $\alpha_s = 0.118$  [31]. The signals studied span a parameter space in  $m_s$  and  $m_{Z'}$  with samples produced at the mass points shown in Figure 4.3.

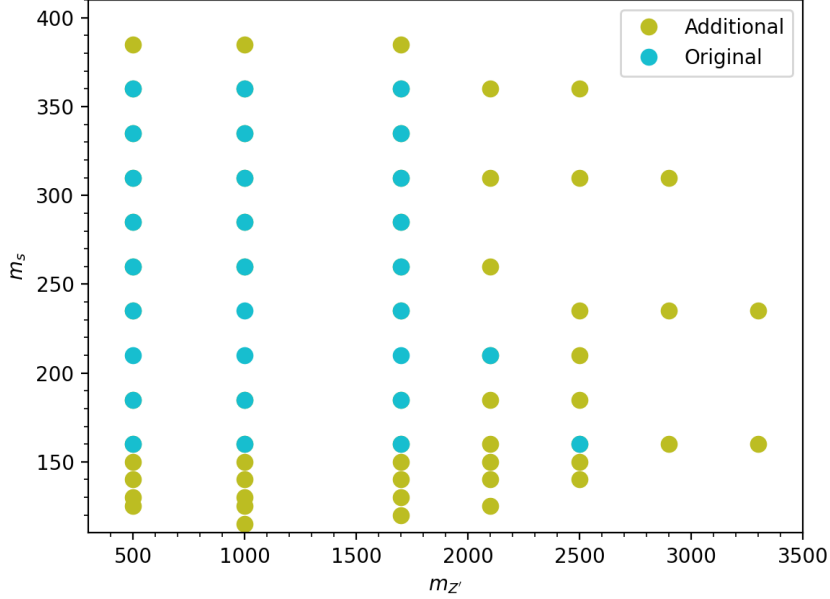


Figure 4.3: Signal point locations in  $m_s$  vs.  $m_{Z'}$  parameter space. Blue points represent the original signal grid, while green points are added in the second signal grid.

## 4.2.2 Background MC

The dominant and sub-dominant SM backgrounds for this analysis are:

- $W$ +jets events, containing a single  $W$  boson and 1 or more jets,
- $t\bar{t}$ , containing with a top and anti-top pair, and
- di-boson events, containing a pair of vector bosons.

Additionally,  $Z$ +jets, tri-boson, and single-top quark<sup>1</sup> events make non-negligible background contributions. All background contributions are estimated using MC simulated data, and the  $W$ +jets and  $t\bar{t}$  backgrounds are further constrained using data-driven control regions.

### $V$ +jets, Di-boson, and Tri-boson

Two distinct sets of  $W$ +jets and  $Z$ +jets ( $V$ +jets) background samples are used. The samples are simulated with the SHERPA v2.2 MC generator [32] with up to

<sup>1</sup>In some plots in this work single-top quark events are denoted by the short form “stop”. This is not to be confused with a supersymmetric partner of a top-quark often referred to as a “stop”.

two additional parton emissions at NLO or four additional parton emissions at LO accuracy [33][34]. Merging of parton showers with matrix elements is achieved using CKKW [30] matching with the MEPS@NLO prescription extending this to NLO [35]. The NNPDF30 NNLO PDF set is used [31].

The difference between these two sets of samples is that first set is produced using SHERPA v2.2.1, and the second with SHERPA v2.2.10. As well, additional statistics-enhanced  $W$ +jets samples are added to the second set with  $m_W > 120$  GeV [34]. This greatly increases the number of simulated events in this small section of phase space that is important to this analysis. Overlap removal in  $m_W$  is performed between these mass-enhanced samples and the other  $W$ +jets samples.

Di-boson and tri-boson samples are similarly generated with only a single set using either SHERPA v2.2.1 or v2.2.2 depending on the process, and CKKW matching is again used to merge parton shower and matrix element simulation.

### $t\bar{t}$ and single-top

MC samples for the  $t\bar{t}$  and single-top standard model background process are modelled using the POWHEGBOX v2 [36] generator, which gives matrix elements at NLO in the strong coupling constant  $\alpha_s$ . This uses the NNPDF30 NLO PDF set [31]. Similar to the signal samples, this is combined with PYTHIA8 8.230 [29] for parton shower modelling using the NNPDF23 LO PDF set [37].

### 4.2.3 Data

This analysis uses data from the full Run-2 period, i.e. from 2015-2018, with individual runs selected using the good runs lists:

- data15\_13TeV.periodAllYear\_DetStatus-v89-pro21-02\_Unknown\_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- data16\_13TeV.periodAllYear\_DetStatus-v89-pro21-01\_DQDefects-00-02-04\_PHYS\_StandardGRL\_All\_Good\_25ns.xml
- data17\_13TeV.periodAllYear\_DetStatus-v99-pro22-01\_Unknown\_PHYS\_StandardGRL\_All\_Good\_25ns\_TriggerNo17e33prim.xml
- data18\_13TeV.periodAllYear\_DetStatus-v102-pro22-04\_Unknown\_PHYS\_StandardGRL\_All\_Good\_25ns\_TriggerNo17e33prim.xml

This results in a total integrated luminosity of  $138.5 \text{ fb}^{-1}$ .

## MC Scaling to Match Data Luminosity

The events in MC simulated samples above must be re-weighted in order to match the total integrated luminosity of the ATLAS data used. In order to achieve this, the weight of each MC event is given by:

$$\text{weight} = \frac{\sigma \times (138.5 \text{ fb}^{-1}) \times (k) \times (\text{filt}_{\text{eff}}) \times (\text{GenWeight}) \times (w_f)}{\sum_{\text{Events in AOD}} \text{GenWeight}} \quad (4.1)$$

where

- $\sigma$  is the cross section for the simulated process in fb,
- $k$  is the k-factor, which corrects for the omission of higher-order terms in event generation,
- $\text{filt}_{\text{eff}}$  is the filter efficiency, the fraction of events passing any filters applied during event generation to emphasize a certain phase space,
- GenWeight is the generator weight from MC production,
- and  $w_f$  is a set of weight factors given by:

$$w_f = \text{EleWt} \times \text{MuoWt} \times \text{JetWtJVT} \times \text{prwWt} \times \text{TriggerWts} \quad (4.2)$$

where the various weights apply small corrections from the generation and reconstruction process.

## 4.3 Object Definition

Once events have been simulated or produced in the ATLAS detector, the information contained in them must be reconstructed into objects which can be analyzed. These objects represent particles or groups of particles, and are used to fully or partially reconstruct events and search for signal signatures. The following section will detail the definition of physics objects used in this analysis.

### 4.3.1 Muons

From the ATLAS detector, muons are reconstructed from measurements in the Inner Detector and the Muon Spectrometer. For this analysis, two different definitions of

muon candidates are used. “*Baseline*” muons use slightly looser selection criteria than “*signal*” muons, which are designed for high purity. *Baseline* muons are selected using the *Loose* identification criteria as defined in [38], which is designed to maximize efficiency. *Signal* muons, meanwhile, use the *Medium* identification criteria, and must also be isolated from other signatures according to the *Tight* track-based working point, with variable radius dependent on  $p_T$ . Additionally, all muons are rejected if they are considered “badly reconstructed” or “cosmic”, and track-to-vertex association cuts are applied. Table 4.1 summarizes the muon selection criteria.

Table 4.1: Muon object definition criteria

| Criterion                   | Baseline Muon   | Signal Muon   |
|-----------------------------|---|---|
| Identification              | Loose   | Medium  |
| Isolation                   | -   | TightTrackOnly_VarRad   |
| Pseudorapidity              | $ \eta  < 2.7$  | $ \eta  < 2.5$  |
| $p_T$                       | $p_T > 7 \text{ GeV}$   | $p_T > 7 \text{ GeV}$   |
| Veto                        | Cosmic or Bad   | Cosmic or Bad   |
| Track-to-vertex association | $\left  \frac{d_0}{\sigma_{d_0}} \right  < 3$<br>$\left  \Delta z_0^{\text{BL}} \sin \theta \right  < 0.5 \text{ mm}$ | $\left  \frac{d_0}{\sigma_{d_0}} \right  < 3$<br>$\left  \Delta z_0^{\text{BL}} \sin \theta \right  < 0.5 \text{ mm}$ |

### 4.3.2 Electrons

Electrons are reconstructed by associating energy deposits in the EM calorimeter with tracks in the inner detector. Similar to muon candidates, two different definitions are used, with “*baseline*” electrons requiring slightly less selective criteria than “*signal*” electrons. *Baseline* and *signal* electrons are selected using a likelihood-based identification described in [39]. *Signal* electrons require the medium identification criteria, while *baseline* electrons require the loose criteria as well as a hit in the innermost layer of the pixel detector. Additionally, *baseline* and *signal* electrons must both satisfy the fixed-cut *Loose* isolation requirement. Like for muons, track-to-vertex association cuts are also applied. Table 4.2 summarizes the electron selection criteria.

Table 4.2: Electron object definition criteria

| Criterion                   | Baseline Electron   | Signal Electron   |
|-----------------------------|---|---|
| Identification              | LooseAndBLayerLLH   | MediumLLH   |
| Isolation                   | FCLoose   | FCLoose   |
| Pseudorapidity              | $ \eta  < 2.47$   | $ \eta  < 2.47$   |
| $p_T$                       | $p_T > 7 \text{ GeV}$   | $p_T > 7 \text{ GeV}$   |
| Track-to-vertex association | $\left  \frac{d_0}{\sigma_{d_0}} \right  < 5$<br>$\left  \Delta z_0^{\text{BL}} \sin \theta \right  < 0.5 \text{ mm}$ | $\left  \frac{d_0}{\sigma_{d_0}} \right  < 5$<br>$\left  \Delta z_0^{\text{BL}} \sin \theta \right  < 0.5 \text{ mm}$ |

### 4.3.3 Small-radius ( $R = 0.4$ ) Jets

In this analysis, small- $R$  jets are used to reconstruct the hadronically-decaying  $W$  boson in some analysis regions. Small- $R$  jets are reconstructed using particle-flow objects [40] clustered with the anti- $k_t$  algorithm [41] with radius  $R = 0.4$ . A jet of radius  $R$  falls within a cone satisfying:

$$\Delta R(\text{axis, slant}) = R \quad (4.3)$$

where  $\Delta R$  between two points is given by:

$$\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} \quad (4.4)$$

with  $\phi$  denoting the azimuthal angle about the beam axis, and  $\eta$  denoting the pseudorapidity given by  $\eta = -\ln[\tan(\frac{\theta}{2})]$  where  $\theta$  is the angle from the beam axis. Following reconstruction, only jets with  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.5 \text{ GeV}$  are used for analysis. Additionally, in order to suppress noise from pileup, the jet vertex tagger [42] is applied with the **Tight** working point.

### **b-jets**

For this analysis, it is also beneficial to identify jets originating from bottom ( $b$ ) quarks. In signal regions these jets are vetoed to reduce background. This is performed using the DL1r [43]  $b$ -tagging algorithm, which employs a deep-learning neural network. A 77% efficient tagger working point is used. Table 4.3 summarizes  $R = 0.4$  jet criteria and tagging.

Table 4.3:  $R = 0.4$  jets object definition criteria

| <b>Criterion</b>     | <b>Requirement</b>     |
|----------------------|------------------------|
| Collection           | AntiKt4EMPFJet         |
| Pseudorapidity       | $ \eta  < 2.5$         |
| $p_T$                | $p_T > 20 \text{ GeV}$ |
| Jet-Vertex-Tagger WP | Tight                  |
| $b$ -tagger          | DL1r                   |

#### 4.3.4 Track-Assisted-Reclustered Jets

Track-Assisted-Reclustered (TAR) jets [44] are used in this analysis to reconstruct hadronically-decaying  $W$  boson candidates in some regions.  $R = 0.2$  jets, constructed with the anti- $k_t$  algorithm [41] from locally-calibrated topological clusters [45] are used as input for building TAR Jets. As well, the TAR algorithm uses tracks fulfilling quality criteria outlined in Table 4.4. The TAR building process involves constructing large- $R$  jets from the constituent small- $R$  jets. Tracks are then matched with the input jets and rescaled using the  $p_T$  of the matched jets. The kinematic properties of the TAR jets are then calculated from the constituent  $R = 0.2$  jets, while the jet substructure variables are calculated from the constituent tracks. Figure 4.4 shows a visual summary of the base TAR algorithm, and the following steps outline the algorithm in more detail:

- Tracks and calibrated anti- $k_t$   $R = 0.2$  jets are chosen as input to the algorithm.
- The anti- $k_t$   $R = 0.2$  jets are reclustered using the anti- $k_t$  algorithm into  $R = 1.0$  jets and trimmed using the  $p_T$  fraction  $f_{cut} = 0.05$ .
- Input tracks are matched to  $R = 0.2$  jets, if possible, using ghost association [46].
- Tracks which remain unassociated are matched to the nearest anti- $k_t$   $R = 0.2$  jet within  $\Delta R < 0.3$
- The  $p_T$  of each track is rescaled using the  $p_T$  of the jet to which it is matched, via the equation:

$$p_T^{\text{track, new}} = p_T^{\text{track, old}} \times \frac{p_T^{\text{subject } j}}{\sum_{i \in j} p_T^{\text{track } i}}, \quad (4.5)$$

where  $j$  is the  $R = 0.2$  subjet that the track being rescaled is matched with, and the index  $i$  runs over all tracks matched to that subjet. This rescaling accounts for the missing neutral momentum, which is measured at calorimeter level but is not present at tracker level.

- Finally, jet substructure variables and  $m^{\text{TAR}}$  are calculated using the rescaled matched tracks.

The parameters of the TAR algorithm used are summarized in Table 4.4.

Table 4.4: TAR jet reconstruction parameters

|                             |   |
|-----------------------------|---|
| Input track selection       | Loose quality<br>$p_T > 0.5 \text{ GeV}$<br>$ \eta  < 2.5$                        |
| Track-to-vertex association | $ d_0  < 2 \text{ mm}$<br>$ \Delta z_0^{\text{BL}} \sin \theta  < 3.0 \text{ mm}$ |
| Input jet selection         | $R = 0.2$ topo jets<br>$p_T > 20 \text{ GeV}$<br>$ \eta  < 2.5$                   |
| Reclustering radius         | $R = 1.0$   |
| TAR jet $p_T$               | $p_T^{\text{TAR}} > 100 \text{ GeV}$  |
| Trimming radius             | $R = 0.2$   |
| Trimming $p_T$ fraction     | $f_{\text{cut}} = 0.05$   |
| Track-to-jet association    | $\Delta R(\text{jet}, \text{track}) < 0.3$  |

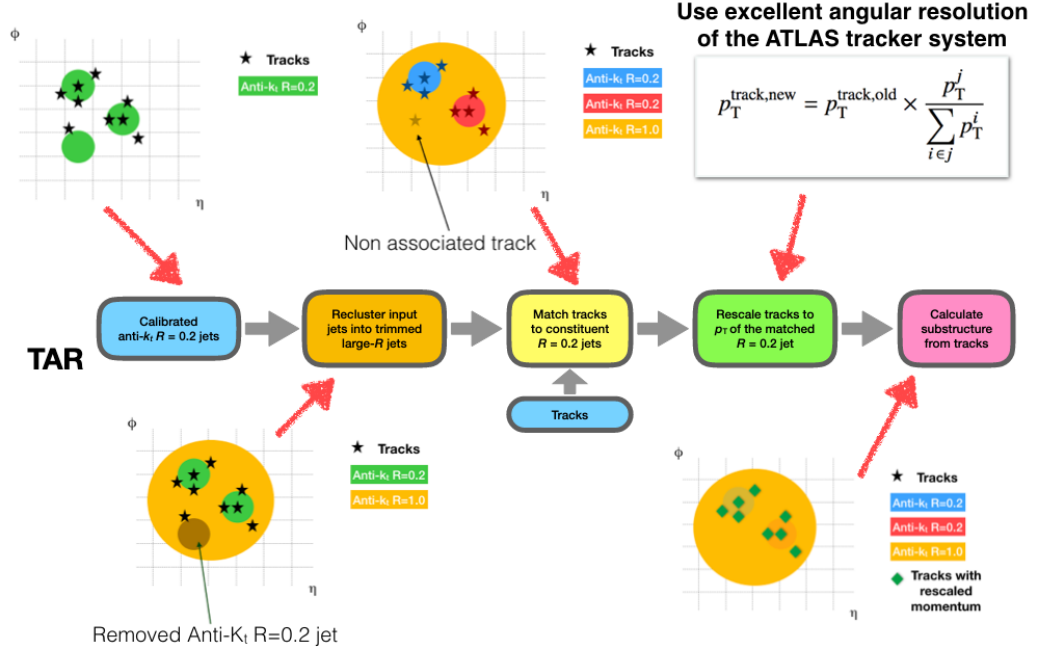


Figure 4.4: A visual summary of the TAR Jet algorithm. [47]

### 4.3.5 Missing Transverse Momentum ( $E_T^{miss}$ )

In the ATLAS detector, the beams collide directly along the longitudinal axis. As a result of the conservation of momentum, all decay products from the collision should thus have a total transverse momentum of 0. Using this information, missing transverse momentum ( $E_T^{miss}$ ) can be used to represent undetected objects. In the signal model described in subsection 2.2.2 both the neutrino from the leptonically decaying  $W$ -boson and the dark matter would not be directly detected in ATLAS. It is impossible to disentangle these objects, and they are reconstructed together using  $E_T^{miss}$ .  $E_T^{miss}$  is calculated for this analysis using baseline electrons and muons as described in subsections 4.3.2 and 4.3.1, as well as  $R = 0.4$  jets with the **Tight** working point applied. No photons or  $\tau$ -leptons are included in the calculation. An additional soft term, which is calculated from tracks not associated with the reconstructed objects, is also added.

#### $E_T^{miss}$ Significance

The significance of the  $E_T^{miss}$  measurement is also assessed using the object-based  $E_T^{miss}$  significance  $\mathcal{S}$ . It is calculated as described in [48] from a combination of the

uncertainties on the reconstructed objects and soft term used to calculate  $E_T^{miss}$ , as well as a pileup correction. A high  $\mathcal{S}$  indicates a higher degree of confidence that the  $E_T^{miss}$  measurement of an event comes from an unseen object rather than from particle mismeasurement.

## 4.4 Data and MC Preparation

After reconstruction (or simulated reconstruction) in the ATLAS detector, data and MC events are stored as Analysis Object Data (AOD) files which contain all reconstructed information from each event. Data and MC samples used for this analysis are initially in ATLAS Derived Analysis Object Data (DAOD format) which are created by skimming and slimming the AODs according to the EXOT27 job options in the Athena framework [49]. The DAODs include information about:

- electrons, photons, muons, tau jets,
- $R = 1.0$  jets,
- AntiKt4EMTopoJets, AntiKt4EMPFLOWJets, AntiKt2LCTopoJets,
- jet b-tagging, and
- InDetTrackParticles associated with TopoJets

The DAOD files are still too large to be convenient for analysis, and some analysis variables have not yet been calculated. In order to further process the DAODs we use the XAMPP analysis framework, which is built on top of the SUSYTools framework within Athena. We use a customized version of the XAMPP framework (XAMPPMonoSlep) maintained by the analysis team to calculate all necessary analysis variables and selects physics objects according to the definitions in Section 4.3. We used the following set of “preselection” criteria to trim the number of events to a convenient size prior to optimization, without cutting into any regions of sensitivity:

- passed  $E_T^{miss}$  trigger,
- (1 signal muon or 1 signal electron) and no additional baseline muons or electrons,
- $E_T^{miss} > 150 \text{ GeV}$ ,

- $\mathcal{S} > 5$ ,
- transverse mass of  $E_T^{miss}$  and lepton  $m_T(\ell, E_T^{miss}) > 50 \text{ GeV}$ ,
- veto on  $b$ -tagged  $R = 0.4$  jets, and
- $N(\text{TAR Jets}) > 0$  or  $N(\text{Jets}) > 1$ .

The trigger requirements beyond the preselection criteria are discussed in further detail in Section 5.7. After processing by the XAMPPMonoSlep framework, the data is stored in CERN ROOT files known as “ntuples”, which are further slimmed to include only necessary variables using XAMPPPlotting. At this stage the data and MC samples are ready to be analyzed.

# Chapter 5

## Analysis

We analyze ATLAS data to search for events which contain a pair of  $W$  bosons with one decaying to  $l\nu$  and the other decaying to  $q\bar{q}$ , and with the pair consistent with having been produced by a new resonance denoted by  $s$  (as described in Chapter 2 subsection 2.2). Following the centrally-performed reconstruction of ATLAS proton-proton collision data, we analyze the reconstructed physics objects to search for evidence of new physics in this mono- $s \rightarrow WW$  semileptonic channel. This thesis focuses on the design of the most sensitive “merged” signal region of the possible phase space of events, as well as the addition of two control regions to constrain the  $t\bar{t}$  background contribution. Other regions used to specify potential signal and the various types of background within the analysis, as well as still further regions used for control and for validation of the analysis, are described in more brief detail in order to provide context to decision-making and a fuller picture of the analysis.

### 5.1 Cut and Count Analysis

This analysis uses the “cut and count” strategy to search for evidence of the signal model. In order to perform this search, we define a set of selection criteria to slice or “cut” the data according to various event variables, such as the amount of missing transverse momentum. We choose these selection criteria in order to maximize the expected number of signal events, and minimize the expected number of standard model background events in the region, based on the MC simulated data. While designing the selection criteria, real ATLAS data is hidden, or blinded, in order to avoid biasing the selection. The final area of phase space defined by the selection criteria

is known as the “signal region”. As the final stage of an analysis, the real ATLAS data can be compared with the simulated MC data, and a statistical conclusion can be reached about the likelihood of the existence of the signal model. The more signal and less background that are present in the signal region, the stronger the statistical evidence for existence of the beyond-SM signal channel will be.

In order to improve the accuracy of the analysis and reduce the uncertainty on the amount of predicted standard model background, a “control region” may also be used. This is a region not overlapping with the signal region which is designed to have similar kinematic properties to signal events, but be more pure in one or more standard model backgrounds and much less rich in actual expected signal events. In this region, we can compare the number of MC simulated background events and ATLAS data events, and any differences provide additional information to correct and constrain the expected number of background events in the signal region.

## 5.2 Analysis Strategy

In this analysis, there are two analysis channels which are combined statistically during fitting. These channels are distinguished by the reconstruction of the dijet system from the hadronically-decaying  $W$  boson, and are named the “merged” channel and the “resolved” channel. In the “merged” channel, we reconstruct the dijet system with a single  $R = 1.0$  Track-Assisted-Reclustered (TAR) jet, which generally means that the dijet system is more boosted, with the  $s$  decay products closer together. In the less-boosted “resolved” channel we reconstruct the dijet system using a pair of  $R = 0.4$  jets. Figure 5.1 provides a visual demonstration of this difference. In each analysis channel there are three analysis regions: the semileptonic signal region, a control region designed to constrain the  $t\bar{t}$  background, and a control region designed to constrain the  $W$ +jets background. To ensure that “merged” and “resolved” regions do not overlap, an event recycling strategy is employed where only events failing the selection for all “merged” regions are considered for selection in “resolved” regions.

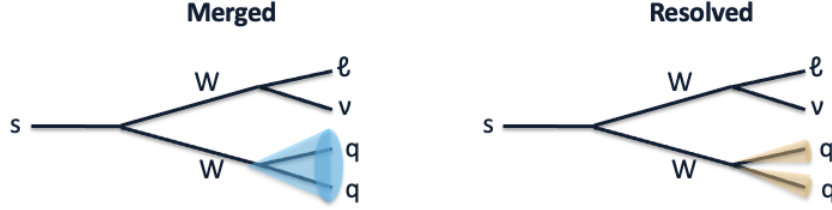


Figure 5.1: Schematic representation of the differing reconstruction technique between merged and resolved regions.

## 5.3 Merged Signal Region

### 5.3.1 Signal and Background Characterization

The first step towards defining signal region selection criteria is to characterize signal and background events, while searching for exploitable differences. Event characteristics explored include the relative positions of analysis objects, the transverse momenta and masses of analysis objects, along with various jet substructure variables. Figures A.2, A.3, and A.4 in Appendix section A.1 show distributions of signal and background events for many of the variables studied at preselection level.

### 5.3.2 TAR Jet Lepton Disentanglement

Reconstructing the hadronically-decaying  $W$  boson as accurately as possible is key in this analysis. As a result, it is given special attention, especially in the more sensitive merged channel. In this channel, the  $W$  is reconstructed by a single  $R = 1.0$  TAR jet. Due to the boosted nature of the  $s$ -decay, however, the leptonically-decaying  $W$  often lies in or very near to this jet. As a result, the charged lepton often overlaps the TAR jet, which can lead to difficulty in jet reconstruction.

In order to resolve this difficulty, we made modifications to the TAR jet building process to disentangle overlapping leptons. We identify input tracks and jets that are likely to be attributable to a final state lepton and not the hadronic  $W$  decay, and remove them. To achieve this, we remove tracks associated with a baseline electron or muon from the input track collection. We then also remove any anti- $k_t$   $R = 0.2$  jet overlapping with a baseline electron (defined here as having  $\Delta R(lep, jet) < 0.2$ ) prior to reclustering into  $R = 1.0$  jets. We do not remove  $R = 0.2$  jets overlapping muons,

as muons do not leave large a calorimeter signature and are therefore unlikely to fake hadronic activity. This results in the following updated TAR jet building algorithm, where steps with a (\*) are added to disentangle leptons:

- Tracks and calibrated anti- $k_t$   $R = 0.2$  jets are chosen as input to the algorithm.
- Tracks associated with a baseline muon or electron are removed from the input collection (\*).
- anti- $k_t$   $R = 0.2$  jets overlapping with a baseline electron ( $\Delta R < 0.2$ ) are removed from the input collection (\*).
- The remaining anti- $k_t$   $R = 0.2$  jets are reclustered into  $R = 1.0$  jets using the anti- $k_t$  algorithm, and trimmed using the  $p_T$  fraction  $f_{cut} = 0.05$ .
- Input tracks are matched to  $R = 0.2$  jets if possible, using ghost association.
- Tracks which remain unassociated are matched to the nearest anti- $k_t$   $R = 0.2$  jet within  $\Delta R < 0.3$ .
- The  $p_T$  of each track is rescaled using the  $p_T$  of the jet to which it is matched using the equation:

$$p_T^{\text{track, new}} = p_T^{\text{track, old}} \times \frac{p_T^{\text{subject } j}}{\sum_{i \in j} p_T^{\text{track } i}}, \quad (5.1)$$

where  $j$  is the  $R = 0.2$  subject that the track being rescaled is matched with, and the index  $i$  runs over all tracks matched to that subject. This rescaling accounts for the momentum of neutral particles (except for neutrinos), which is measured at calorimeter level but is not visible in the tracker.

- Finally, jet substructure variables and  $m^{\text{TAR}}$  are calculated using the rescaled matched tracks.

In order to study the potential benefit of lepton-disentanglement, we produced near-identical sets of signal MC samples, with one implementing disentanglement and the other not. We then compared the mass of the reconstructed jets in the two samples in a region where a lepton overlaps the  $R = 1.0$  TAR Jet ( $\Delta R(e, \text{TAR}) < 1$ ), to determine how well the  $W$  boson mass is reconstructed in this case. Figure 5.2 demonstrates the improvement in mass resolution achieved by disentangling leptons

from TAR jets. A clearly enhanced resolution in the TAR jet mass peak around the  $W$  mass is visible for the lepton-disentangled jets in the electron channel. This significantly improves sensitivity in the merged signal region by allowing a much tighter selection on the TAR jet mass. Very little changes in the muon channel.

We additionally compared the performance of building the TAR jets from constituent anti- $k_t$   $R = 0.4$  jets, rather than anti- $k_t$   $R = 0.2$  jets. In the electron channel we found that for high  $s$  mass,  $W$  mass reconstruction performance was similar between the two methods, but for low  $s$  mass using anti- $k_t$   $R = 0.4$  jets significantly impairs resolution around the  $W$  mass. Again in the muon channel we observed few significant differences. Figure 5.3 shows a comparison of the TAR jet mass distribution built from anti- $k_t$   $R = 0.4$  and anti- $k_t$   $R = 0.2$  jets, for two selected signal points, in a signal-enriched region with 1 electron,  $\Delta R(e, \text{TAR}) < 1$ , and  $m_T(\ell, E_T^{\text{miss}}) > 150$  GeV. This clearly demonstrates the improved mass resolution for the signal point at  $m_s = 160$  GeV.

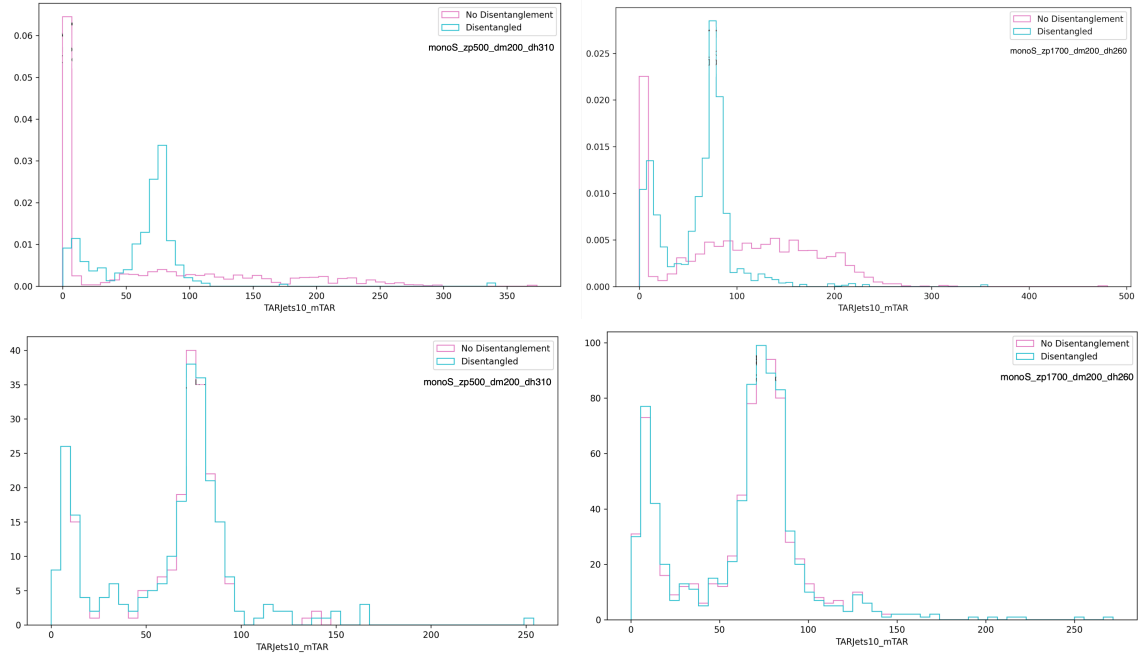


Figure 5.2: Normalized  $R = 1.0$  TAR jet mass distributions with and without lepton disentanglement applied for four representative signal points. **Upper row:** electron channel, **lower row:** muon channel.

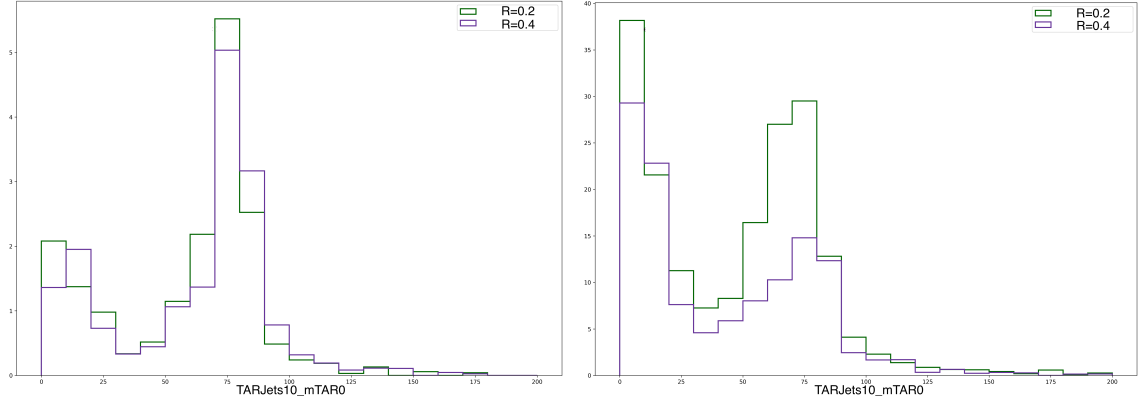


Figure 5.3: TAR jet mass distributions for two representative signal points, compared with constituent anti- $k_t$   $R = 0.4$  vs. anti- $k_t$   $R = 0.2$  jets. Left: sample signal point at  $(m_s, m_{Z'}) = (310, 500)$  GeV. Right: sample signal point at  $(m_s, m_{Z'}) = (160, 1700)$  GeV.

### 5.3.3 Merged Signal Region Optimization

#### Figure of Merit

We optimized the “merged” signal region using the expected discovery significance  $Z$  as the figure of merit [50]. This is given by:

$$Z = \left[ 2(s+b) \left( \ln \left[ \frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[ 1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{\frac{1}{2}}, \quad (5.2)$$

where  $b$  is the expected total number of background events,  $s$  is the expected total number of signal events, and  $\sigma_b$  is the uncertainty on the expected total number of background events. Many MC generators, including Sherpa, produce widely varying event weights, including negative weights. As a result, the statistical uncertainty of a sample of weighted events is given by:

$$\sigma = \sqrt{\sum (w^2)}. \quad (5.3)$$

For a given set of selection criteria and signal point, this metric is calculated from the MC simulated data by applying the criteria, counting the number of remaining signal and background events, and computing their uncertainties, then calculating the significance. It provides a metric to assess the expected sensitivity of the signal region to the signal model.

## Optimization Approach

In order to optimize the signal region selection, we employed a combination of an iterative visual approach and a computational approach. Visually, we studied the distributions of analysis variables, comparing signal and background distributions, to find areas of good discrimination where a cut placement might improve sensitivity. We then compiled a list of possible cuts in each of these areas of discrimination to form a basis for the computational optimization. To determine the optimal combination of these possible cuts, we created a script to test all possible combinations while computing the expected significance of each. When calculating significance during optimization, in addition to statistical uncertainty we applied a flat 20% systematic uncertainty on the number of background events to more realistically reflect the expected total uncertainty. Different signal points in the  $m_s$ - $m_{Z'}$  plane have different kinematic properties, and cross-sections and therefore have different optimal selection. As a result, we selected the following four signal points at the edge of the search window for optimization:

- $(m_s, m_{Z'}) = (310 \text{ GeV}, 500 \text{ GeV}),$
- $(m_s, m_{Z'}) = (335 \text{ GeV}, 1000 \text{ GeV}),$
- $(m_s, m_{Z'}) = (285 \text{ GeV}, 1700 \text{ GeV}),$
- and  $(m_s, m_{Z'}) = (210 \text{ GeV}, 2100 \text{ GeV}).$

We then considered the cut combination with the highest mean expected significance over these four points to be optimal. In early iterations of optimization, the number of signal events selected using the optimal selection criteria was undesirably low. As a result, we implemented an additional requirement forcing the chosen criteria to maintain an expected yield of at least 15 signal events in the “merged” signal region.

In order to avoid over-tuning the optimization on single events or statistical fluctuations, we spaced the lattice of possible cuts widely enough to allow several MC events to fall between each placement. Optimizing a selection in this manner also creates a systematic bias toward an under-prediction of background in the signal region, because a low expected background yield is considered desirable. To counter this bias, we used statistically-independent optimization and fitting data sets. We performed all optimization using a set of MC samples produced when using Sherpa

2.2.1 as a training set. A set of samples produced when using Sherpa 2.2.10 then provides a statistically-independent and more statistically rich data set which we used for validation and fit model testing. This means that the MC data used for fit model testing is isolated from the optimization of selection criteria.

### Optimization Variables

We examined many analysis variables for potential sensitivity, and selected those having the best discrimination potential to be tested in the final optimization. Presented below is a list of variables tested and their meaning:

- $E_T^{miss}$  and  $\mathcal{S}$ : Missing transverse momentum; and the significance of its difference from a value of zero.
- $p_T(\ell)$ : Transverse momentum of signal lepton.
- $m_T(\ell, E_T^{miss})$ : Transverse mass of lepton and  $E_T^{miss}$  system, given by

$$m_T(\ell, E_T^{miss}) = \sqrt{2p_{T,\ell}E_T^{miss} \left(1 - \cos(\phi_\ell - \phi_{E_T^{miss}})\right)} \quad (5.4)$$

- $m^{\text{TAR Jet}}$ : Mass of the  $p_T$ -leading TAR jet calculated from the rescaled matched tracks.
- $p_T^{\text{TAR Jet}}$ : Transverse momentum of the leading TAR jet.
- $D_2^{\beta=1}(\text{TAR Jet})$ : Energy correlation function of the TAR jet, which helps to identify two-pronged jet substructure [51].
- **TAR Jet  $\tau_{42}^{WTA}$  and  $\tau_{21}^{WTA}$** : “ $n$ -Subjettiness” ratios of the leading TAR jet, to identify two-pronged substructure [52].

In addition, we considered a two-dimensional cut above or below a line with variable slope and intercept in the  $E_T^{miss}$ - $p_T(\ell)$  plane during optimization.

### Optimized Selection

We found the maximum mean expected significance across the four selected signal points to be achieved using the selection shown in Table 5.1.

| variable                                | requirement            | reason  |
|---|------------------------|---|
| $N(\text{TAR Jets})$                    | $> 0$                  | $W_{\text{Cand}}$ reconstruction                      |
| $m_{\text{T}}(\ell, E_T^{\text{miss}})$ | $> 220 \text{ GeV}$    | $W$ +jets background reduction                        |
| $m^{\text{TAR Jet}}$                    | $[68, 89] \text{ GeV}$ | $W_{\text{Cand}}$ reconstruction                      |
| $\mathcal{S}$                           | $> 16$                 | Select for high $E_T^{\text{miss}}$                   |
| $\Delta R(\text{TAR Jet}, \ell)$        | $< 1.2$                | Select for boosted $W_{\text{Cand}} + \ell\nu$ system |
| $D_2^{\beta=1}(\text{TAR Jet})$         | $< 1.1$                | 2-pronged $W \rightarrow qq$ reconstruction           |

Table 5.1: Optimized selection criteria for the “merged” category signal region.

In order to validate this choice of selection criteria, we produced “ $N - 1$ ” plots which show the distribution of a variable with all but the cut on that variable applied. We also calculated the expected significance of placing a cut at each bin edge in the “ $N - 1$ ” histogram. We then compared the optimized cut placement with the cut placement showing maximal expected significance in the plot. These plots were produced using the testing and validation data ( $W$ +jets and  $Z$ +jets samples produced in Sherpa 2.2.10), and are shown in Figure 5.4. They show that the cuts optimized using Sherpa 2.2.1 samples are still placed at optimal or near-optimal locations using the final samples, and therefore validate the choice of selection criteria. For visual interpretation of these plots, a single event with a weight of (-6) was removed from the background MC, as its weight overwhelmed the shape of the background distributions and produced sharp spikes in estimated significance. This event is **not** removed later during fitting. A secondary set of  $N-1$  plots with this event included are shown in Appendix section A.4

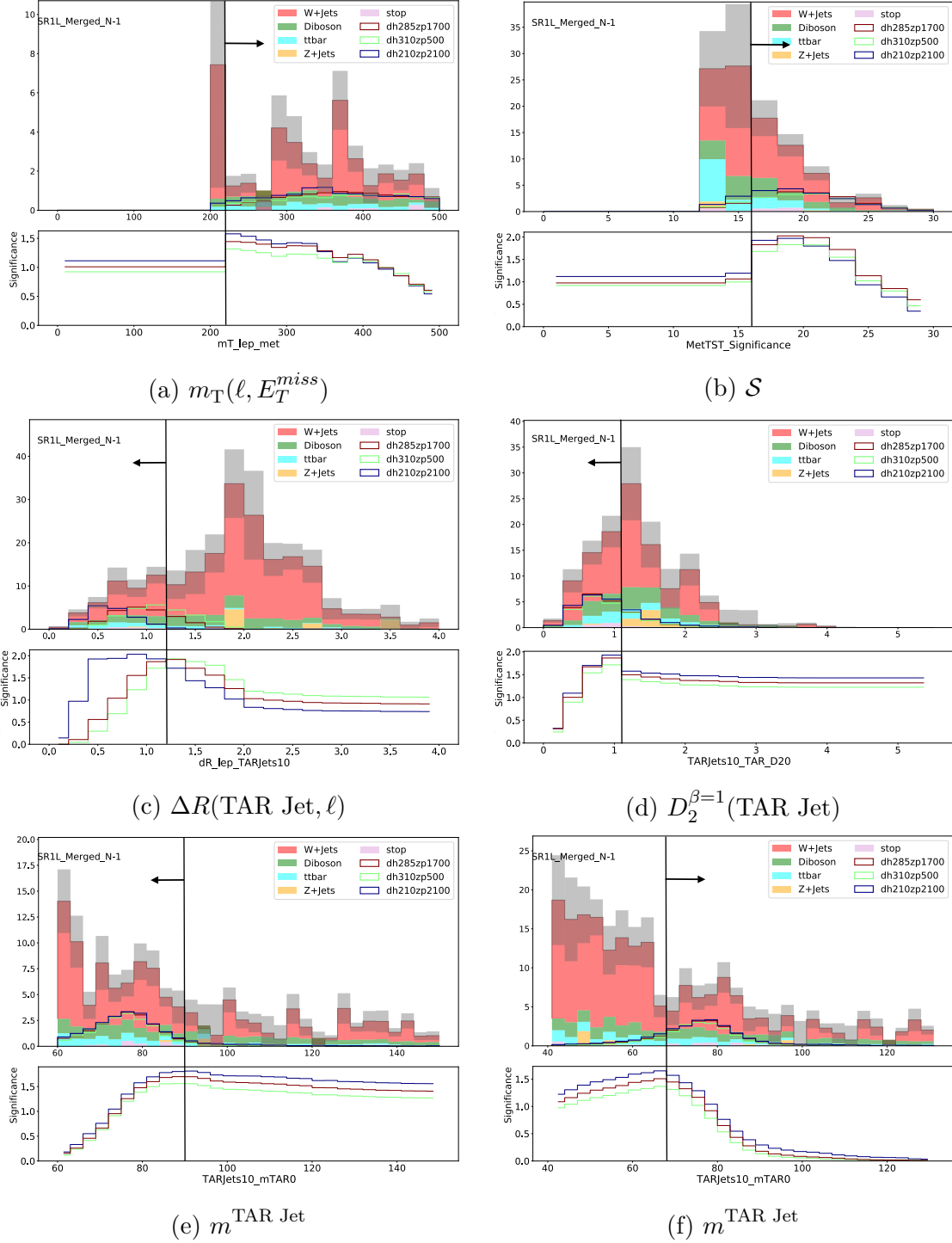


Figure 5.4: “ $N - 1$ ” plots in the “merged” signal region showing the distribution of variables with all cuts applied except that of the variable shown. Grey bands represent MC statistical uncertainty on each bin.

## 5.4 Resolved Signal Region

Due to the highly boosted nature of most signal events, the hadronic  $W$  decay products are often best reconstructed as an  $R = 1.0$  TAR jet. To capture events for which this is not the best reconstruction we define our “resolved” signal region. Being a secondary region, the “resolved” signal region is considerably less sensitive to the signal model than the “merged” signal region, but does provide some enhancement especially at low  $s$  mass where the least boost is experienced. The “resolved” signal region selection criteria are listed in Table 5.2.

Table 5.2: Overview of “resolved” SR cuts.

| cut  | reason  |
|--|---|
| Fails “merged” analysis region selections        | Recycling strategy                                    |
| $N(\text{Jets}) \geq 2$                          | $W_{\text{Cand}}$ reconstruction                      |
| $m(W_{\text{Cand}}) \in [65, 95] \text{ GeV}$    | $W_{\text{Cand}}$ reconstruction                      |
| $p_T(W_{\text{Cand}}) > 150 \text{ GeV}$         | $W$ +jets Background Reduction                        |
| $m_T(\ell, E_T^{\text{miss}}) > 200 \text{ GeV}$ | $W$ +jets Background Reduction                        |
| $E_T^{\text{miss}} > 250 \text{ GeV}$            | Select for high $E_T^{\text{miss}}$                   |
| $\mathcal{S} > 16$                               | Select for good $E_T^{\text{miss}}$                   |
| $\Delta R(W_{\text{Cand}}, \ell) < 1.4$          | Select for boosted $W_{\text{Cand}} + \ell\nu$ system |

## 5.5 $t\bar{t}$ Control Regions

A control region is used to help constrain one or more sources of background in an analysis. It is designed as a region that covers a similar but non-overlapping area of phase space to the signal region, and is expected to contain many fewer signal events. As a result, the number of observed ATLAS data and MC-predicted SM events should match very closely. Any difference provides information about how MC predictions differ from data observations and is used to derive a normalization factor to constrain the SM background predictions of MC in the signal region, and provide a more accurate projection.

To design an effective control region, we attempted to satisfy four important criteria:

- **Non-overlap with the signal region:** This ensures that the sample used to constrain the signal region background is statistically-independent from the

signal region.

- **High  $t\bar{t}$  purity:** This selects for the background of interest in the control region to better individually constrain that background.
- **Low signal contamination:** This ensures that the comparison of ATLAS data and MC in this region is not affected by the signal model and effectively constrains only the SM background.
- **Similar kinematics to SR:** This ensures that a similar region of phase space is being studied in the control and signal regions, so that the normalization factors derived apply well to both regions.

The fact that  $t\bar{t}$  is one of the sub-dominant backgrounds in both signal regions motivated me to create one such control region for each signal region. As a consequence of top quarks decaying primarily to a  $W$  boson and  $b$ -quark, the principal cut excluding  $t\bar{t}$  events from the signal regions is the veto of any events containing a  $b$ -tagged anti- $k_t$   $R = 0.4$  jet. We therefore chose to try defining a similar but non-overlapping region rich in  $t\bar{t}$  events by simply reversing this cut and also requiring at least 1  $b$ -tagged jet. We found this proposed region to be too rich in expected signal events, however, with signal yields as high as 10% of background yields in the “merged” region.

To improve this, we further extended the  $b$ -jet cut to require 2 or more  $b$ -jets in the control region. This change sufficiently reduced signal contamination, but in the “merged” control region it had the additional consequence of substantially reducing the number of events in the region, thereby reducing its statistical power. In response we also loosened the  $\mathcal{S}$  cut in the “merged”  $t\bar{t}$  control region to  $\mathcal{S} > 12$ , considerably improving the  $t\bar{t}$  yield. To inform this choice, we used figure 5.5, which shows the number of weighted  $t\bar{t}$  events expected within the “merged” control region for a given  $\mathcal{S}$  cut.

Table 5.3 summarizes the differences in selection criteria between the “merged” and “resolved” signal and  $t\bar{t}$  control regions. Figures A.6 and A.8 in Appendix section A.2 show comparisons of the distributions of MC and ATLAS data in the  $t\bar{t}$  control regions. These demonstrate a slight over-prediction of background from MC in the “merged” region, which is used during fitting to constrain the  $t\bar{t}$  background in the signal region. The difference between data and MC distributions does not show a strong correlation with any of the analysis variables. To ensure that the kinematic

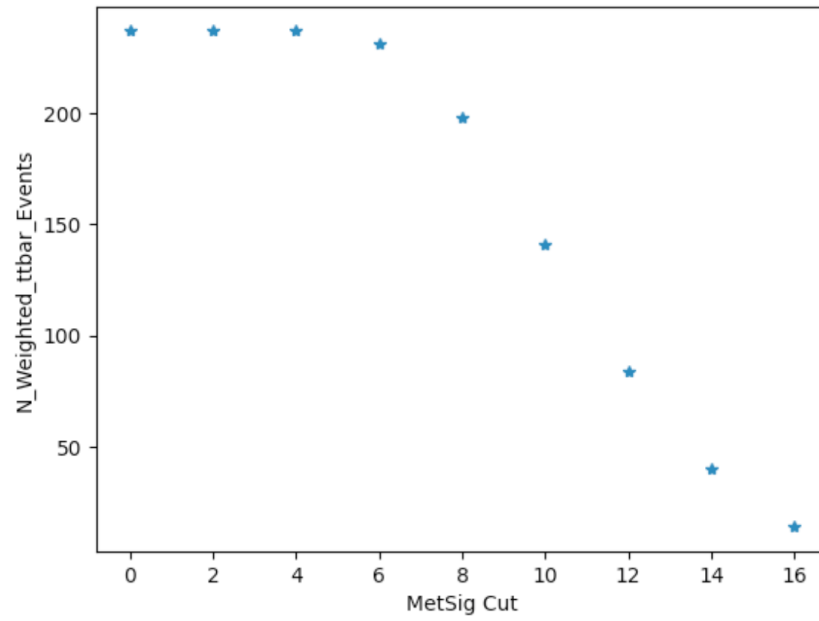


Figure 5.5:  $t\bar{t}$  yield in the “merged”  $t\bar{t}$  control region for various  $E_T^{miss}$  significance cut placements.

| <b>Region</b>        | $N(b\text{-jets})$ Cut | $\mathcal{S}$ Cut  |
|----------------------|------------------------|--------------------|
| <b>Merged SR</b>     | $N(b\text{-jets}) = 0$ | $\mathcal{S} > 16$ |
| <b>Resolved SR</b>   | $N(b\text{-jets}) = 0$ | $\mathcal{S} > 16$ |
| <b>Merged CRtt</b>   | $N(b\text{-jets}) > 1$ | $\mathcal{S} > 12$ |
| <b>Resolved CRtt</b> | $N(b\text{-jets}) > 1$ | $\mathcal{S} > 16$ |

Table 5.3: Summary of differences in selections on  $N(b\text{-jets})$  and  $\mathcal{S}$  between signal regions and  $t\bar{t}$  control regions.

qualities of the signal and control regions are sufficiently similar to provide a good constraint, Figures A.14 and A.16 in Appendix section A.3 show comparisons of distributions of key kinematic variables between the “merged” and “resolved” control and signal regions.

## 5.6 $W$ + jets Control Region

This analysis employs an additional pair of control regions in order to constrain the dominant  $W$ +jets background in the signal regions. The signal regions occupy a realm of phase space very high in  $m_T(\ell, E_T^{miss})$ , and we sought to better constrain the background prediction in the tail of this distribution. To achieve this we defined control regions by reversing the  $\Delta R(\text{TAR Jet}, \ell)$  and  $\Delta R(W_{\text{Cand}}, \ell)$  cuts in the “merged” and “resolved” signal regions. This reverses the direction of the neutrino and charged lepton in the leptonic  $W$  decay without substantially altering the remaining kinematics of the event, keeping the region high in  $m_T(\ell, E_T^{miss})$ . In the “merged”  $W$ +jets control region we again reduced the  $\mathcal{S}$  cut to  $\mathcal{S} > 12$  in order to improve statistical power. We then further extended the  $\Delta R(\text{TAR Jet}, \ell)$  requirement to  $\Delta R(\text{TAR Jet}, \ell) > 1.8$  in the merged control region to reduce signal contamination. Table 5.4 summarizes the differences in selection criteria between the “merged” and “resolved” signal and  $W$ +jets control regions.

Figures A.10 and A.12 in Appendix section A.2 show checks of the distributions of MC and ATLAS data in the  $W$ +jets control regions. As in the  $t\bar{t}$  control regions, no strong correlation is observed with any particular variable.

| Region       | $\Delta R$ Cut                          | $\mathcal{S}$ Cut  |
|--------------|---|--------------------|
| Merged SR    | $\Delta R(\text{TAR Jet}, \ell) < 1.2$  | $\mathcal{S} > 16$ |
| Resolved SR  | $\Delta R(W_{\text{Cand}}, \ell) < 1.4$ | $\mathcal{S} > 16$ |
| Merged CRW   | $\Delta R(\text{TAR Jet}, \ell) > 1.8$  | $\mathcal{S} > 12$ |
| Resolved CRW | $\Delta R(W_{\text{Cand}}, \ell) > 1.4$ | $\mathcal{S} > 16$ |

Table 5.4: Summary of differences in selections on  $\Delta R(W_{\text{Cand}}, \ell)$  and  $\mathcal{S}$  between signal regions and  $W$ +jets control regions.

## 5.7 Trigger Strategy

The ATLAS High Level Triggers (HLT) each represent a set of criteria used by the L2 trigger and event filter to select which events passing the L1 trigger are potentially interesting and should be further analyzed. Throughout the process of optimization and signal and control region definition, we used the unrescaled  $E_T^{\text{miss}}$  triggers listed in Table 5.5, which select events with high  $E_T^{\text{miss}}$ . Upon closer examination, however, we found that these triggers were unexpectedly inefficient even in our analysis regions with  $E_T^{\text{miss}} > 200$  GeV. Figures 5.6 and 5.7 show the  $E_T^{\text{miss}}$  trigger efficiency in the  $t\bar{t}$  control regions in the electron and muon decay channels. In the electron channel we observe full efficiency, but not in the muon channel.

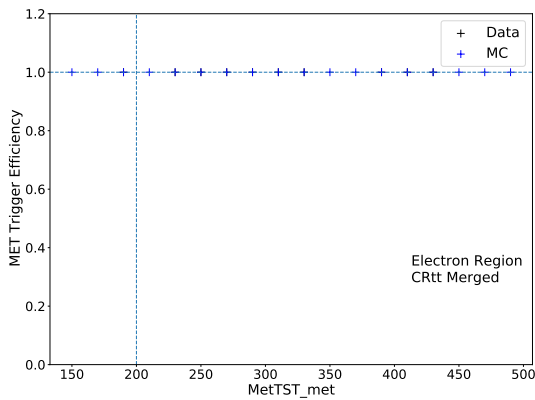
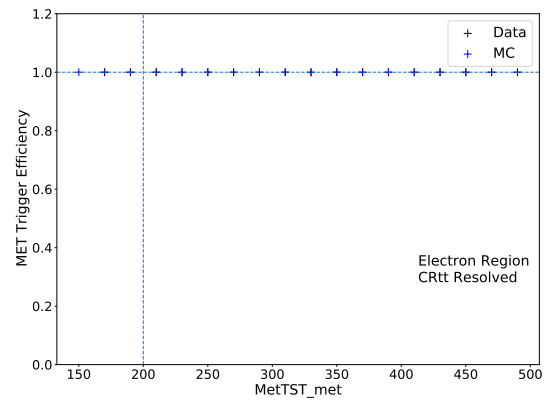
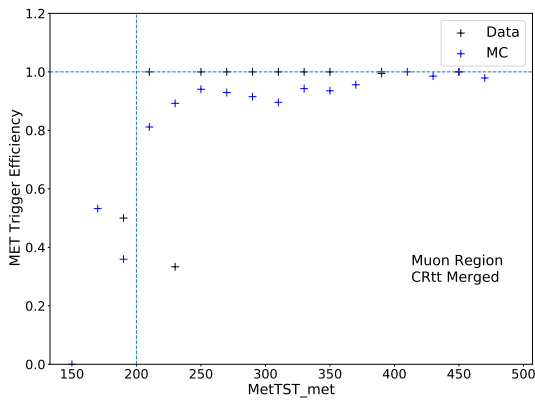
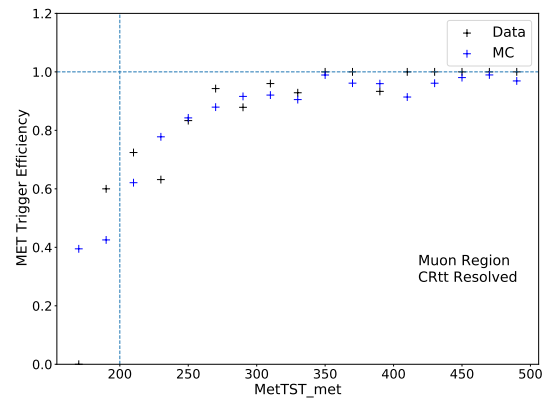
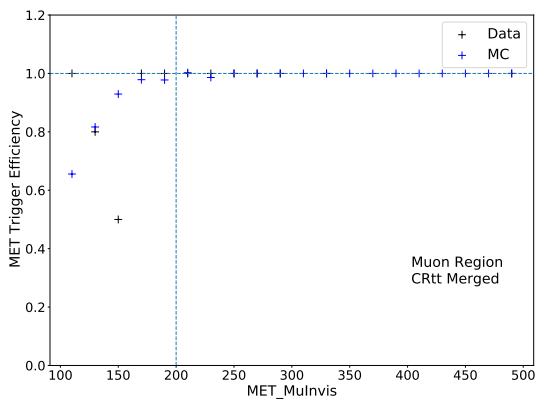
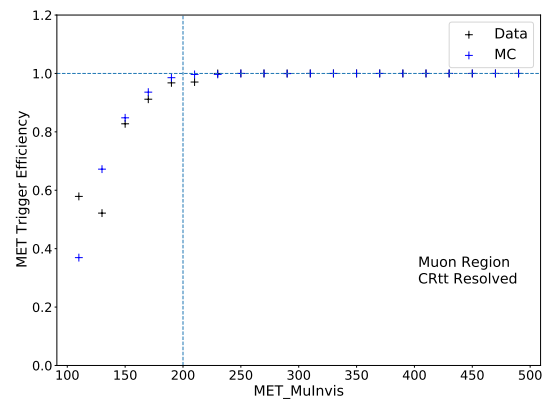
We hypothesized that the inefficiency was caused by high- $p_T$  muons, which are not seen by the trigger-level  $E_T^{\text{miss}}$  calculation. As a result, if a muon travels in the opposite direction to the true  $E_T^{\text{miss}}$ , the measured trigger-level  $E_T^{\text{miss}}$  will be reduced and the trigger criteria may be failed. To confirm this, Figure 5.8 shows the  $E_T^{\text{miss}}$  trigger efficiency plotted against  $E_T^{\text{miss}}$  with signal muons treated as invisible (i.e. having their  $p_T$  subtracted from the  $E_T^{\text{miss}}$ ). In this case full or near-full trigger efficiency for  $E_T^{\text{miss}}(\text{no}\mu) > 200$  GeV is recovered. We find it desirable to obtain full trigger efficiency in order to avoid the uncertainty associated with calculating trigger scale factors to match the trigger behaviour of data and MC events. Therefore, to counter the trigger inefficiency we moved to a logical OR between the unrescaled  $E_T^{\text{miss}}$  triggers and the single muon triggers in Table 5.6. This is designed to ensure that the events with a high- $p_T$  muon that fail the  $E_T^{\text{miss}}$  trigger will pass the single muon trigger and be considered. Figure 5.9 shows that this recovers full trigger efficiency.

| Period          | MET Trigger                 |
|-----------------|-----------------------------|
| 2015            | HLT_XE70_MHT                |
| 2016<br>(A-D3)  | HLT_XE90_MHT_L1XE50         |
| 2016<br>(D4-F1) | HLT_XE100_MHT_L1XE50        |
| 2016<br>(F2-)   | HLT_XE110_MHT_L1XE50        |
| 2017<br>(B-D5)  | HLT_XE110_PUFIT_L1XE55      |
| 2017<br>(D6-K)  | HLT_XE110_PUFIT_L1XE50      |
| 2018<br>(B-C5)  | HLT_XE110_PUFIT_XE70_L1XE50 |
| 2018<br>(C5-)   | HLT_XE110_PUFIT_XE65_L1XE50 |

Table 5.5: ATLAS unrescaled  $E_T^{miss}$  triggers used in Run-2 data-taking.

| Periods   | Single Muon Trigger    |
|---|------------------------|
| 2015  | HLT_MU20_ILOOSE_L1MU15 |
| 2016 (A, B-D3, D4-E, F-G2, G3-I3, I4-), 2017 (B-), 2018 | HLT_MU50               |
| 2016  | HLT_MU24_ILOOSE        |
| 2015, 2016 (A)  | HLT_MU40               |
| 2016 (B-D3, D4-E)                                       | HLT_MU24_IVARMEDIUM    |
| 2016 (D4-E, F-G2, G3-I3, I4-), 2017 (B-), 2018          | HLT_MU26_IVARMEDIUM    |

Table 5.6: ATLAS single muon triggers used in Run-2 data-taking.

(a)  $t\bar{t}$  CR Merged(b)  $t\bar{t}$  CR ResolvedFigure 5.6:  $E_T^{miss}$  trigger efficiency in the electron channel.(a)  $t\bar{t}$  CR Merged(b)  $t\bar{t}$  CR ResolvedFigure 5.7:  $E_T^{miss}$  trigger efficiency in the muon channel.(a)  $t\bar{t}$  CR Merged(b)  $t\bar{t}$  CR ResolvedFigure 5.8:  $E_T^{miss}$  trigger efficiency plotted against  $E_T^{miss}$  with muons invisible.

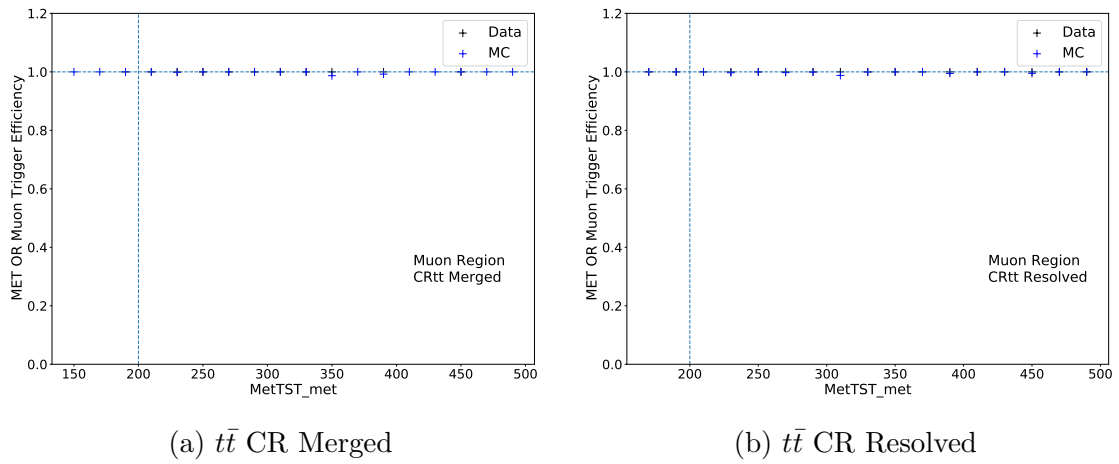


Figure 5.9: Efficiency of the logical OR of  $E_T^{miss}$  and single muon triggers.

# Chapter 6

## Results

In order to provide a comprehensive prediction of the sensitivity of the analysis regions to the signal model, we perform an exclusion fit. In an exclusion fit, the signal model plus the standard model background is taken as the null hypothesis and is tested against the alternative hypothesis of the background-only model. The following chapter describes the process of the fit, and provides the results.

### 6.1 Statistical Framework

In each signal region, we create a histogram of all events in the sample that are within the region, consisting of several bins of a chosen variable  $x$ . The expectation value for the number of events in a single bin is:

$$E[n_i] = \mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}), \quad (6.1)$$

where  $\mu$  is the parameter of interest, the signal strength, which is 0 in the background-only case and 1 in the nominal signal model. The vector of variables  $\boldsymbol{\theta}$  represents the set of nuisance parameters, and  $s_i$  and  $b_i$  are the expected number of signal and background events in bin  $i$ :

$$\begin{aligned} s_i &= s_{tot} \int_{\text{bin } i} f_s(x|\boldsymbol{\theta}) dx \\ b_i &= b_{tot} \int_{\text{bin } i} f_b(x|\boldsymbol{\theta}) dx. \end{aligned} \quad (6.2)$$

Here,  $f_s(x|\boldsymbol{\theta})$  and  $f_b(x|\boldsymbol{\theta})$  are the probability density functions of the variable  $x$  for signal and background events. The variables  $s_{tot}$  and  $b_{tot}$  contain the total mean numbers of signal and background events, where  $b_{tot}$  is allowed to vary and  $s_{tot}$  is fixed to the value predicted by the MC signal samples. In each control region, we build a similar histogram with a single bin with the expected value:

$$E[m_i] = b_i(\boldsymbol{\theta}). \quad (6.3)$$

We can then construct a likelihood function from the product of Poisson probabilities for each bin:

$$L(\mu, \boldsymbol{\theta}) = C_{sys}(\boldsymbol{\theta}^0, \boldsymbol{\theta}) \prod_{j=1}^N \frac{E[n_j]}{n_j!} e^{-E[n_j]} \prod_{k=1}^M \frac{E[m_k]}{m_k!} e^{-E[m_k]}, \quad (6.4)$$

where  $C_{sys}(\boldsymbol{\theta}^0, \boldsymbol{\theta})$  is a product of probability distributions for the measurements describing each systematic uncertainty. These are taken to be Gaussian distributed, and are each scaled to have unit standard deviations, giving:

$$C_{sys}(\boldsymbol{\theta}^0, \boldsymbol{\theta}) = \prod_{m \in S} G(\boldsymbol{\theta}_m^0 - \boldsymbol{\theta}_m, \sigma = 1). \quad (6.5)$$

We then test a hypothesized value of the signal strength  $\mu$  using the profile log-likelihood ratio as the test statistic:

$$t_\mu = -2 \log \left( \frac{L(\mu_{sig}, \hat{\boldsymbol{\theta}})}{L(\hat{\mu}_{sig}, \hat{\boldsymbol{\theta}})} \right), \quad (6.6)$$

where  $\hat{\boldsymbol{\theta}}$  is the set of nuisance parameters which maximize the likelihood function for the chosen signal strength  $\mu_{sig}$ , and  $\hat{\boldsymbol{\theta}}$  and  $\hat{\mu}_{sig}$  are the unconditional maximum-likelihood estimators for  $L$ . From this we compute the  $p$ -value corresponding to a given signal strength  $\mu_{sig}$  by integrating the probability distribution of  $t_\mu$ :

$$p_\mu = \int_{t_{\mu_{obs}}}^{\infty} f(t_\mu | \mu_{sig}, \boldsymbol{\theta}) dt_\mu. \quad (6.7)$$

This probability distribution can be estimated by generating toy MC experiments, however in the ‘‘asymptotic’’ regime with sufficiently high statistics (usually at least  $\mathcal{O}(5)$  expected data events) this distribution is known to follow a  $\chi^2$  distribution

according to Wilks' theorem [53]. Using this  $p$ -value formula, we follow the  $\text{CL}_s$  technique to present results [54]. We calculate  $\text{CL}_{s+b}$  as the  $p$ -value for the existence of the signal model with some  $\mu_{\text{sig}} > 0$  and  $1 - \text{CL}_b$  as the  $p$ -value for the background-only case  $\mu_{\text{sig}} = 0$ :

$$\begin{aligned}\text{CL}_{s+b}(\mu_{\text{sig}}) &= \int_{t_{\mu_{\text{obs}}}}^{\infty} f(t_{\mu} | \mu_{\text{sig}}, \boldsymbol{\theta}) dt_{\mu} \\ 1 - \text{CL}_b &= \int_{t_{\mu_{\text{obs}}}}^{\infty} f(t_{\mu} | 0, \boldsymbol{\theta}) dt_{\mu}.\end{aligned}\tag{6.8}$$

$\text{CL}_s$  is then given by the ratio of these two quantities:

$$\text{CL}_s(\mu_{\text{sig}}) = \frac{\text{CL}_{s+b}(\mu_{\text{sig}})}{1 - \text{CL}_b}.\tag{6.9}$$

To perform an exclusion fit, we evaluate  $\text{CL}_s$  for the nominal signal strength  $\mu_{\text{sig}} = 1$ , and exclude signal models with  $\text{CL}_s(1) > 0.05$ . We perform this statistical analysis of the finalized analysis regions using the HistFitter software package [55].

## 6.2 Minimum $m_s$

As a discrimination variable for the exclusion fit, we use a partial reconstruction of the  $s$  mass. The  $s$  can not be fully reconstructed because the  $E_T^{\text{miss}}$  derived from the neutrino cannot be distinguished from the  $E_T^{\text{miss}}$  derived from dark matter. Instead, we calculate the minimum possible  $s$  mass  $m_s^{\text{min}}$  consistent with the other measured  $s$  decay products. This variable shows good discrimination power, but is not used in the signal region selection because the distribution shape varies widely across signal points according to their  $s$  mass.

To calculate  $m_s^{\text{min}}$ , we begin by using the  $W$  mass constraint on the charged lepton-neutrino system to solve for the neutrino energy  $E_{\nu}$  as a function of the angle  $\theta_{\ell\nu}$  between the charged lepton and neutrino:

$$\begin{aligned}E_{\nu} &= \frac{m_W^2}{2E_{\ell}(1 - \cos \theta_{\ell\nu})}, \text{ where} \\ m_W^2 &= (p_{\ell} + p_{\nu})^2 = 2p_{\ell}p_{\nu} = 2E_{\ell}E_{\nu}(1 - \cos \theta_{\ell\nu}).\end{aligned}\tag{6.10}$$

We then rotate the coordinate system to place the neutrino 3-momentum along the  $z$ -axis and the hadronic  $W$  ( $W_H$ ) 3-momentum in the  $x$ - $z$  plane. The neutrino 4-

momentum is then:

$$p_\nu = \frac{m_W^2}{2E_l(1 - \cos \theta_{l\nu})} (\sin \theta_{l\nu}, \sin \theta_{l\nu} \sin \phi_\nu, \cos \theta_{l\nu}, 1). \quad (6.11)$$

We can then write the  $s$  mass as:

$$\begin{aligned} m_s^2 &= (p_{W_H} + p_l + p_\nu)^2 \\ m_s^2 &= (E_{W_H} + E_l + E_\nu)^2 - (p_{W_{Hx}} + E_\nu \sin \theta_{l\nu} \cos \phi_\nu)^2 - (E_\nu \sin \theta_{l\nu} \sin \phi_\nu)^2 \\ &\quad - (E_l + p_{W_{Hz}} + E_\nu \cos \theta_{l\nu})^2. \end{aligned} \quad (6.12)$$

It is clear from this equation that  $m_s$  is minimized with  $\phi_\nu = 0$ , and substituting this and Equation 6.10 for  $E_\nu$  we are left with an expression for  $m_s$  with  $\theta_{l\nu}$  as the only remaining unknown variable:

$$\begin{aligned} m_s^2 &= \left( E_l + \frac{m_W^2}{2E_l(1 - \cos \theta_{l\nu})} + E_{W_H} \right)^2 - \left( |p_{\vec{W}_H}| \sin \theta_{W_l} + \frac{\sqrt{1 - \cos^2 \theta_{l\nu}} m_W^2}{2E_l(1 - \cos \theta_{l\nu})} \right)^2 \\ &\quad - \left( E_l + |p_{\vec{W}_H}| \cos \theta_{W_l} + \frac{\cos \theta_{l\nu} m_W^2}{2E_l(1 - \cos \theta_{l\nu})} \right)^2. \end{aligned} \quad (6.13)$$

We then vary  $\theta_{l\nu}$  on the interval  $[0, \pi]$  to search for the minimum  $m_s$ . Figure 6.1 shows the signal and background distributions of  $m_s^{\min}$  in both signal regions, demonstrating the strong discrimination potential and large variance between signal samples that make it an ideal model fitting variable. During model fitting we place events in 4 bins in the “merged” region and 5 bins in the “resolved” region in the range  $125 \leq m_s \leq 375$  GeV. The bins in the “merged” region have edges at 125, 165, 190, 225, and 375 GeV, and were motivated by the requirement of at least 5 expected background events in each bin to ensure the validity of the asymptotic approximation of Wilks’ theorem. In the “resolved” region, the bins are equidistant for simplicity.

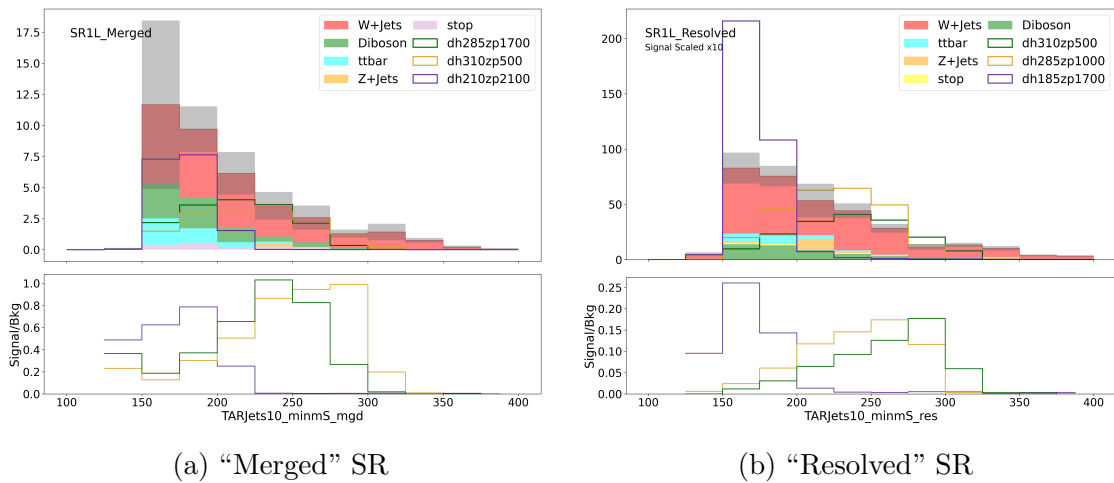


Figure 6.1: Distributions of  $m_s^{\min}$  in the "merged" and "resolved" signal regions. Grey bands represent MC statistical uncertainty on each bin.

## 6.3 Systematic Uncertainties

Statistical fluctuations are uncorrelated between measurements that use non-overlapping datasets, and arise from a measurement consisting of a limited number of observations. Systematic uncertainties, however, are often correlated in measurements that do not necessarily use overlapping datasets, generally do not scale with the sample size, and may derive from the nature of the experiment or uncertainty in the model used to make conclusions about the data. For this analysis, we further categorize systematic uncertainties into theory uncertainties on the modelling and experimental uncertainties. The evaluation of theory uncertainties is not within the scope of this thesis, and we assign them a flat value of 20% of the yield on each background category. We chose a value of 20% based on the magnitude of the SM modelling uncertainties observed in the fully-hadronic channel in [26] as well as ongoing work internal to the analysis. What follows is a short description of the various types of experimental systematic uncertainties that we consider in this thesis.

### 6.3.1 $R = 0.4$ Jets

Uncertainties on  $R = 0.4$  jets are divided into jet energy scale (JES) and jet energy resolution (JER) uncertainties. These include, but are not limited to, uncertainties on pileup, flavour composition, and punch-through of jets.

### 6.3.2 Track Uncertainties

These are uncertainties on the reconstruction of tracks in the ATLAS inner detector. In this analysis they are primarily propagated into the reconstruction of TAR jets in the “merged” signal and control regions.

### 6.3.3 Muon and Electron

For muons and electrons, we consider uncertainties on the reconstruction efficiency, isolation and identification, and energy or momentum scale and resolution.

### 6.3.4 $E_T^{miss}$

We propagate uncertainties on the aforementioned objects into the  $E_T^{miss}$  calculation, however we consider separate  $E_T^{miss}$  systematics which affect the reconstruction of the

$E_T^{miss}$  soft term.

### 6.3.5 Luminosity

The integrated luminosity measured across the run-2 data taking period is known to a precision of 1.7% [56]. This is therefore applied as an overall systematic across the normalization of all MC events.

## 6.4 Fit Results

### 6.4.1 Background Only Fit

We first perform a background-only fit to examine the effects of systematic uncertainties and control region normalization on background predictions in the signal regions. The signal regions for this analysis remain blinded; therefore ATLAS data in the signal regions is mimicked by fake data with a yield set equal to the nearest integer value to the pre-fit total background yield in each bin.

Figures 6.2 and 6.3 show comparisons of the pre-fit and post-fit yields in each region for the background only fit, while Tables 6.1 and 6.2 give a more detailed numerical breakdown. In the “merged” signal region, the fit increases the expected number of SM events primarily due to an increase in expected  $W$ +jets events from the normalization derived in the “merged”  $W$ +jets control region. In the “resolved” signal region, the fit decreases the expected background yield, again primarily due to normalization this time from the “resolved”  $W$ +jets control region. In both cases the uncertainty on the predicted  $W$ +jets yield is reduced. In both the “merged” and “resolved” categories the  $t\bar{t}$  control regions slightly reduce the predicted  $t\bar{t}$  yield and substantially reduce the uncertainty on those predictions.

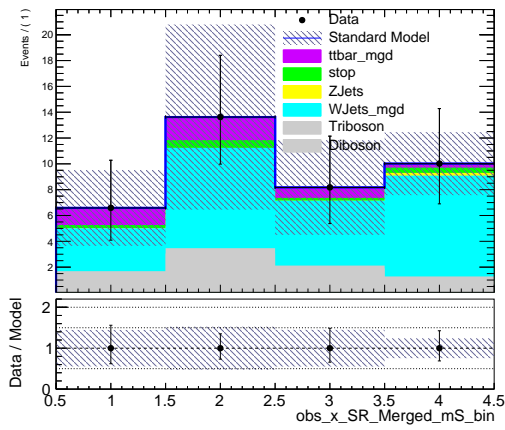
Tables 6.3 and 6.4 summarize the uncertainty on the total yield in each analysis region from statistical and systematic sources. Uncertainties preceded by “ $\mu$ ”, “ $\alpha$ ”, and “ $\gamma$ ” represent normalization, systematic, and statistical uncertainties respectively. The leading sources of uncertainty in both signal regions are normalization factors and theory uncertainties on the  $W$ +jets background, followed by a mix of JER experimental systematics and bin-by-bin statistical uncertainties.

| <b>table.results.yields channel</b> | SR_Merged        | CRW_Merged         | CRTT_Merged            |
|-------------------------------------|------------------|--------------------|------------------------|
| Observed events                     | 38               | 205                | 59                     |
| Fitted bkg events                   | $33.80 \pm 7.94$ | $204.88 \pm 14.28$ | $59.00 \pm 7.67$       |
| Fitted Diboson events               | $7.40 \pm 2.22$  | $26.49 \pm 5.96$   | $0.22 \pm 0.06$        |
| Fitted WJets events                 | $20.52 \pm 5.74$ | $150.71 \pm 16.85$ | $2.21^{+3.58}_{-2.21}$ |
| Fitted ttbar events                 | $3.05 \pm 0.92$  | $6.99 \pm 1.59$    | $51.02 \pm 8.46$       |
| Fitted ZJets events                 | $0.22 \pm 0.06$  | $15.74 \pm 4.13$   | $0.00 \pm 0.00$        |
| Fitted stop events                  | $1.46 \pm 0.47$  | $3.10 \pm 0.77$    | $5.55 \pm 1.28$        |
| Fitted Triboson events              | $1.15 \pm 0.23$  | $1.85 \pm 0.20$    | $0.00 \pm 0.00$        |
| MC exp. SM events                   | $38.39 \pm 9.45$ | $232.60 \pm 43.80$ | $78.98 \pm 15.07$      |
| MC exp. Diboson events              | $7.40 \pm 2.22$  | $26.49 \pm 6.00$   | $0.22 \pm 0.06$        |
| MC exp. WJets events                | $23.93 \pm 6.94$ | $175.75 \pm 40.10$ | $2.58^{+4.09}_{-2.58}$ |
| MC exp. ttbar events                | $4.22 \pm 1.34$  | $9.68 \pm 2.36$    | $70.63 \pm 14.45$      |
| MC exp. ZJets events                | $0.22 \pm 0.06$  | $15.74 \pm 4.15$   | $0.00 \pm 0.00$        |
| MC exp. stop events                 | $1.46 \pm 0.47$  | $3.10 \pm 0.77$    | $5.55 \pm 1.29$        |
| MC exp. Triboson events             | $1.15 \pm 0.23$  | $1.85 \pm 0.20$    | $0.00 \pm 0.00$        |

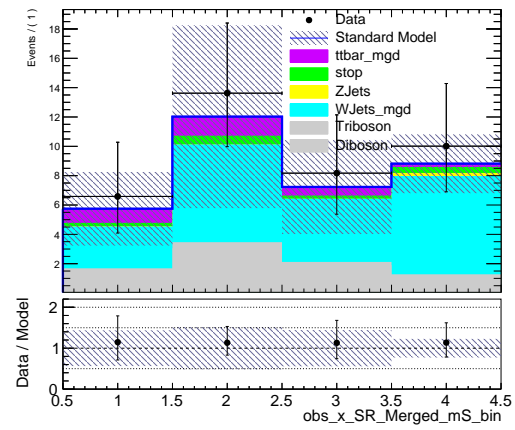
Table 6.1: Pre and post-fit yields for each background category in the merged signal and control regions for a background-only fit.

| <b>table.results.yields channel</b> | SR_Resolved        | CRW_Resolved        | CRTT_Resolved          |
|-------------------------------------|--------------------|---------------------|------------------------|
| Observed events                     | 330                | 717                 | 87                     |
| Fitted bkg events                   | $321.79 \pm 36.33$ | $716.53 \pm 26.74$  | $86.99 \pm 9.30$       |
| Fitted Diboson events               | $50.03 \pm 10.88$  | $105.76 \pm 21.93$  | $0.58 \pm 0.14$        |
| Fitted WJets events                 | $226.84 \pm 33.14$ | $560.43 \pm 35.98$  | $2.27^{+3.00}_{-2.27}$ |
| Fitted ttbar events                 | $24.05 \pm 4.79$   | $22.64 \pm 4.01$    | $73.24 \pm 10.11$      |
| Fitted ZJets events                 | $14.32 \pm 3.38$   | $18.85 \pm 5.33$    | $0.00 \pm 0.00$        |
| Fitted stop events                  | $3.74 \pm 0.95$    | $3.26 \pm 0.77$     | $10.89 \pm 2.58$       |
| Fitted Triboson events              | $2.81 \pm 0.24$    | $5.59 \pm 0.27$     | $0.01^{+0.01}_{-0.01}$ |
| MC exp. SM events                   | $330.48 \pm 60.09$ | $742.55 \pm 124.74$ | $78.06 \pm 13.93$      |
| MC exp. Diboson events              | $50.04 \pm 10.95$  | $105.77 \pm 22.09$  | $0.58 \pm 0.14$        |
| MC exp. WJets events                | $238.49 \pm 56.08$ | $589.23 \pm 120.85$ | $2.38^{+3.20}_{-2.38}$ |
| MC exp. ttbar events                | $21.08 \pm 4.93$   | $19.84 \pm 4.29$    | $64.21 \pm 13.17$      |
| MC exp. ZJets events                | $14.32 \pm 3.40$   | $18.85 \pm 5.36$    | $0.00 \pm 0.00$        |
| MC exp. stop events                 | $3.74 \pm 0.96$    | $3.26 \pm 0.77$     | $10.89 \pm 2.59$       |
| MC exp. Triboson events             | $2.81 \pm 0.24$    | $5.59 \pm 0.28$     | $0.01^{+0.01}_{-0.01}$ |

Table 6.2: Pre and post-fit yields for each background category in the resolved signal and control regions for a background-only fit.



(a) Merged SR pre-fit



(b) Merged SR post-fit

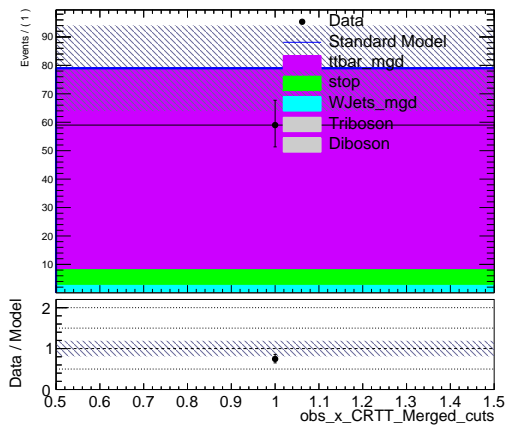
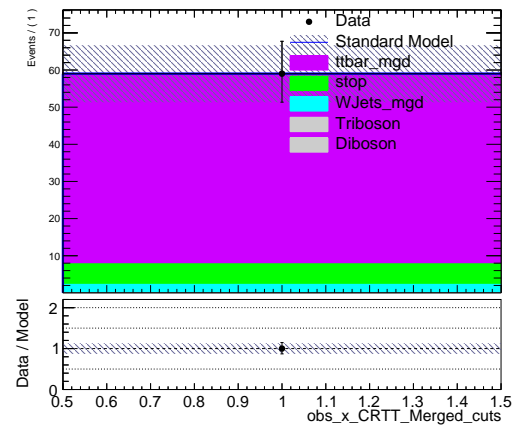
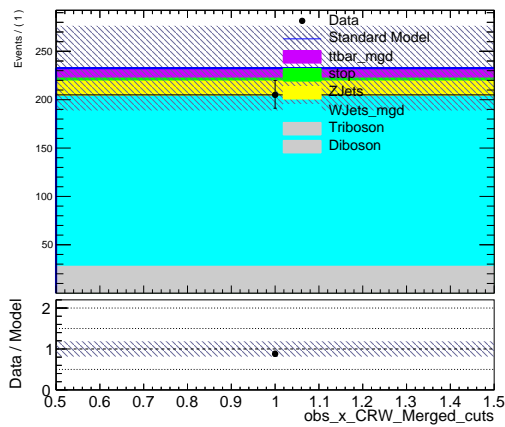
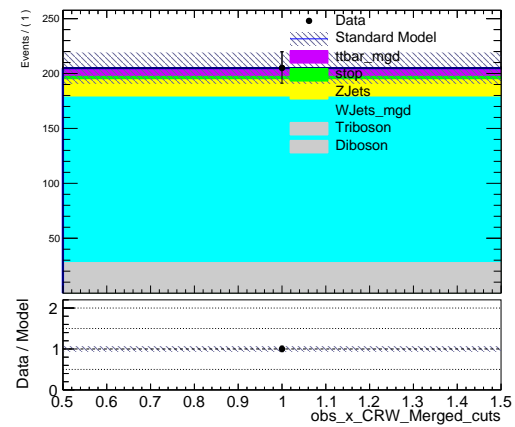
(c) Merged CR  $t\bar{t}$  pre-fit(d) Merged CR  $t\bar{t}$  post-fit(e) Merged CR  $W$ +jets pre-fit(f) Merged CR  $W$ +jets post-fit

Figure 6.2: Pre and post-fit yields in the merged signal and control regions for a background-only fit.

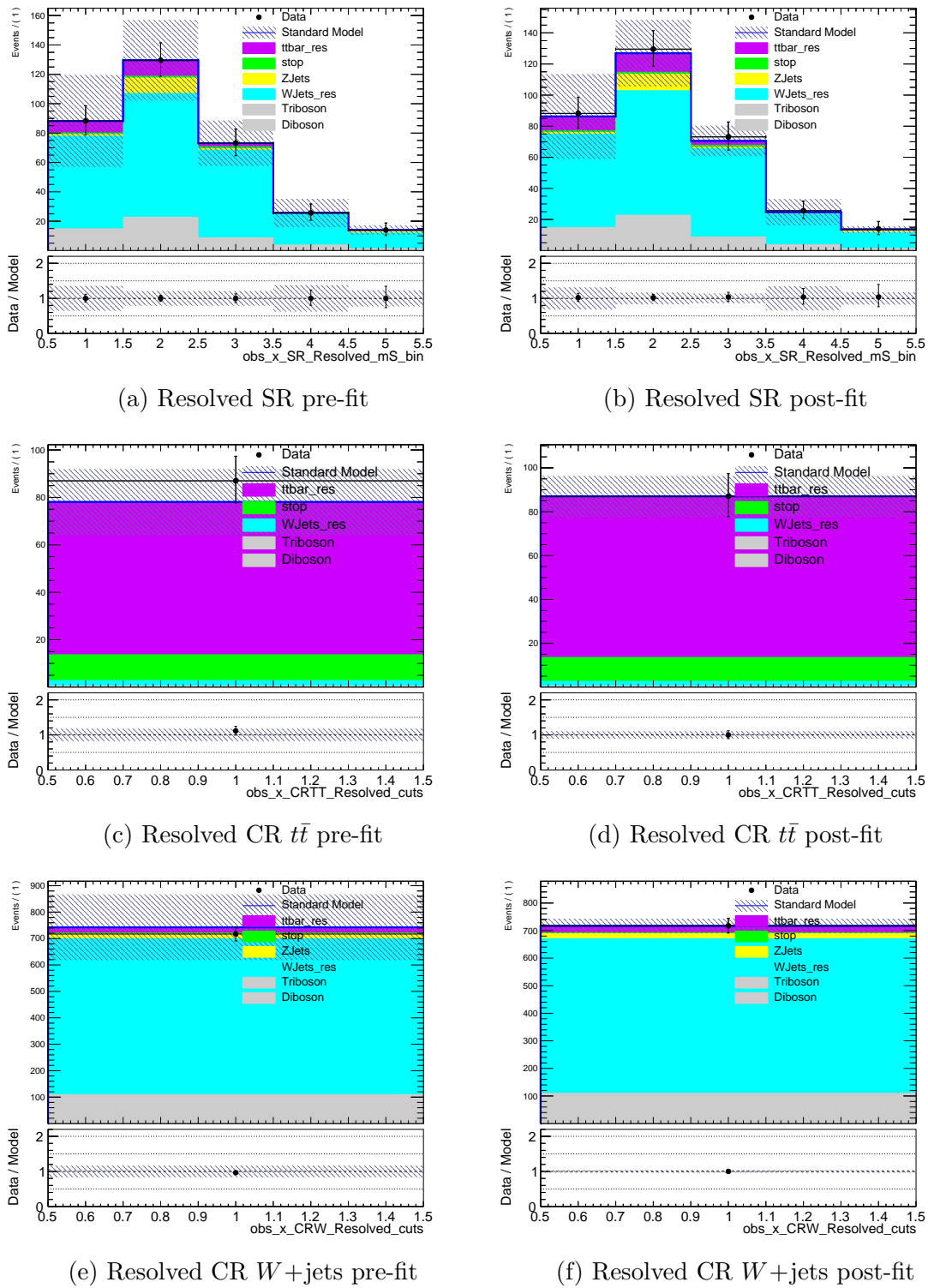


Figure 6.3: Pre and post-fit yields in the resolved signal and control regions for a background-only fit.

| <b>Uncertainty of channel</b>                  | SR_Merged           | CRW_Merged          | CRTT_Merged         |
|--|---------------------|---------------------|---------------------|
| Total background expectation                   | 33.80               | 204.88              | 59.00               |
| Total statistical ( $\sqrt{N_{\text{exp}}}$ )  | $\pm 5.81$          | $\pm 14.31$         | $\pm 7.68$          |
| Total background systematic                    | $\pm 7.94$ [23.48%] | $\pm 14.28$ [6.97%] | $\pm 7.67$ [13.00%] |
| $\gamma_{\text{stat\_SR\_Merged\_mS\_bin\_1}}$ | $\pm 5.82$          | $\pm 0.00$          | $\pm 0.00$          |
| $\mu_{\text{Wjets\_mgd}}$                      | $\pm 5.49$          | $\pm 40.31$         | $\pm 0.59$          |
| $\alpha_{\text{WJetsTheorySys}}$               | $\pm 4.00$          | $\pm 29.39$         | $\pm 0.43$          |
| $\gamma_{\text{stat\_SR\_Merged\_mS\_bin\_2}}$ | $\pm 1.86$          | $\pm 0.00$          | $\pm 0.00$          |
| $\gamma_{\text{stat\_SR\_Merged\_mS\_bin\_0}}$ | $\pm 1.59$          | $\pm 0.00$          | $\pm 0.00$          |
| $\alpha_{\text{DibosonTheorySys}}$             | $\pm 1.47$          | $\pm 5.26$          | $\pm 0.04$          |
| $\gamma_{\text{stat\_SR\_Merged\_mS\_bin\_3}}$ | $\pm 1.45$          | $\pm 0.00$          | $\pm 0.00$          |
| $\alpha_{\text{JET\_Pileup\_RhoTopology}}$     | $\pm 1.27$          | $\pm 1.64$          | $\pm 0.79$          |
| $\alpha_{\text{JET\_EtaIntercalibration}}$     | $\pm 1.05$          | $\pm 0.44$          | $\pm 0.12$          |
| $\alpha_{\text{JET\_Pileup\_OffsetMu}}$        | $\pm 0.92$          | $\pm 0.47$          | $\pm 0.02$          |
| $\mu_{\text{ttbar\_mgd}}$                      | $\pm 0.81$          | $\pm 1.85$          | $\pm 13.52$         |
| $\alpha_{\text{MET\_SoftTrk\_ResoPerp}}$       | $\pm 0.75$          | $\pm 0.19$          | $\pm 0.14$          |
| $\alpha_{\text{JET\_Flavor\_Composition}}$     | $\pm 0.75$          | $\pm 2.00$          | $\pm 1.06$          |
| $\alpha_{\text{JET\_Pileup\_OffsetNPV}}$       | $\pm 0.74$          | $\pm 1.62$          | $\pm 1.06$          |
| $\alpha_{\text{JET\_EffectiveNP\_Modelling1}}$ | $\pm 0.66$          | $\pm 1.50$          | $\pm 0.93$          |
| $\alpha_{\text{ttbarTheorySys}}$               | $\pm 0.61$          | $\pm 1.39$          | $\pm 10.14$         |
| $\alpha_{\text{lumiSys}}$                      | $\pm 0.57$          | $\pm 3.46$          | $\pm 1.00$          |
| $\alpha_{\text{TRK\_FAKE\_RATE\_LR}}$          | $\pm 0.47$          | $\pm 0.00$          | $\pm 0.05$          |
| $\alpha_{\text{MET\_SoftTrk\_ResoPara}}$       | $\pm 0.43$          | $\pm 0.21$          | $\pm 0.79$          |
| $\alpha_{\text{JET\_EffectiveNP\_Detector1}}$  | $\pm 0.39$          | $\pm 0.52$          | $\pm 0.00$          |
| $\alpha_{\text{JET\_EffectiveNP\_Mixed3}}$     | $\pm 0.34$          | $\pm 0.00$          | $\pm 0.00$          |

Table 6.3: Effect of systematic uncertainties in the merged signal and control regions for a background-only fit. Uncertainties with a magnitude  $<1\%$  of total yield are excluded.

| <b>Uncertainty of channel</b>                    | SR_Resolved          | CRW_Resolved        | CRTT_Resolved       |
|--|----------------------|---------------------|---------------------|
| Total background expectation                     | 321.79               | 716.53              | 86.99               |
| Total statistical ( $\sqrt{N_{\text{exp}}}$ )    | $\pm 17.94$          | $\pm 26.77$         | $\pm 9.33$          |
| Total background systematic                      | $\pm 36.33$ [11.29%] | $\pm 26.74$ [3.73%] | $\pm 9.30$ [10.69%] |
| $\mu_{\text{Wjets\_res}}$                        | $\pm 48.68$          | $\pm 120.27$        | $\pm 0.49$          |
| $\alpha_{\text{WJetsTheorySys}}$                 | $\pm 44.24$          | $\pm 109.30$        | $\pm 0.44$          |
| $\gamma_{\text{stat\_SR\_Resolved\_mS\_bin\_1}}$ | $\pm 17.28$          | $\pm 0.00$          | $\pm 0.00$          |
| $\gamma_{\text{stat\_SR\_Resolved\_mS\_bin\_0}}$ | $\pm 13.76$          | $\pm 0.00$          | $\pm 0.00$          |
| $\alpha_{\text{JET\_Pileup\_PtTerm}}$            | $\pm 10.14$          | $\pm 0.75$          | $\pm 0.05$          |
| $\alpha_{\text{MET\_SoftTrk\_ResoPerp}}$         | $\pm 10.14$          | $\pm 0.15$          | $\pm 0.20$          |
| $\alpha_{\text{DibosonTheorySys}}$               | $\pm 9.93$           | $\pm 20.99$         | $\pm 0.12$          |
| $\alpha_{\text{JET\_Pileup\_OffsetNPV}}$         | $\pm 9.34$           | $\pm 0.14$          | $\pm 0.36$          |
| $\alpha_{\text{JET\_Flavor\_Composition}}$       | $\pm 8.00$           | $\pm 1.66$          | $\pm 0.42$          |
| $\gamma_{\text{stat\_SR\_Resolved\_mS\_bin\_2}}$ | $\pm 7.38$           | $\pm 0.00$          | $\pm 0.00$          |
| $\mu_{\text{ttbar\_res}}$                        | $\pm 5.99$           | $\pm 5.64$          | $\pm 18.25$         |
| $\alpha_{\text{lumiSys}}$                        | $\pm 5.43$           | $\pm 12.10$         | $\pm 1.47$          |
| $\alpha_{\text{ttbarTheorySys}}$                 | $\pm 4.78$           | $\pm 4.50$          | $\pm 14.55$         |
| $\alpha_{\text{JET\_Pileup\_RhoTopology}}$       | $\pm 4.60$           | $\pm 2.50$          | $\pm 0.85$          |
| $\alpha_{\text{JET\_EffectiveNP\_Modelling2}}$   | $\pm 4.45$           | $\pm 0.03$          | $\pm 1.13$          |
| $\alpha_{\text{JET\_EffectiveNP\_Modelling3}}$   | $\pm 4.22$           | $\pm 0.03$          | $\pm 1.21$          |
| $\alpha_{\text{JET\_EtaIntercalibration}}$       | $\pm 3.84$           | $\pm 1.09$          | $\pm 0.11$          |
| $\alpha_{\text{JET\_EffectiveNP\_Statistical4}}$ | $\pm 3.74$           | $\pm 0.03$          | $\pm 1.12$          |
| $\alpha_{\text{JET\_EffectiveNP\_Statistical3}}$ | $\pm 3.43$           | $\pm 0.03$          | $\pm 0.00$          |

Table 6.4: Effect of systematic uncertainties in the resolved signal and control regions for a background-only fit. Uncertainties with a magnitude  $<1\%$  of total yield are excluded.

### 6.4.2 Exclusion Fit

We finally perform an exclusion fit using the signal model. We evaluate the  $CL_s$  for the nominal signal strength  $\mu_{\text{sig}} = 1$  for each signal mass point in the  $m_s$ - $m_{Z'}$  plane, and interpolate between mass points using a cubic interpolation. The signal regions remain blinded, so we once again replace ATLAS data with fake data equal to the nearest integer value to the pre-fit MC expected SM yield in those regions. We then build a contour along  $CL_s = 0.05$ , and the area outside this contour ( $CL_s > 0.05$ ) would be excluded at a 95% confidence level with the fake data. This provides the strongest estimate of the designed regions' sensitivity to the dark Higgs signal model prior to unblinding.

We first perform an exclusion fit on the “merged” analysis channel alone to assess the sensitivity of the “merged” signal region selection criteria we designed. The expected exclusion contour for the “merged” channel is shown in Figure 6.4. We then perform an exclusion fit of the “resolved” channel alone, with the expected exclusion contour shown in Figure 6.5 to see how it complements the “merged” channel. Finally, in Figure 6.6 we show the expected exclusion contour for the combined “merged” and “resolved” channels, demonstrating the full expected exclusion power of the analysis. Figure 6.7 shows the expected and observed exclusion limits obtained in the fully hadronic  $s \rightarrow WW$  decay channel for comparison.

The expected exclusion contours show that the merged region dominates sensitivity. On its own, the resolved region has very little exclusion power, but it helps to complement and strengthen the results from the merged region. As a result, we expect to substantially improve on the limits set in the fully hadronic decay channel in our analysis.

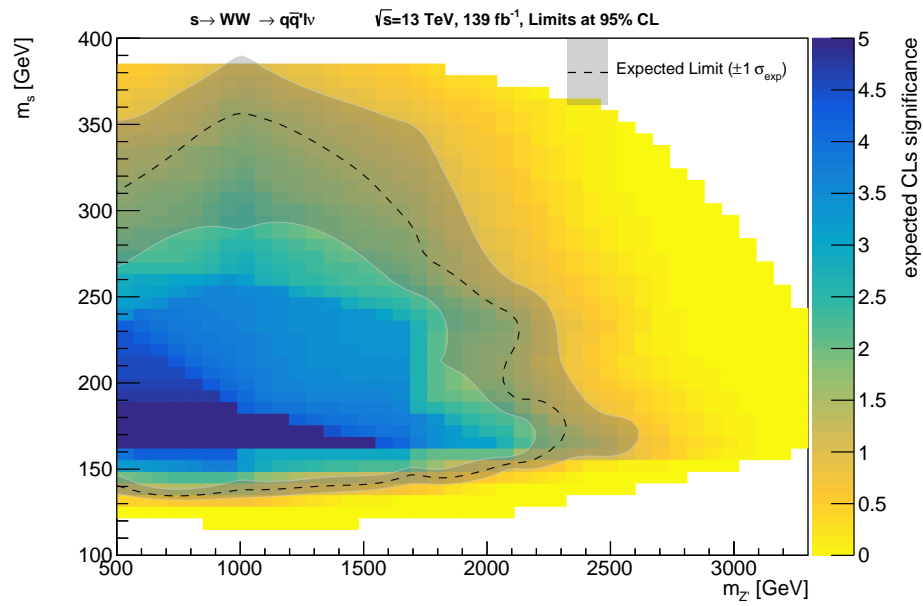


Figure 6.4: Expected exclusion limits for the dark Higgs signal model obtained using the “merged” signal and control regions.

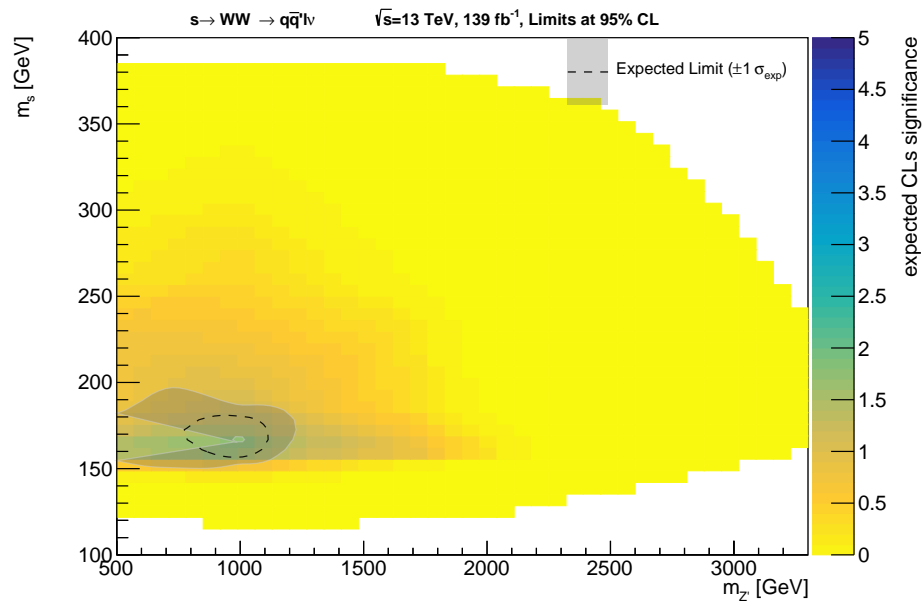


Figure 6.5: Expected exclusion limits for the dark Higgs signal model obtained using the “resolved” signal and control regions.

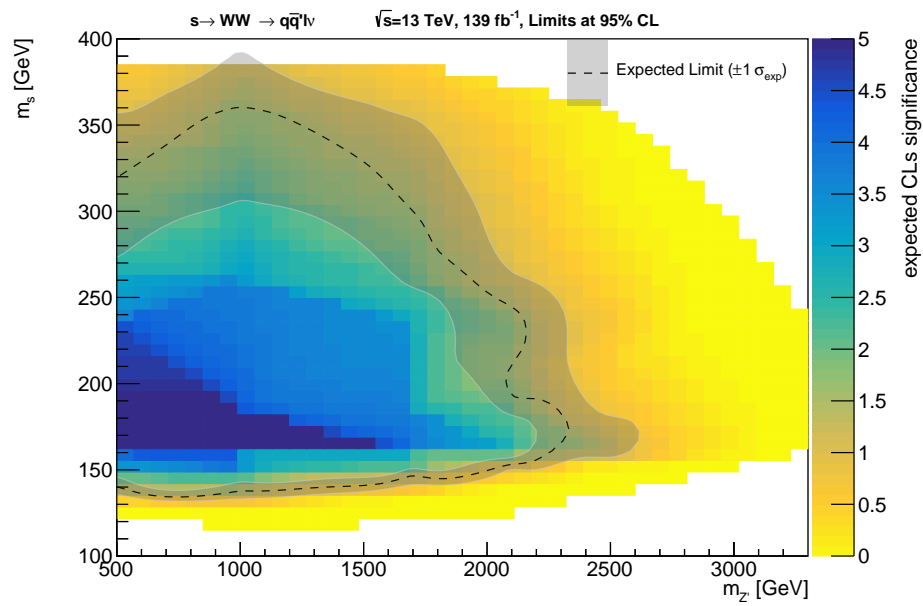


Figure 6.6: Expected exclusion limits for the dark Higgs signal model obtained using the “merged” and “resolved” signal and control regions.

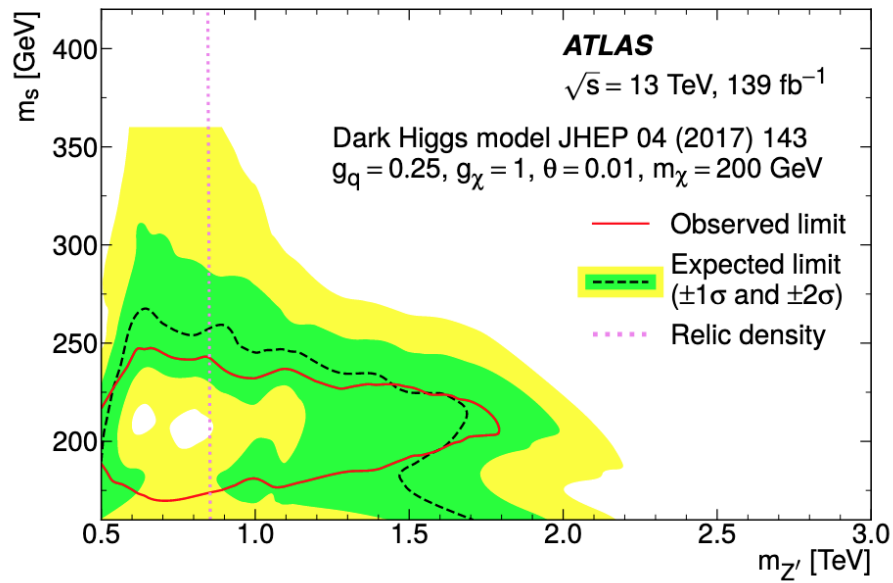


Figure 6.7: Expected and observed exclusion limits for the dark Higgs signal model from the fully hadronic  $s \rightarrow WW$  channel from [26].

# Chapter 7

## Conclusions

This thesis has presented the design of the most sensitive “merged” signal region in the search for dark matter produced with a new heavy vector boson,  $Z'$ , and a scalar “dark Higgs” boson,  $s$ , decaying to a pair of  $W$  bosons with a semileptonic ( $q\bar{q}\ell\nu$ ) final state. It has also presented the design of control regions to constrain the  $t\bar{t}$  SM background, as well as the important supporting analysis work. The search uses data collected by the ATLAS experiment in 2015-2018, from proton-proton collisions at a center-of-mass energy of 13 TeV.

The search for the dark Higgs model [5] is motivated by the desire to detect and characterize dark matter, as well as the requirement of a mechanism to generate the masses of particles in the dark sector. The SM is an extremely successful theory of the fundamental particles and their interactions, however astrophysical observations demand the existence of dark matter beyond the current confines of the model. The dark Higgs model introduces dark matter as a Majorana fermion, with a mass generated by the Higgs mechanism from a new Higgs field with an associated dark Higgs boson  $s$ . Through the mixing of the SM Higgs field and the dark Higgs field, the  $s$  can interact with all massive SM particles and can thus decay to SM final states that are observable in the ATLAS detector.

To search for this signature, we have analyzed proton-proton collision events measured in the ATLAS detector. We have begun by defining a set of physics objects representing particles and groups of particles in the ATLAS detector. We have then used those definitions to reconstruct and characterize events in the ATLAS detector. Using the reconstructed events we have defined a signal region consisting of an optimized set of event-by-event criteria that, based on MC simulated data, are expected to have a high efficiency for selecting events containing signal (mono- $s$  decay)

processes, and to select few SM background events.

We have subsequently defined the control regions as regions near to, but not overlapping with, the signal regions, which are statistically dominated by the  $t\bar{t}$  background process. We compared the MC simulated SM background to the measured ATLAS data in these regions to gain information about the accuracy of our MC predictions and constrain our predictions in the signal region. Based on the SM and signal predictions given by MC in the signal region, and the constraints derived in the control regions, we have finally calculated and presented the expected sensitivity of the signal regions to the signal model, which represents our ability to make conclusions about its existence using observations in those regions.

Using the designed regions described in this thesis, the search for the semileptonic  $s$  decay will be able to substantially improve upon the sensitivity to the dark Higgs model achieved by prior searches in other decay channels. As shown in Figure 6.6 in the previous chapter, in this search we will be able to exclude  $s$  masses from approximately 150 GeV up to a range between approximately 290 to 360 GeV, with  $Z'$  masses up to a range between approximately 1200 to 2200 GeV, using the Run-2 ATLAS  $140 \text{ fb}^{-1}$  dataset at the proton-proton center-of-mass energy of 13 TeV.

# Appendix A

## Additional Information

### A.1 Preselection level plots

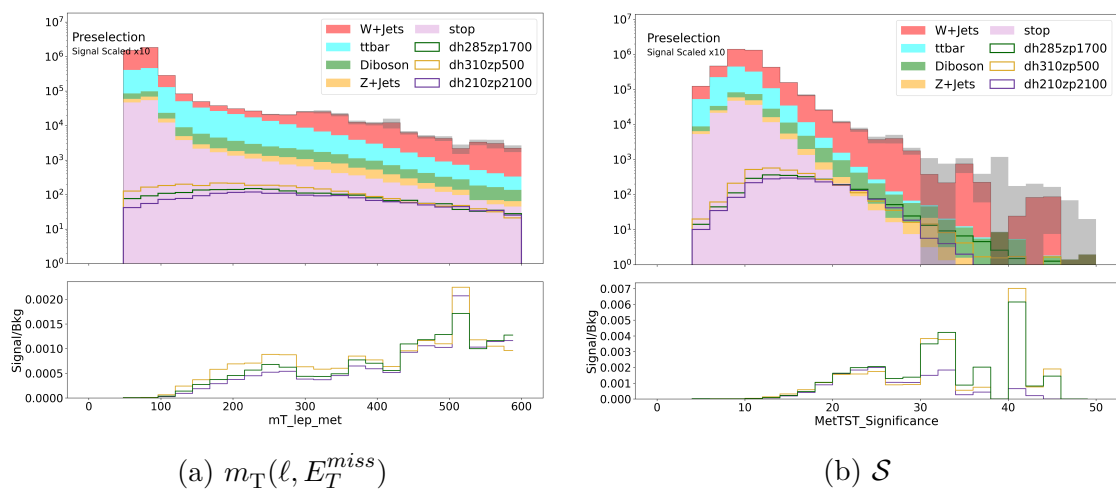


Figure A.1: Preselection level distributions. Grey bands represent MC statistical uncertainty on each bin.

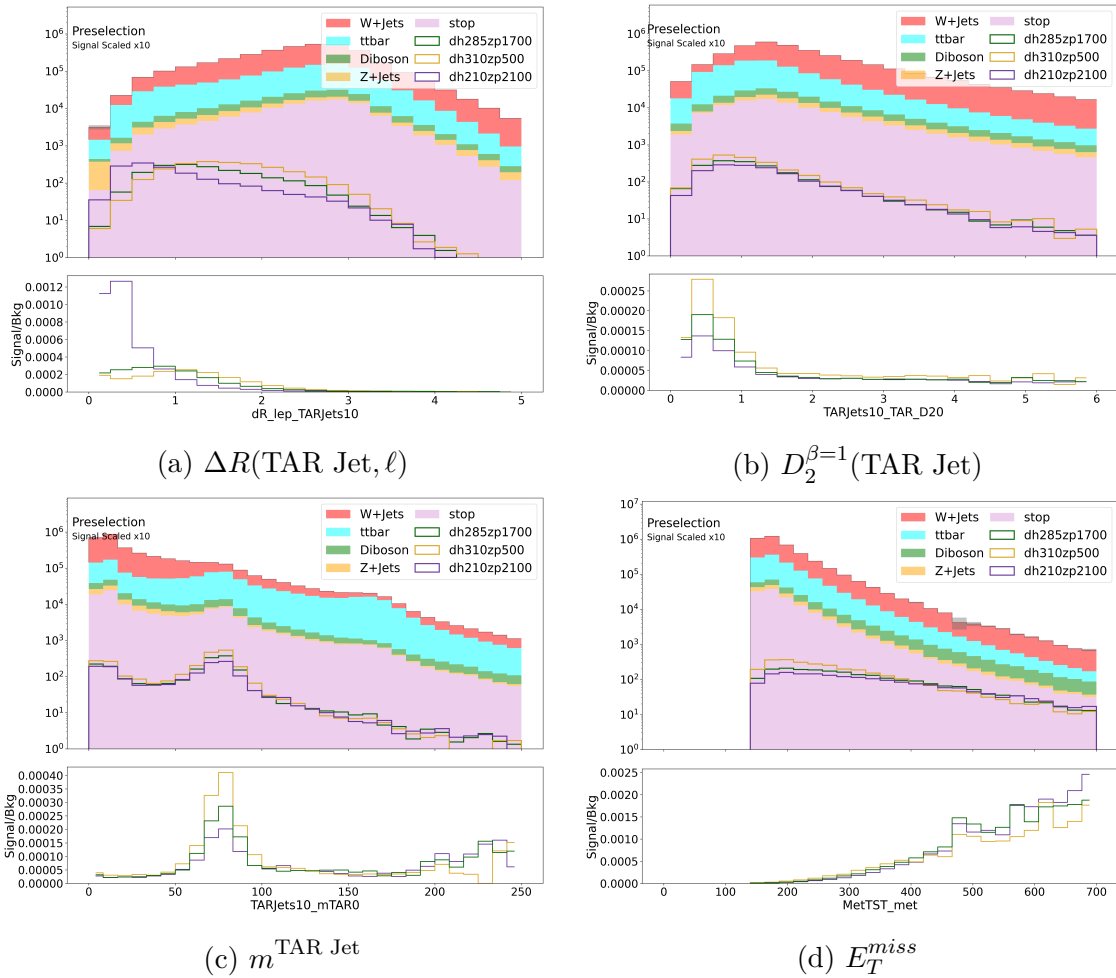


Figure A.2: Preselection level distributions. Grey bands represent MC statistical uncertainty on each bin.

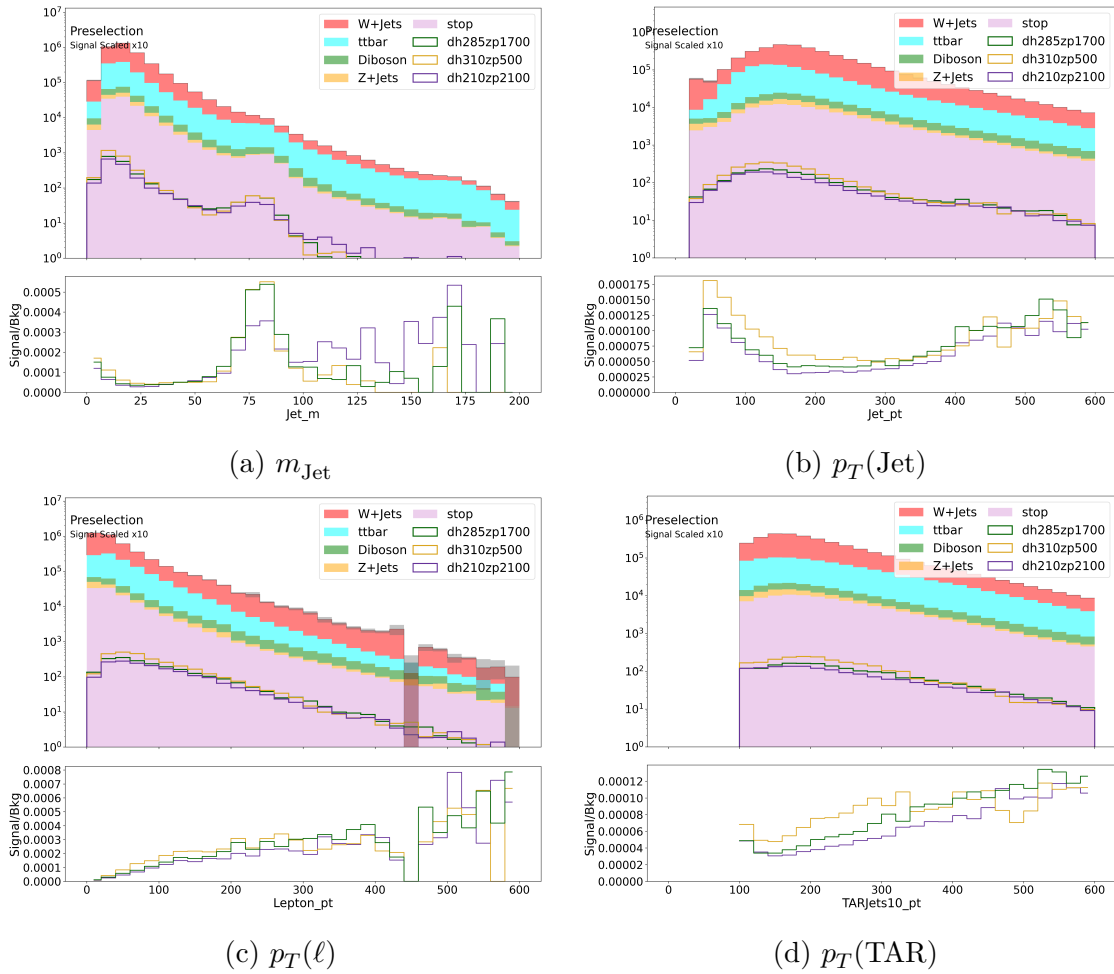


Figure A.3: Preselection level distributions. Grey bands represent MC statistical uncertainty on each bin.

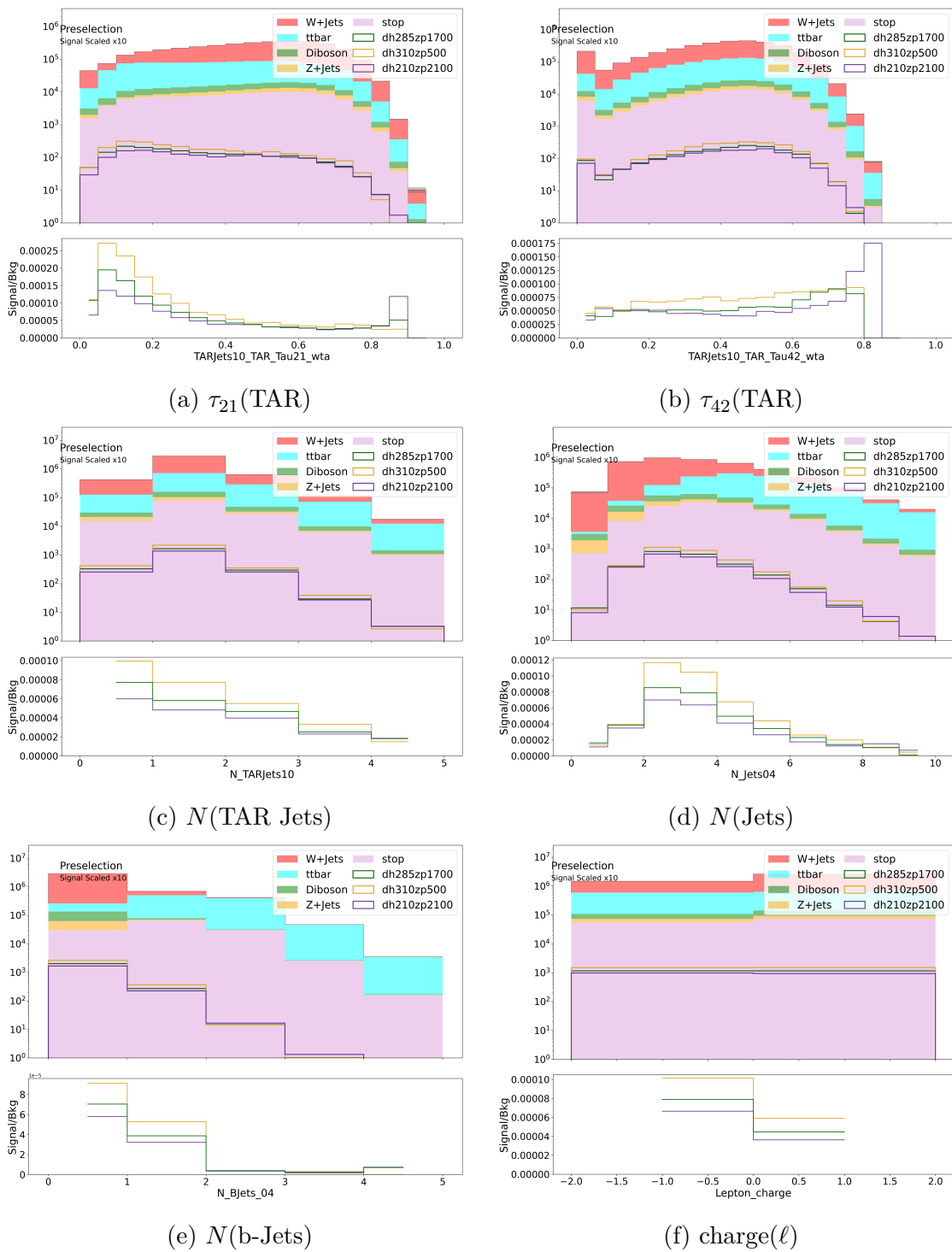


Figure A.4: Preselection level distributions. Grey bands represent MC statistical uncertainty on each bin.

## A.2 Data-MC Comparisons in Control Regions

The following section contains plots comparing data and MC distributions in the control regions. MC events are normalized to the luminosity measured in data as described in Chapter 4 subsection 4.2.3. The uncertainties shown on both data and MC events in these plots are purely statistical and do not include uncertainties on the modelling of background events, which impact the cross-section of the different categories of background events. Some statistically significant differences between data and MC are therefore not unexpected.

### A.2.1 $t\bar{t}$ Control Regions

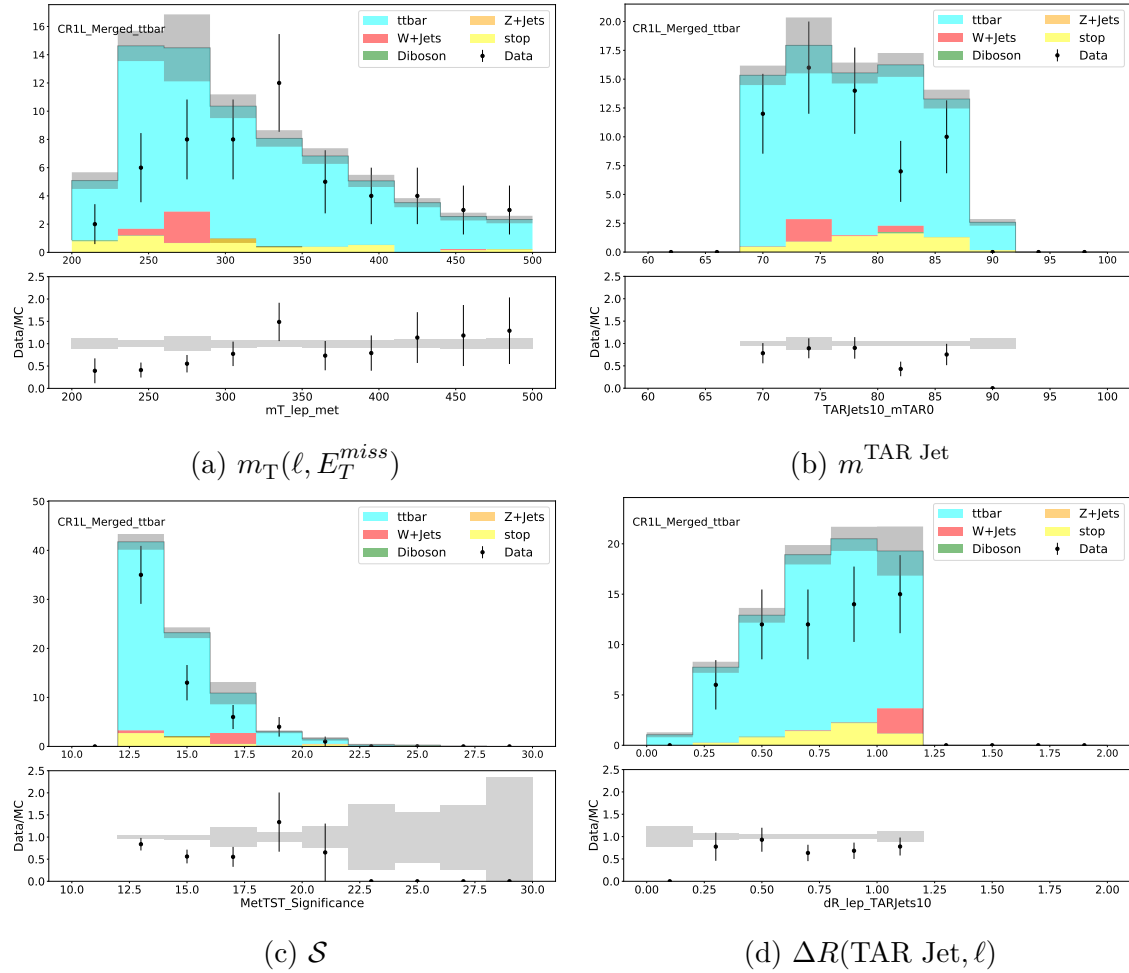


Figure A.5: Data-MC Comparisons in the “merged”  $t\bar{t}$  control region. Grey bands represent MC statistical uncertainty on each bin.

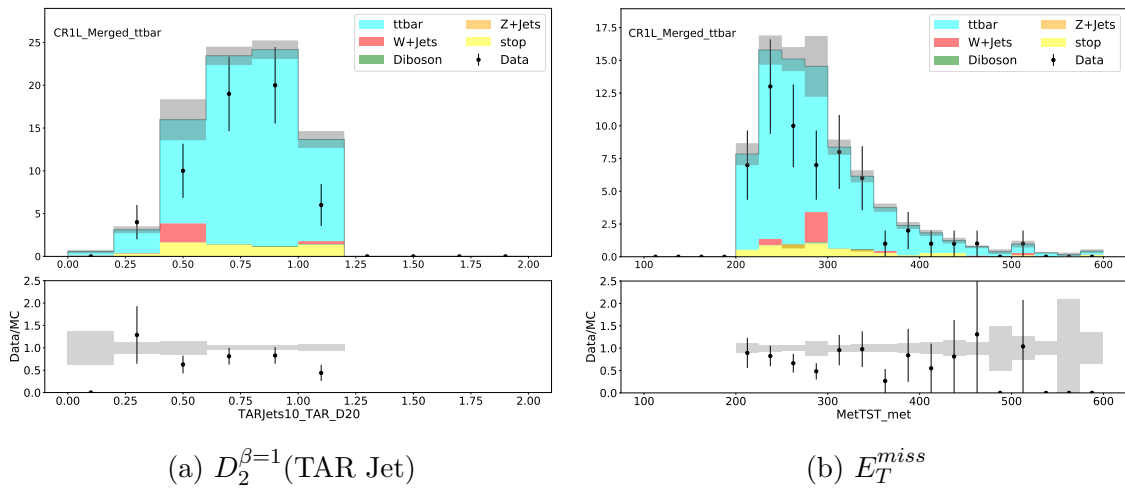


Figure A.6: Data-MC Comparisons in the “merged”  $t\bar{t}$  control region. Grey bands represent MC statistical uncertainty on each bin.

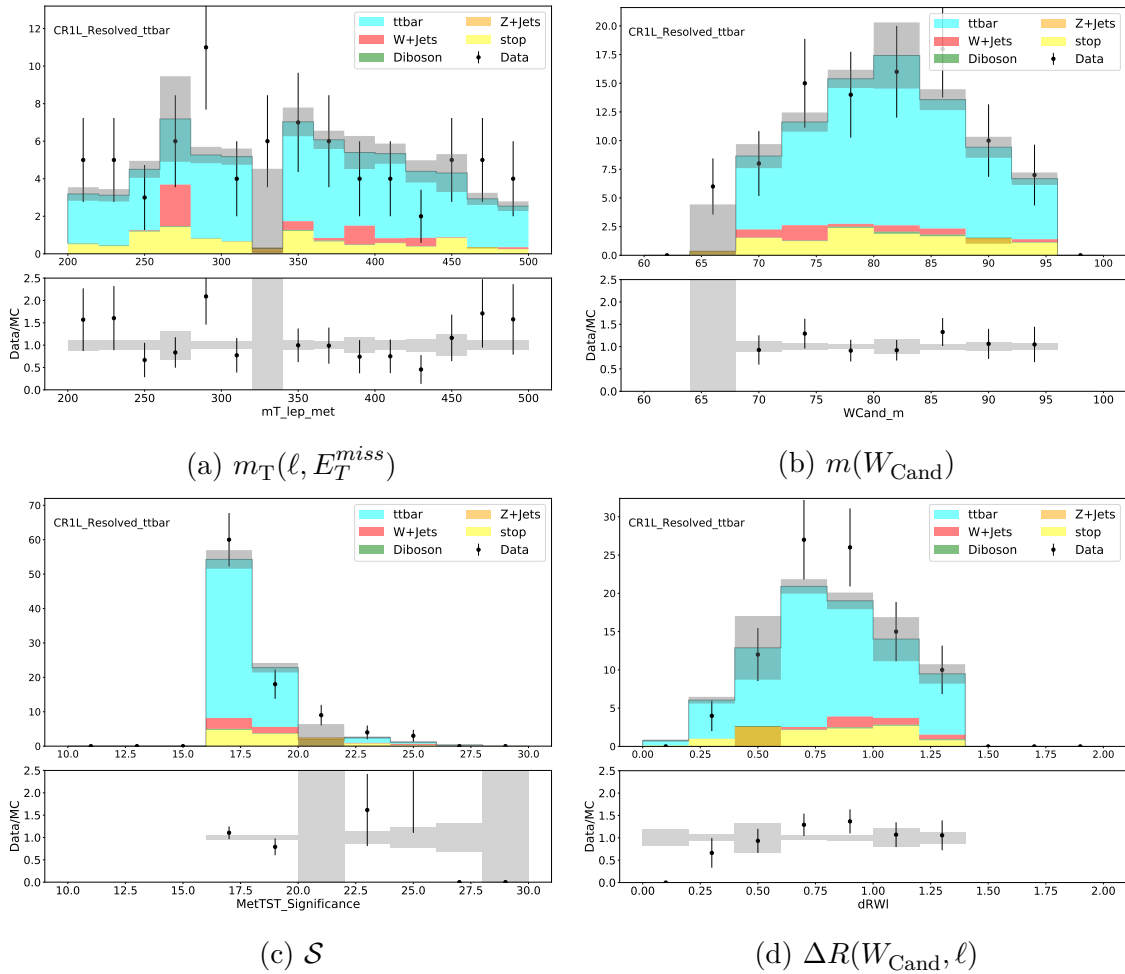


Figure A.7: Data-MC Comparisons in the “resolved”  $t\bar{t}$  control region. Grey bands represent MC statistical uncertainty on each bin.

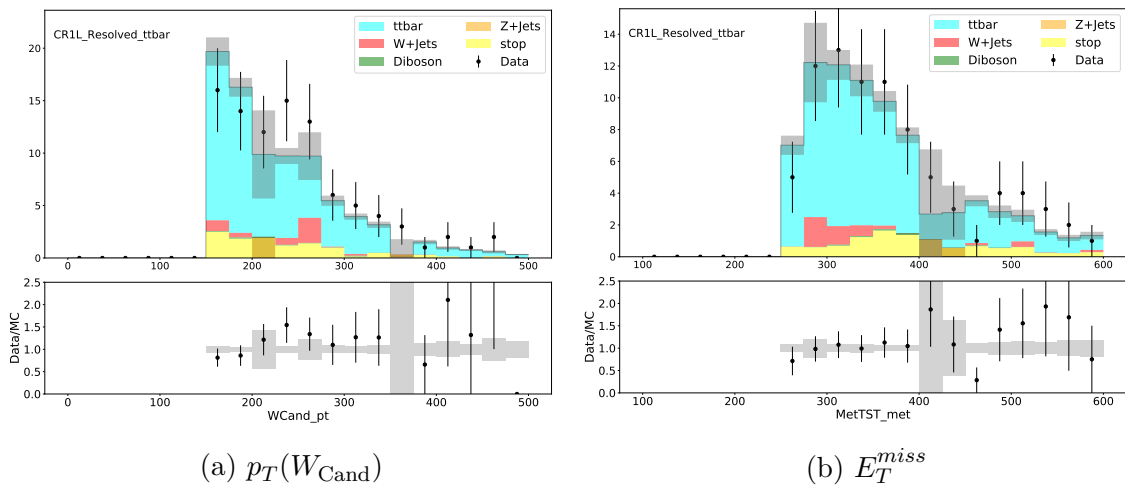


Figure A.8: Data-MC Comparisons in the “resolved”  $t\bar{t}$  control region. Grey bands represent MC statistical uncertainty on each bin.

## A.2.2 $W$ +jets Control Regions

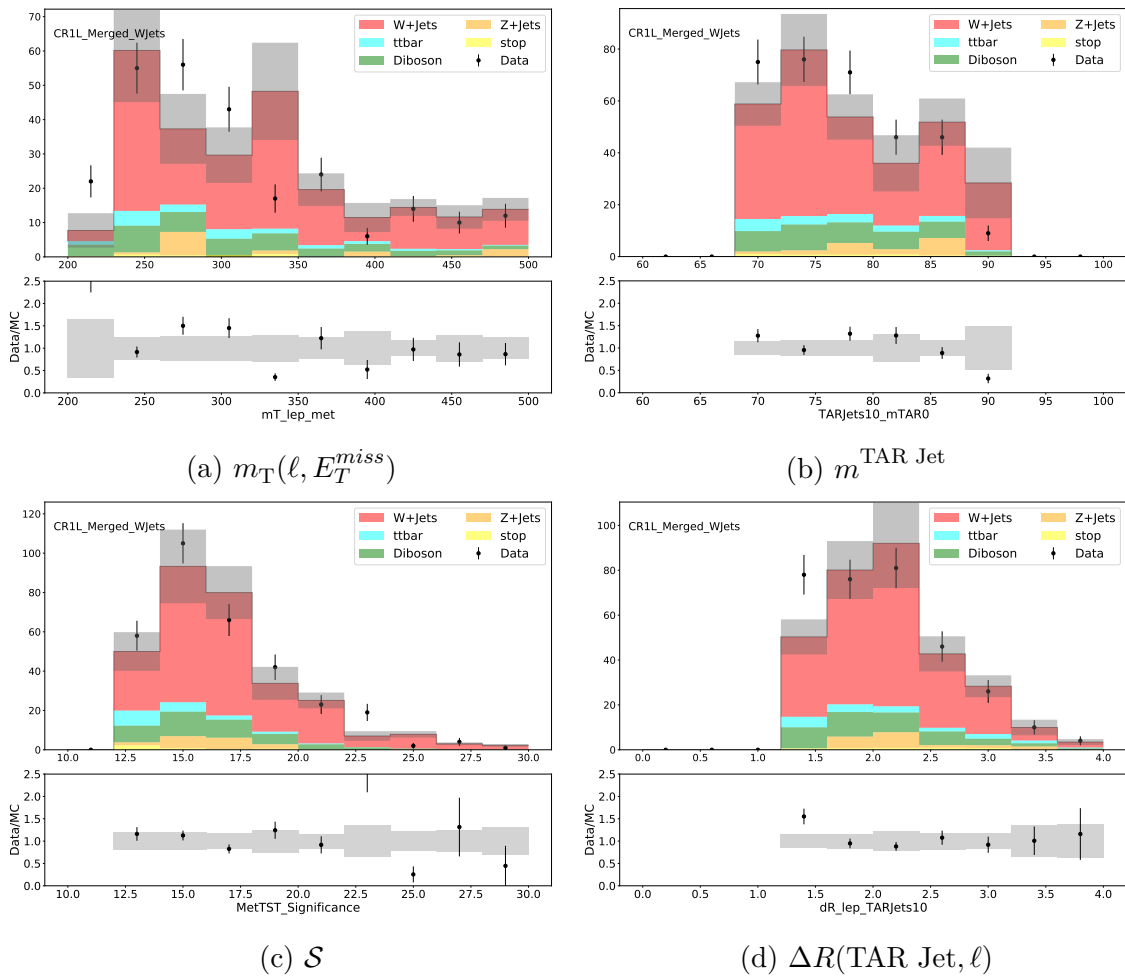


Figure A.9: Data-MC comparisons in the “merged”  $W$ +jets control region. Grey bands represent MC statistical uncertainty on each bin.

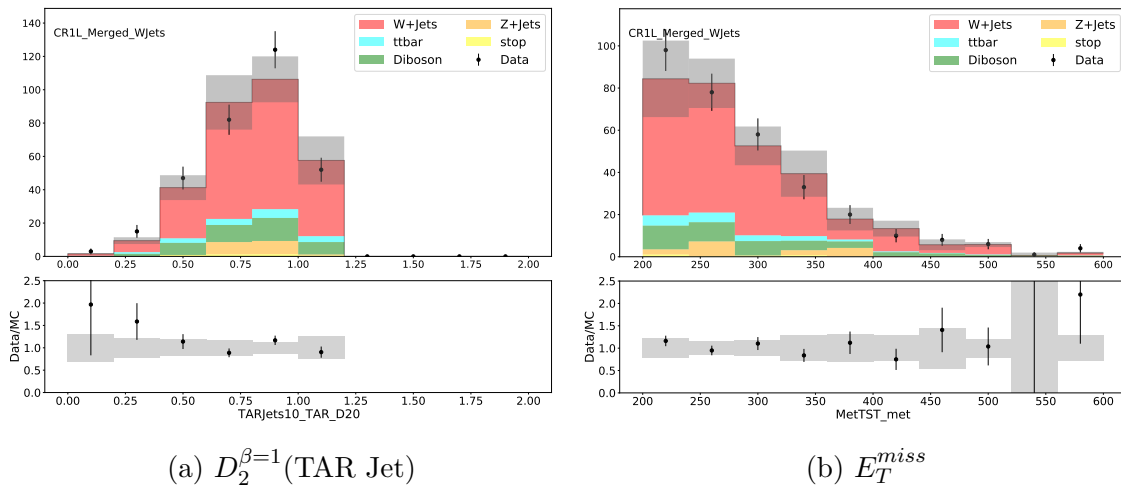


Figure A.10: Data-MC comparisons in the “merged”  $W$ +jets control region. Grey bands represent MC statistical uncertainty on each bin.

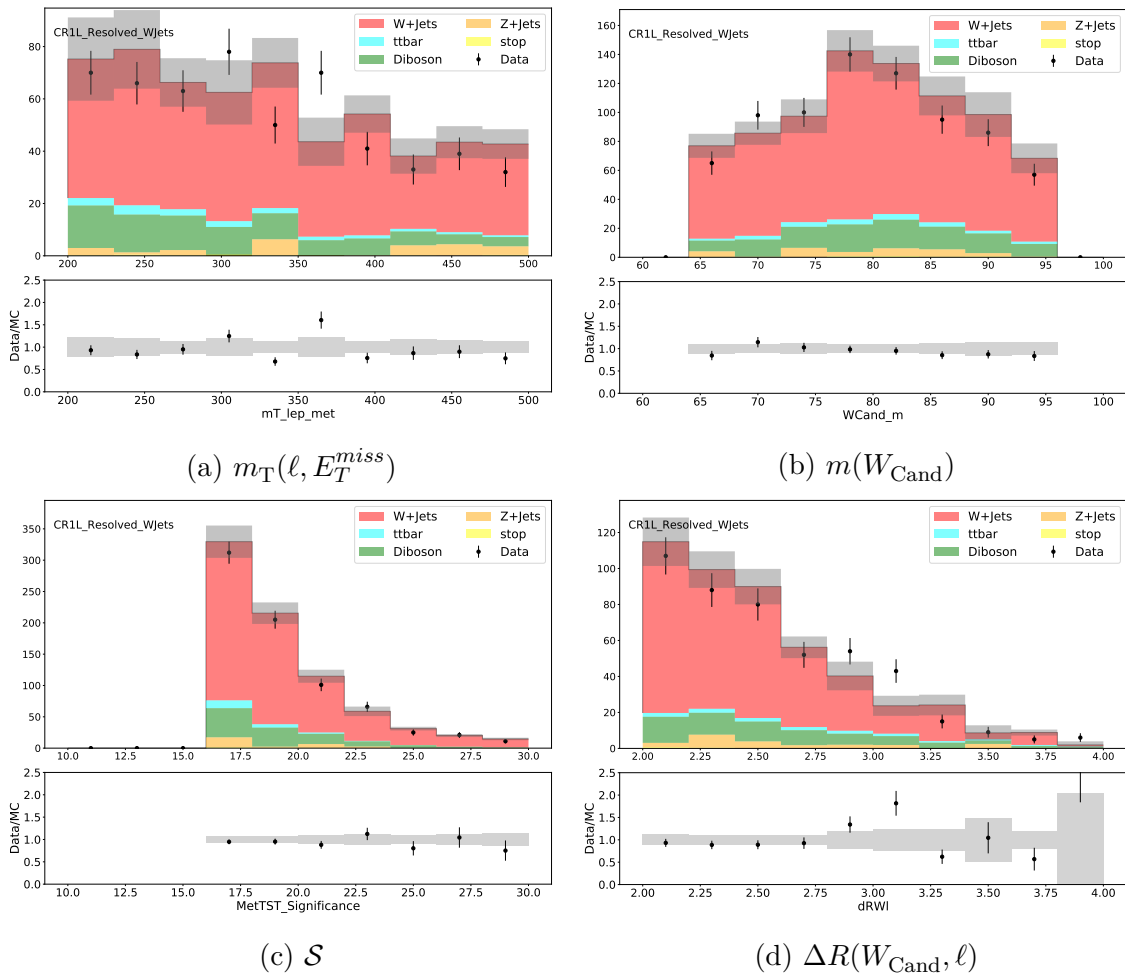


Figure A.11: Data-MC comparisons in the “resolved”  $W$ +jets control region. Grey bands represent MC statistical uncertainty on each bin.

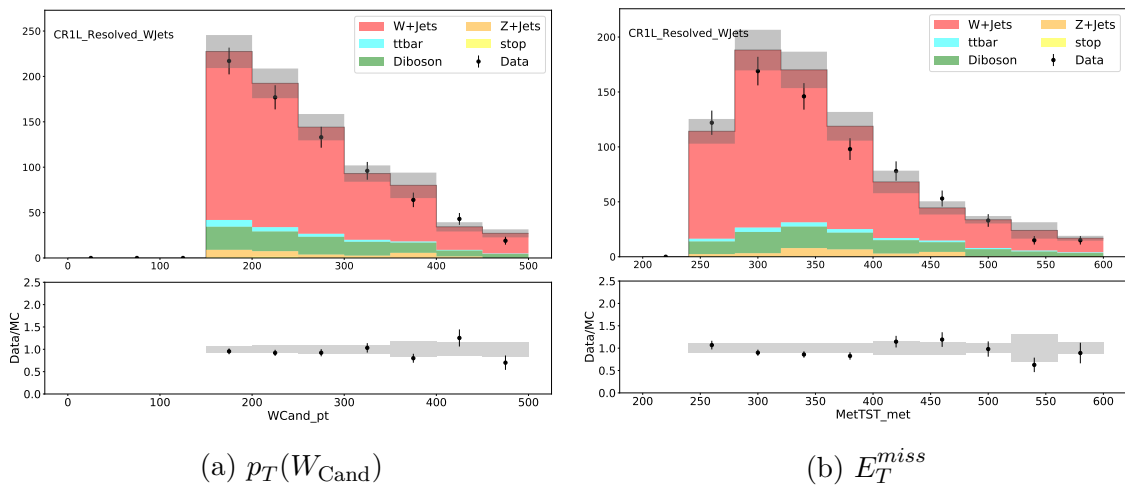


Figure A.12: Data-MC comparisons in the “resolved”  $W$ +jets control region. Grey bands represent MC statistical uncertainty on each bin.

### A.3 CR-SR Distribution Comparisons

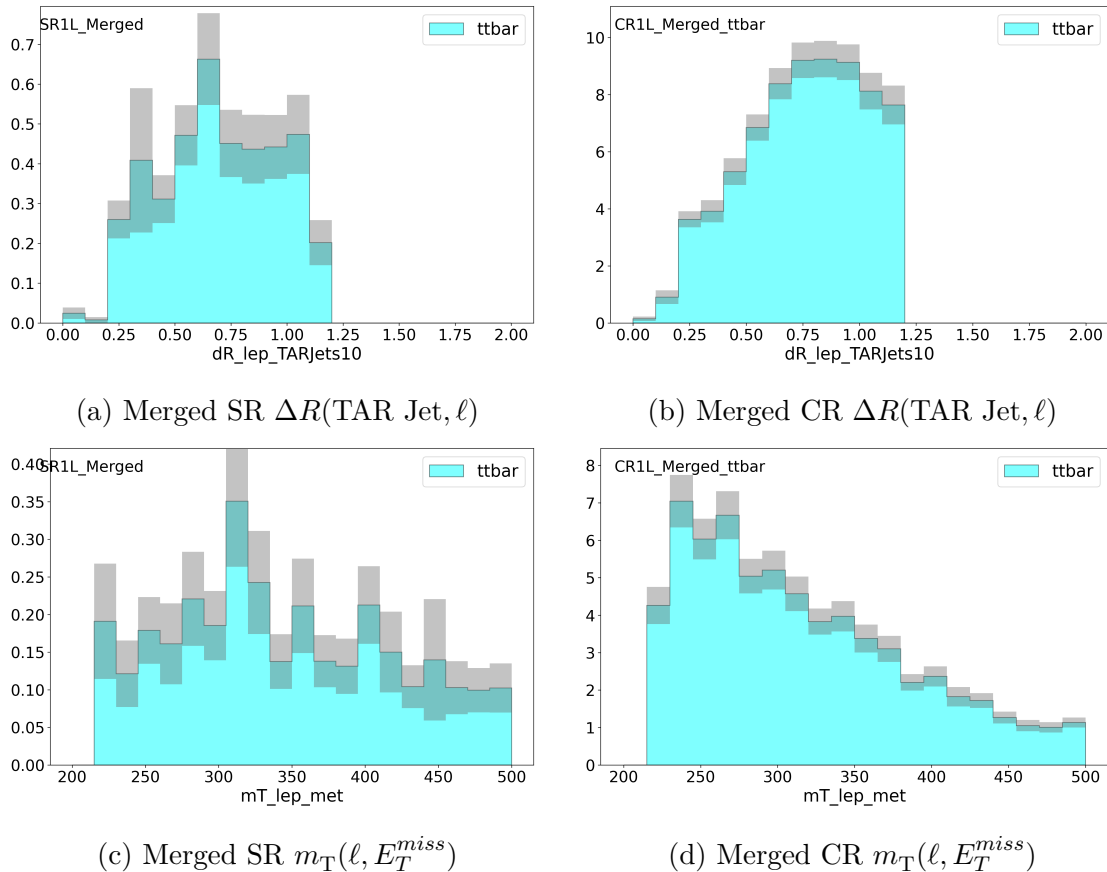


Figure A.13: Comparisons of kinematic variables between the “merged” signal region and  $t\bar{t}$  control region. **Left column:** signal region, **Right column:** control region. Grey bands represent MC statistical uncertainty on each bin.

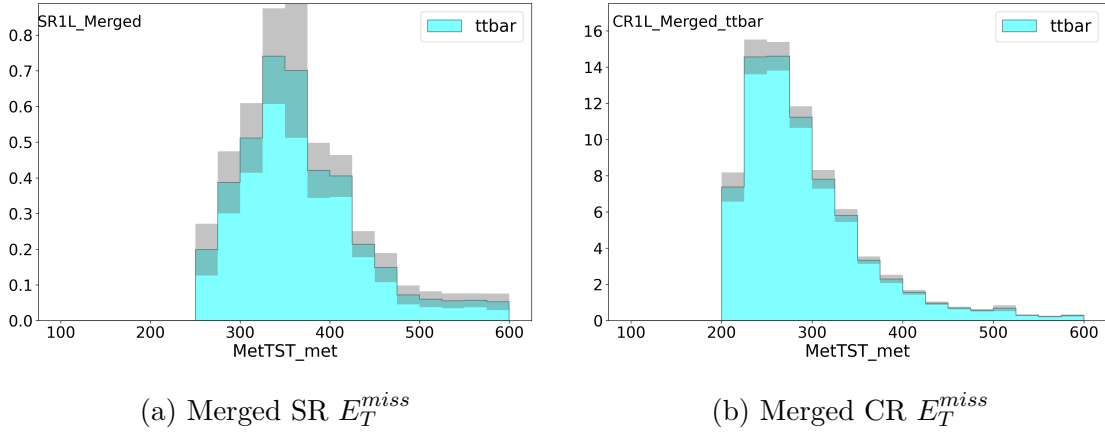


Figure A.14: Comparisons of kinematic variables between the “merged” signal region and  $t\bar{t}$  control region. **Left column:** signal region, **Right column:** control region. Grey bands represent MC statistical uncertainty on each bin.

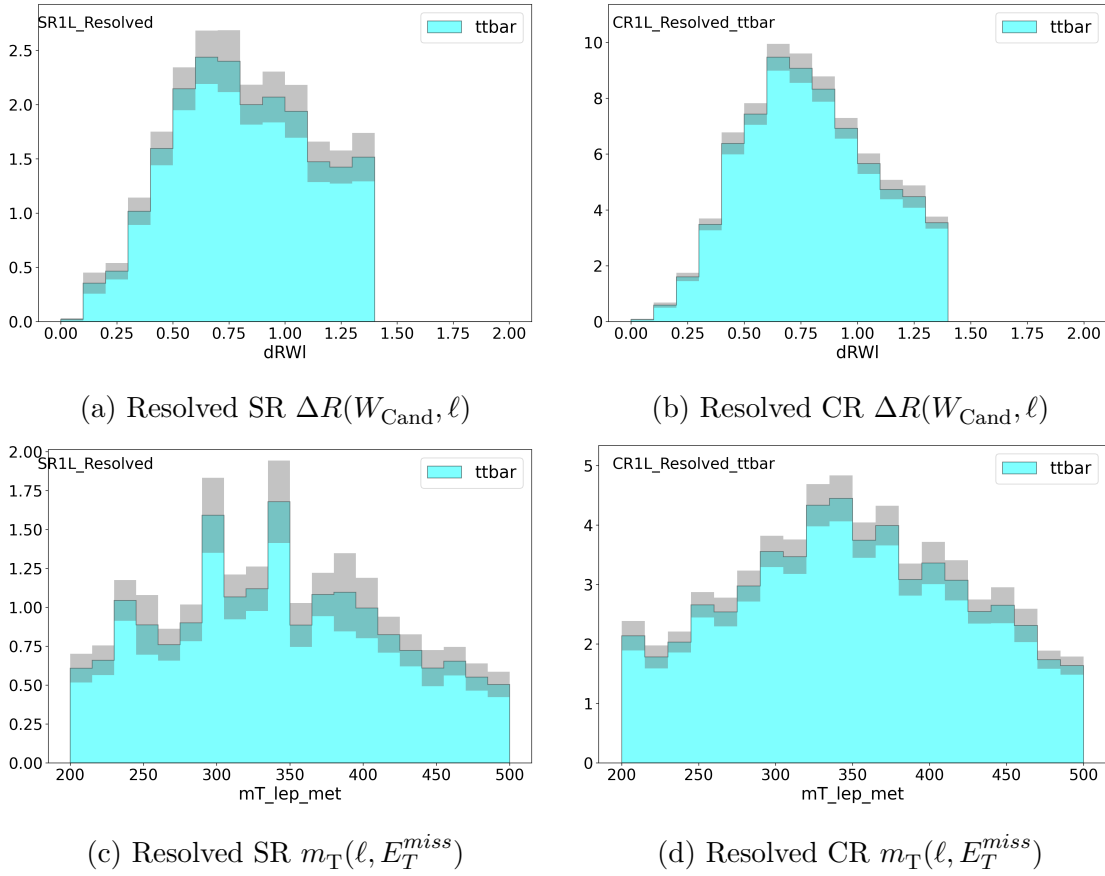


Figure A.15: Comparisons of kinematic variables between the “resolved” signal region and  $t\bar{t}$  control region. **Left column:** signal region, **Right column:** control region. Grey bands represent MC statistical uncertainty on each bin.

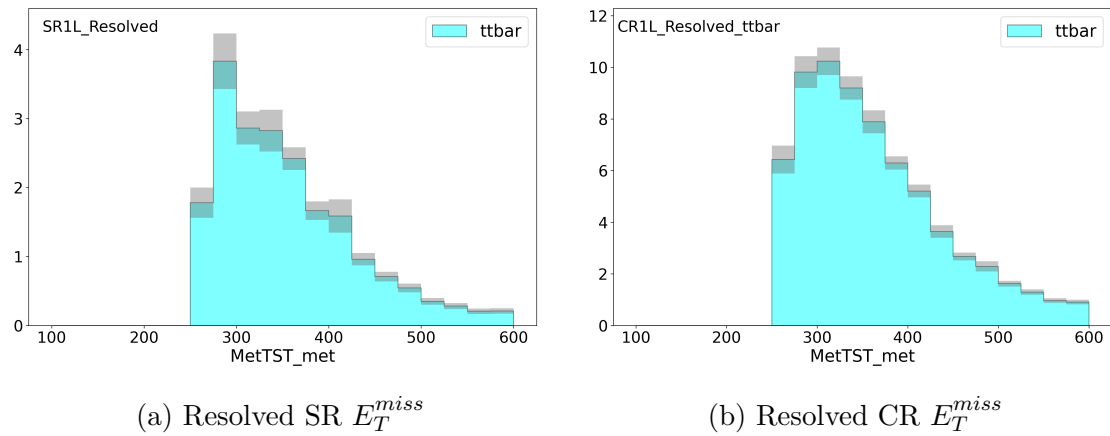


Figure A.16: Comparisons of kinematic variables between the “resolved” signal region and  $t\bar{t}$  control region. **Left column:** signal region, **Right column:** control region. Grey bands represent MC statistical uncertainty on each bin.

## A.4 $N - 1$ Plots - Highly Weighted Event Included

The following section contains  $N - 1$  plots in the “merged” signal region, with the inclusion of the single highly negatively weighted event excluded in Section 5.3.

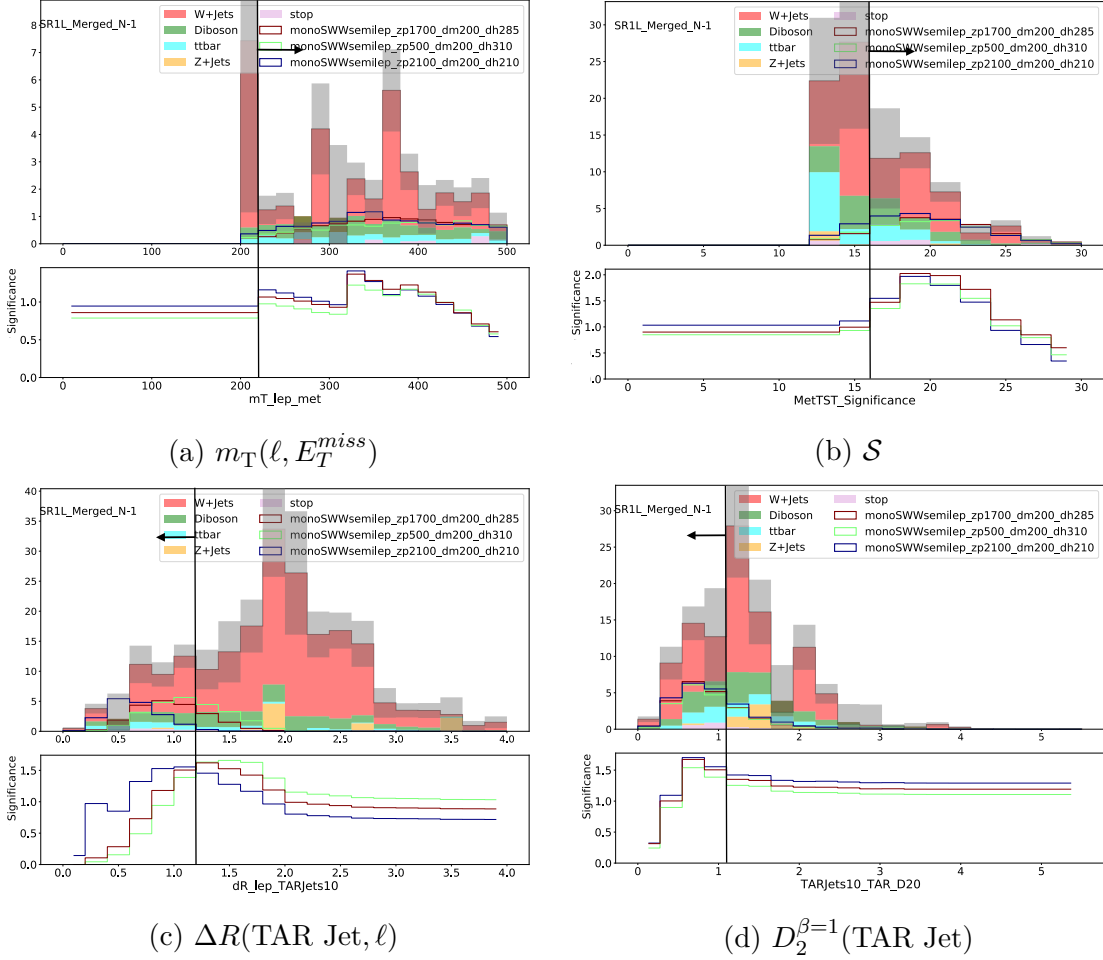


Figure A.17:  $N - 1$  plots in the “merged” signal region with the inclusion of a highly negatively weighted background event. Grey bands represent MC statistical uncertainty on each bin.

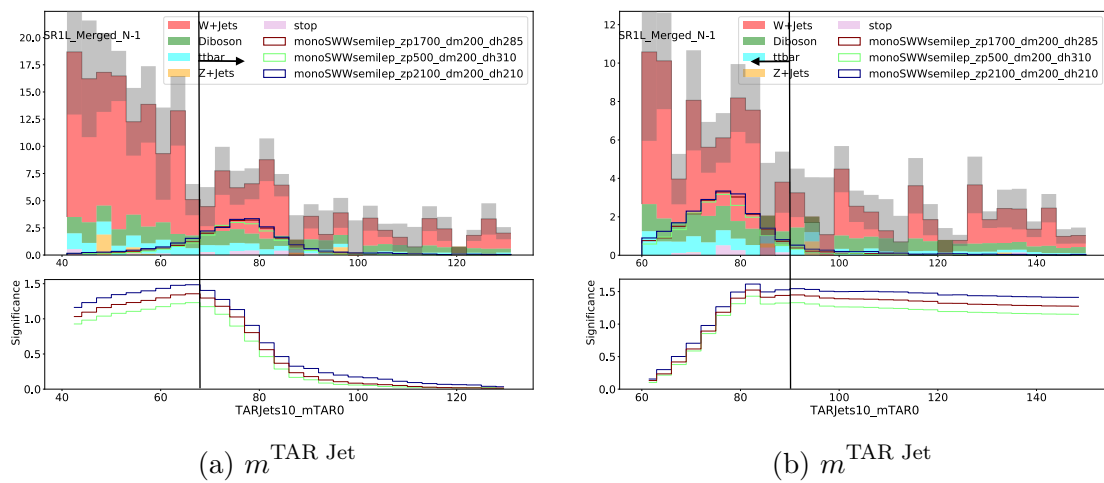


Figure A.18:  $N - 1$  plots in the “merged” signal region with the inclusion of a highly negatively weighted background event. Grey bands represent MC statistical uncertainty on each bin.

# Bibliography

1. Peskin, M. & Schroeder, D. *An Introduction To Quantum Field Theory* ISBN: 9780813345437 (Avalon Publishing, 1995).
2. Schwartz, M. *Quantum Field Theory and the Standard Model* ISBN: 9781107034730 (Cambridge University Press, 2014).
3. Group, P. D. *et al.* Review of Particle Physics. *Progress of Theoretical and Experimental Physics* **2020**. ISSN: 2050-3911 (Aug. 2020).
4. Griffiths, D. J. *Introduction to elementary particles; 2nd rev. version* (Wiley, New York, NY, 2008).
5. Duerr, M. *et al.* Hunting the dark Higgs. *JHEP* **04**, 143. arXiv: 1701.08780 [hep-ph] (2017).
6. Aad, G., Abajyan, T., Abbott, B., Abdallah, J. & Zwalinski, L. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B* **716**, 1–29. ISSN: 0370-2693 (2012).
7. Commons, W. *Standard Model of Elementary Particles* [https://commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg).
8. Feynman, R. P. *QED: The Strange Theory of Light and Matter* ISBN: 9780691164090 (Princeton University Press, Princeton, NJ, 2008).
9. Glashow, S. L. Partial-symmetries of weak interactions. *Nuclear Physics* **22**, 579–588. ISSN: 0029-5582 (1961).
10. Goldstone, J., Salam, A. & Weinberg, S. Broken Symmetries. *Phys. Rev.* **127**, 965–970 (3 1962).
11. Weinberg, S. A Model of Leptons. *Phys. Rev. Lett.* **19**, 1264–1266 (21 1967).

12. Campbell, J., Huston, J. & Krauss, F. *The black book of quantum chromodynamics: a primer for the LHC era* (Oxford University Press, Oxford, 2018).
13. Englert, F & Brout, R. Broken symmetry and the masses of gauge vector mesons. *Phys. Rev. Lett.* **13**, 321–323 (1964).
14. Higgs, P. W. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.* **13** (ed Taylor, J. C.) 508–509 (1964).
15. Higgs, P. W. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.* **145**, 1156–1163 (1966).
16. Vicente Montesinos, A. *Phenomenology of supersymmetric neutrino mass models* Other thesis (Apr. 2011). arXiv: 1104.0831 [hep-ph].
17. Zwicky, F. On the Masses of Nebulae and of Clusters of Nebulae. *Astrophysical Journal* **86**, 217 (Oct. 1937).
18. Rubin, V. C., Ford W. K., J. & Thonnard, N. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc). *Astrophysical Journal* **238**, 471–487 (June 1980).
19. Clowe, D. *et al.* A direct empirical proof of the existence of dark matter. *Astrophys. J. Lett.* **648**, L109–L113. arXiv: astro-ph/0608407 (2006).
20. NASA. The Bullet Cluster. General Photo. <https://chandra.harvard.edu/photo/2006/1e0657/> (2006).
21. LHC Machine. *JINST* **3** (eds Evans, L. & Bryant, P.) S08001 (2008).
22. Mobs, E. The CERN accelerator complex. Complexe des accélérateurs du CERN. General Photo. <https://cds.cern.ch/record/2197559> (2016).
23. Aad, G *et al.* The ATLAS Experiment at the CERN Large Hadron Collider. *JINST* **3**. Also published by CERN Geneva in 2010, S08003. 437 p. <https://cds.cern.ch/record/1129811> (2008).
24. Aad, G. *et al.* Drift Time Measurement in the ATLAS Liquid Argon Electromagnetic Calorimeter using Cosmic Muons. *Eur. Phys. J. C* **70**, 755–785. arXiv: 1002.4189 [physics.ins-det] (2010).

25. *RECAST framework reinterpretation of an ATLAS Dark Matter Search constraining a model of a dark Higgs boson decaying to two b-quarks* tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-032> (CERN, Geneva, 2019). <http://cds.cern.ch/record/2686290>.
26. Aad, G. *et al.* Search for Dark Matter Produced in Association with a Dark Higgs Boson Decaying into  $W^\pm W^\mp$  or  $ZZ$  in Fully Hadronic Final States from  $\sqrt{s} = 13$  TeV pp Collisions Recorded with the ATLAS Detector. *Phys. Rev. Lett.* **126**, 121802. arXiv: 2010.06548 [hep-ex] (2021).
27. Dittmaier, S. *et al.* Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. arXiv: 1101.0593 [hep-ph] (Jan. 2011).
28. Alwall, J. *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP* **07**, 079. arXiv: 1405.0301 [hep-ph] (2014).
29. Sjöstrand, T. *et al.* An introduction to PYTHIA 8.2. *Comput. Phys. Commun.* **191**, 159–177. arXiv: 1410.3012 [hep-ph] (2015).
30. Lonblad, L. & Prestel, S. Matching Tree-Level Matrix Elements with Interleaved Showers. *JHEP* **03**, 019. arXiv: 1109.4829 [hep-ph] (2012).
31. Ball, R. D. *et al.* Parton distributions for the LHC Run II. *JHEP* **04**, 040. arXiv: 1410.8849 [hep-ph] (2015).
32. Bothmann, E. *et al.* Event Generation with Sherpa 2.2. *SciPost Phys.* **7**, 034. arXiv: 1905.09127 [hep-ph] (2019).
33. *ATLAS simulation of boson plus jets processes in Run 2* tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-006> (CERN, Geneva, 2017). <https://cds.cern.ch/record/2261937>.
34. Gignac, M. *et al.* *Single Vector-Boson Monte Carlo setups for ATLAS Run 2 physics* tech. rep. (CERN, Geneva, 2021). <https://cds.cern.ch/record/2753199>.
35. Hoeche, S., Krauss, F., Schonherr, M. & Siegert, F. QCD matrix elements + parton showers: The NLO case. *JHEP* **04**, 027. arXiv: 1207.5030 [hep-ph] (2013).

36. Alioli, S., Nason, P., Oleari, C. & Re, E. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP* **06**, 043. arXiv: 1002.2581 [hep-ph] (2010).
37. Ball, R. D. *et al.* Parton distributions with LHC data. *Nucl. Phys. B* **867**, 244–289. arXiv: 1207.1303 [hep-ph] (2013).
38. Aad, G. *et al.* Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J. C* **76**, 292. arXiv: 1603.05598 [hep-ex] (2016).
39. Aaboud, M. *et al.* Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J. C* **79**, 639. arXiv: 1902.04655 [physics.ins-det] (2019).
40. Aaboud, M. *et al.* Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C* **77**, 466. arXiv: 1703.10485 [hep-ex] (2017).
41. Cacciari, M., Salam, G. P. & Soyez, G. The anti- $k_t$  jet clustering algorithm. *JHEP* **04**, 063. arXiv: 0802.1189 [hep-ph] (2008).
42. *Tagging and suppression of pileup jets with the ATLAS detector* tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-018> (CERN, Geneva, 2014). <https://cds.cern.ch/record/1700870>.
43. *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment* tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-003> (CERN, Geneva, 2017). <https://cds.cern.ch/record/2255226>.
44. *Track assisted techniques for jet substructure* tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-012> (CERN, Geneva, 2018). <https://cds.cern.ch/record/2630864>.
45. Aad, G. *et al.* Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C* **77**, 490. arXiv: 1603.02934 [hep-ex] (2017).

46. Cacciari, M., Salam, G. P. & Soyez, G. The Catchment Area of Jets. *JHEP* **04**, 005. arXiv: 0802.1188 [hep-ph] (2008).
47. Napolitano, F. Improving Dark Matter searches using Track Assisted Reclustered (TAR) jets with the ATLAS detector at  $\sqrt{s} = 13$  TeV. <https://cds.cern.ch/record/2683856> (2019).
48. *Object-based missing transverse momentum significance in the ATLAS detector* tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2018-038> (CERN, Geneva, 2018). <https://cds.cern.ch/record/2630948>.
49. Calafiura, P, Lavrijsen, W, Leggett, C, Marino, M & Quarrie, D. The Athena Control Framework in Production, New Developments and Lessons Learned. <http://cds.cern.ch/record/865624> (2005).
50. Cowan, G., Cranmer, K., Gross, E. & Vitells, O. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C* **71**. [Erratum: *Eur.Phys.J.C* **73**, 2501 (2013)], 1554. arXiv: 1007.1727 [physics.data-an] (2011).
51. Larkoski, A. J., Salam, G. P. & Thaler, J. Energy Correlation Functions for Jet Substructure. *JHEP* **06**, 108. arXiv: 1305.0007 [hep-ph] (2013).
52. Aaboud, M. *et al.* Measurement of jet-substructure observables in top quark,  $W$  boson and light jet production in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *JHEP* **08**, 033. arXiv: 1903.02942 [hep-ex] (2019).
53. Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics* **9**, 60–62. <https://doi.org/10.1214/aoms/1177732360> (1938).
54. Read, A. L. Presentation of search results: the CL<sub>s</sub> technique. *Journal of Physics G: Nuclear and Particle Physics* **28**, 2693–2704. <https://doi.org/10.1088/0954-3899/28/10/313> (2002).
55. Baak, M. *et al.* HistFitter software framework for statistical data analysis. *Eur. Phys. J. C* **75**, 153. arXiv: 1410.1280 [hep-ex] (2015).
56. Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC, ATLAS-CONF-2019-021. <https://cds.cern.ch/record/2677054> (June 2019).