

1 Search for Dark Matter Produced in Association with a 2 Hypothetical Dark Higgs Boson Decaying to W^+W^- in 3 the $q\bar{q}\ell\nu$ Final State Using pp Collisions Recorded with 4 the ATLAS Detector

5 Candidacy Report

Danika MacDonell

July 15, 2020

7 1 Introduction

The proposed thesis is a search for dark matter (DM) using high energy proton-proton (pp) collision data recorded with the ATLAS detector at the Large Hadron Collider (LHC). The search targets a final state signature of DM production in association with the emission of a hypothetical higgs boson s in the dark sector, which subsequently decays to a pair of W bosons. The search is motivated by and optimized with a “dark higgs model” [1] shown in figure 1, in which the s is emitted from a hypothetical Z' gauge boson in the dark sector, which itself mediates the production of DM particles from the annihilation of “initial state quarks” (IS qq) constituting the colliding protons.

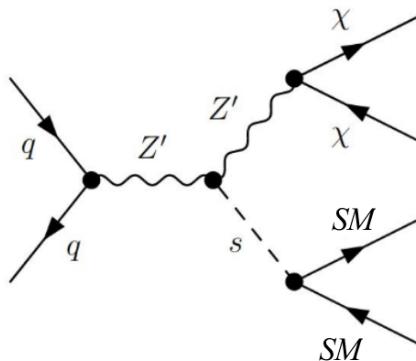


Figure 1: Dark higgs model

Any DM that may be produced in the qq collisions is expected not to interact to any measurable extent with the ‘normal’ matter constituting the detector. As such, it is assumed that the momentum

18 carried by the DM would escape the detector undetected. The law of momentum conservation,
 19 which requires the vector sum of momenta of all measured particles produced by the collision to
 20 match that of the initial state quarks, can be used to infer the presence of undetected particles if the
 21 IS quark momenta are known. As the fraction of proton momentum carried by each of the IS quarks
 22 - described by “parton distribution functions” (PDFs) [2] - is statistical in nature, their momenta
 23 in the direction of the pp beam line cannot be known precisely. However, IS quark momentum
 24 in the plane transverse to the beam line is in general negligibly small compared with the collision
 25 energy, so it can be expected to a high degree of precision that the final state particle momenta in
 26 this transverse plane will sum to zero. This expectation implies that DM produced at the LHC will
 27 exhibit a signature of high missing transverse momentum in the final state. This two dimensional
 28 missing transverse momentum vector is typically denoted “ \vec{E}_T^{miss} ”, and its magnitude “ E_T^{miss} ”:

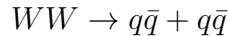
$$\vec{E}_T^{\text{miss}} = - \sum_i \vec{p}_{x,i} + \vec{p}_{y,i} \quad (1)$$

29 where the sum is over all visible particles in the event.

30 The aim of the proposed thesis work is to apply selections, including a requirement of high
 31 E_T^{miss} , to data from the ATLAS detector to optimize the sensitivity of the data to the dark higgs
 32 signal model. The data will subsequently be compared with standard model (SM) background and
 33 signal processes simulated with Monte Carlo to search for an above-background excess in the data
 34 consistent with the signal model.

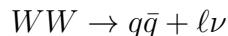
35 The search is split into two separate analyses:

- 36 1. **The hadronic decay channel:** In this channel, each of the two W bosons in the final state
 37 decay to quarks, which subsequently hadronize in the detector and are measured as jets in
 38 the ATLAS calorimeter.



39 Given that the branching fraction for hadronic W decay is 0.68 [3], the fully hadronic WW
 40 decay occurs with a branching fraction of 0.46 ($= 0.68^2$). In addition to its substantial
 41 branching fraction, this channel has the advantage of being able to fully reconstruct the
 42 momenta of both W bosons, and as such the s mass. The drawback is that there are SM
 43 processes with large production cross section which also produce a final state of multiple
 44 jets in the detector, a subset of which pass the other signal selection criteria and represent a
 45 sizeable background in the analysis.

- 46 2. **The semileptonic decay channel:** One W boson decays to quarks in this channel, and the
 47 other decays to a lepton and a neutrino, where the lepton ℓ is either an electron or a muon.



48 Despite its lower branching fraction of 0.29 [3] compared with the fully hadronic decay
 49 channel, the semileptonic channel has the advantage that the requirement of having one
 50 lepton in the final state reduces the background of SM processes. However, this comes at the
 51 cost of additional E_T^{miss} from the neutrino production, which both inhibits full reconstruction

52 of the leptonically-decaying W boson and cannot easily be distinguished from the E_T^{miss}
53 associated with DM production in the signal model.

54 There is a mature analysis searching for the dark higgs signal in the hadronic decay channel,
55 and the proposed thesis will focus on the semileptonic decay channel.

56 The WW decay can also proceed via a fully leptonic channel, in which both W bosons decay to
57 a lepton and a neutrino. This channel would give a clean signature of two leptons and no jets, but
58 is currently not considered as a viable search channel due to its relatively low branching fraction
59 of 0.046 [3].

60 Once the analysis is fully developed in the semileptonic decay channel, this channel will be
61 statistically combined with the hadronic search channel. In the event that the combined search
62 does not show an above-background excess, the analysis will place exclusion limits - at some
63 specified level of confidence, typically 95% - on the range of Z' and s masses in the dark higgs
64 model to which the combined search is sensitive.

65 **2 Motivation**

66 The proposed thesis work would contribute to the extensive ongoing DM search program at the
67 LHC. The LHC search program is motivated by compelling evidence from observational astron-
68 omy for the existence of DM constituting 85% [4] of all matter in the universe. Despite clear
69 evidence from observational astronomy for its gravitational interactions with normal matter, DM
70 has yet to be detected through the weak, strong or electromagnetic interactions. As such, its com-
71 position and non-gravitational interactions - if any - remain largely a mystery. The current most
72 widely accepted and theoretically motivated DM candidates take the form of fundamental particles,
73 yet there is no particle in the standard model of particle physics that could represent a viable DM
74 candidate [5]. DM is therefore widely assumed to be a “beyond-standard-model” (BSM) particle.

75 **2.1 Dark Matter Search Methods**

76 There are three complementary approaches used to search for particle DM: direct and indirect
77 detection, and collider searches. Direct detection searches [6, 7] aim to directly detect evidence of
78 a recoil induced by elastic scattering between a DM particle in the galactic halo passing through
79 the detector and a target particle in the detector. Indirect searches [8, 9] use observational data
80 to search for evidence of products produced by DM annihilation or decay in particular regions of
81 the observable universe expected to have a high DM density. Collider searches [10], of which
82 the proposed thesis work is an example, study the decay products from high-energy collisions of
83 subatomic particles to search for an above-background excess of events that could be consistent
84 with DM having been produced in some of the collisions.

85 **2.2 Models of Dark Matter Production at Colliders**

86 Models of DM production in colliders can range in complexity from an effective field theory (EFT),
87 where the DM production mechanism is completely unspecified, to a complete model such as

88 supersymmetry [11], which predicts viable DM candidates as part of a hypothesized extension to
 89 the SM designed to address a range of phenomena unexplained by the SM.

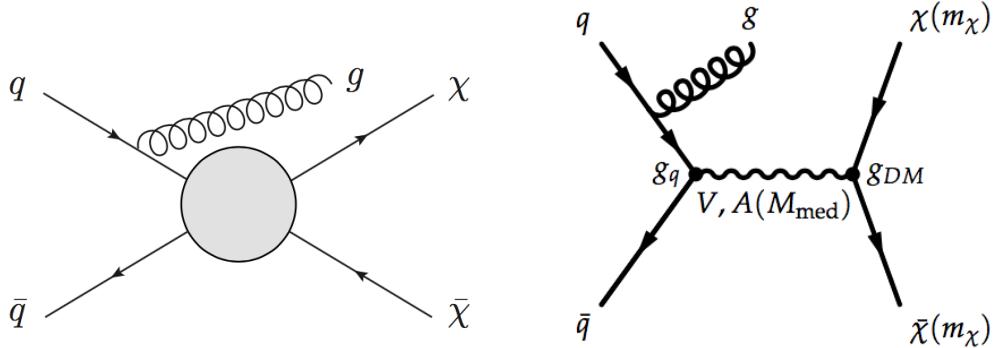


Figure 2: Left: Mono-jet process in the EFT framework (source: [12]). Right: Mono-jet process in a simplified model framework, where the pair production of DM occurs via a new vector or axial-vector (V, A) mediator of mass M_{med} , which couples to quarks and DM with coupling constants g_q and g_{DM} , respectively (source: [13])

90 In principle, complete theories of physics beyond the SM, such as the minimal supersymmetric
 91 SM (MSSM) [14] can offer theoretically motivated and experimentally accessible models which
 92 specify the details of candidate processes by which the colliding partons may annihilate to produce
 93 DM. However, these theories tend to be quite complex, with many free parameters - over 100 in
 94 the case of MSSM [10] - most of which need to be fixed to generate a reasonably testable model.
 95 Relying on complete theories alone to guide experimental signatures may run the risk of missing
 96 important parameter space of new physics for which a complete theory has not yet been developed.

97 Simplified models, widely used in recent and ongoing DM searches at the LHC, are designed
 98 to bridge the gap between EFT and complete theories. They provide a ‘first-order’ description of
 99 theoretically motivated new physics scenarios that could be accessible at collider energies. They
 100 provide guidance for experimental searches without fully specifying the details of any additional
 101 new physics at energies above the collider scale that would be needed for a complete theory [10].
 102 In terms of DM production at the LHC, one or more new mediators associated with new physics
 103 scenarios may be considered which allow for mixing between SM particles and DM. The process
 104 by which the mixing occurs is represented with a tree-level diagram whose experimental signature
 105 would be accessible at LHC energies, such as the diagram shown in figure 2 which represents a
 106 DM benchmark model featured in the 2015 report of the ATLAS/CMS Dark Matter Forum [13].

107 2.3 Dark Higgs Model

108 The proposed search is motivated by and interpreted with a simplified model described in [1] where
 109 DM is pair produced from a Z boson in the dark sector - known as the Z' boson - in association
 110 with the emission of a dark sector higgs boson s . Figure 1 shows one of the leading order Feynman
 111 diagram representing this simplified model. The dark sector higgs boson subsequently decays to a
 112 pair of massive SM particles through a small mixing with the SM higgs boson.

113 The generic hypothesis of ‘dark sector’ mediators which exhibit a small mixing with SM parti-
114 cles is motivated in part by the theoretical need for creation and annihilation mechanisms between
115 DM and SM particles in the well-motivated and popular ‘thermal freeze-out’ hypothesis [15]. In
116 the thermal freeze-out hypothesis, DM and SM particles interacted at a sufficient rate in the high
117 temperature and matter density of the early universe to remain in thermal equilibrium. Once the
118 DM-SM interaction rate dropped below the expansion rate of the universe, the DM rapidly “froze
119 out” of thermal equilibrium with the SM particles, at which point the relic abundance of DM was
120 fixed to the value observed in the present-day universe.

121 The dark higgs boson in particular is motivated by the need to generate the masses of DM and
122 any other particles residing in the dark sector [1]. If dark sector particle masses are generated via
123 a ‘dark sector higgs mechanism’, this would naturally imply the existence of a dark higgs boson.

124 The model presented in [1] considers a scenario in which there is another dark sector mediator
125 - taken to be the Z' - with a mass within the range accessible to measurement at the LHC, which
126 mediates DM production from parton collisions. In this case, the Z' could radiate a dark higgs
127 boson. Theoretically, the observed relic abundance of DM in the universe suggests that couplings
128 between particles in the dark sector should be quite large, which could result in a sufficiently high
129 probability of dark higgs emission from the Z' to produce a measurable signature in LHC collisions
130 [1].

131 Experimentally, the dark higgs signature would be distinct from generic mono-X signatures
132 where the SM particle that the DM recoils against is produced via initial state radiation (as shown
133 in figure 2, because as long as the mass of the s is well below the typical momentum transfer of the
134 collisions, the s will tend to be highly boosted in the plane transverse to the beam line. As a result,
135 the SM decay products will tend to be much higher in transverse momentum and highly collimated
136 compared with particles produced via initial state radiation.

137 2.4 Experimental Background

138 This section discusses a previous experimental search which set exclusion limits on the dark higgs
139 model described in Section 2.3, in the case where the dark higgs decays to a pair of b quarks.
140 It goes on to describe the m_s range that the ongoing and proposed searches in the $s \rightarrow WW$
141 decay channel target, and explain how this target is complementary to the m_s range explored by
142 the previous search in the $s \rightarrow bb$ channel.

143 2.4.1 Re-interpretation of Mono-H(bb) Search with a Dark Higgs Mediator

144 Last year, the “mono-H(bb)” DM search, originally published in 2018 [16] with 79.8 fb^{-1} of data
145 taken with the ATLAS detector at a 13 TeV centre of mass energy, was re-interpreted [17] using
146 the RECAST framework [18] to set exclusion limits on the dark higgs model in the case of the
147 dark higgs decaying to a pair of b-quarks.

148 The mono-H(bb) search selects for a final state of missing transverse energy along with two
149 jets in the calorimeter, both of which must be tagged as having originated from the hadronization
150 of b quarks. When the mono-H(bb) analysis was originally published, it had been developed and
151 interpreted using a simplified model shown in figure 3 left, in which the DM is pair produced from
152 a new pseudoscalar higgs boson, along with the emission of a SM higgs boson, which subsequently
153 decays to two b quarks.

154 The original analysis was preserved using RECAST [18], a framework designed to preserve
 155 searches for new physics with high energy collision data in such a way that the searches can
 156 be readily re-analyzed with alternative models of new physics which predict the same final state
 157 signature. The RECAST framework was used to re-interpret the mono-H(bb) search using the
 158 dark higgs model shown in figure 3 right, where the dark higgs, whose mass is left as a floating
 159 parameter, decays to a pair of b quarks.

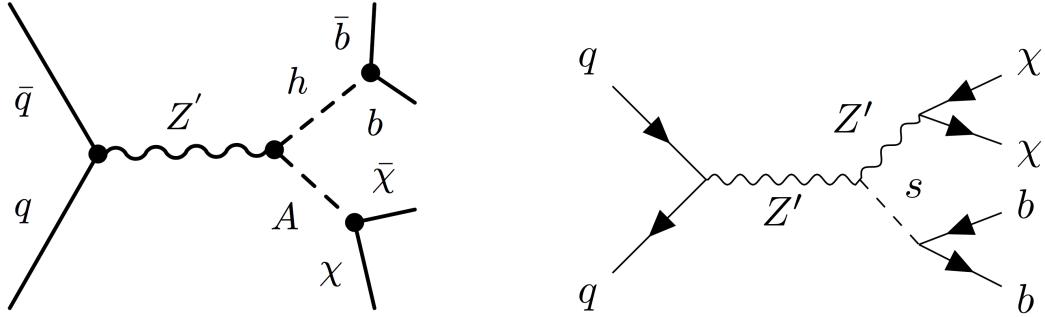


Figure 3: Left: Leading-order Feynman diagram representing the model used to interpret the mono-H(bb) DM search. Right: A leading-order Feynman diagram representing the model used to guide and interpret the mono-s(bb) re-interpretation of the mono-H(bb) search. The dark higgs mediator mass m_S is allowed to float in the dark higgs model.

160 The re-interpreted mono-H(bb) search set limits, shown in figure 4, on the phase space of Z'
 161 and s masses excluded by the data. To set these exclusion limits, several free parameters in the
 162 dark higgs model need to be fixed, as detailed in Section 6.1.1.

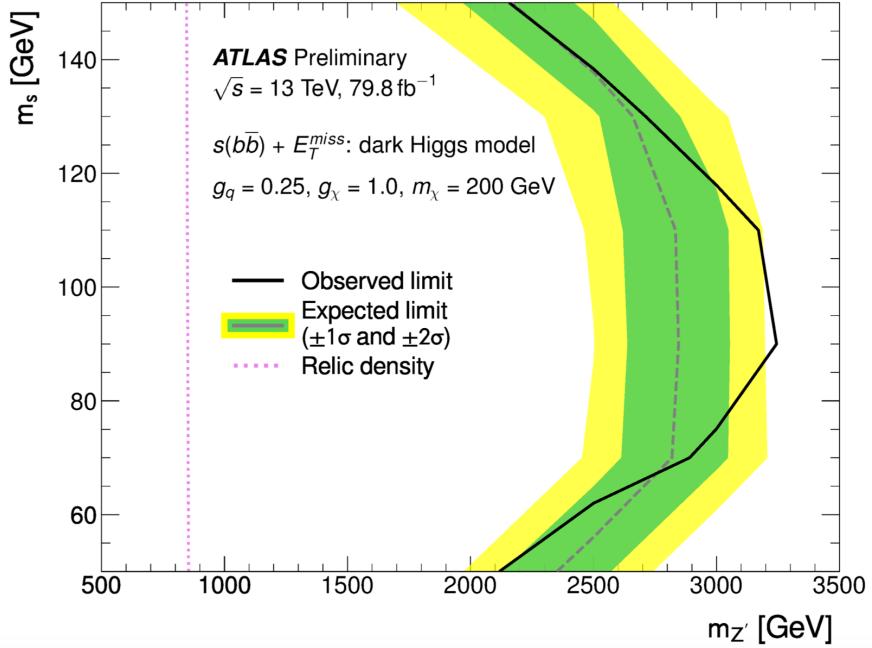


Figure 4: Exclusion limits set on the Z' and s mediator masses by the mono-H(bb) search re-interpreted with the dark higgs model. Mass combinations to the left of the solid black curve are excluded at a 95% confidence level. Source: [17]

163 2.4.2 Dark Higgs Decay Modes

164 The re-interpreted mono-H(bb) search probed the dark higgs model in the case where the s emitted
165 from the Z' mediator decays to a pair of b quarks. However, predictions of the dark higgs branching
166 ratio presented in the re-interpretation paper [17] and shown in figure 5 indicate that the $s \rightarrow bb$
167 decay mode is only sensitive to a range of dark higgs masses up to ~ 150 GeV, above which the
168 WW decay mode dominates in sensitivity.

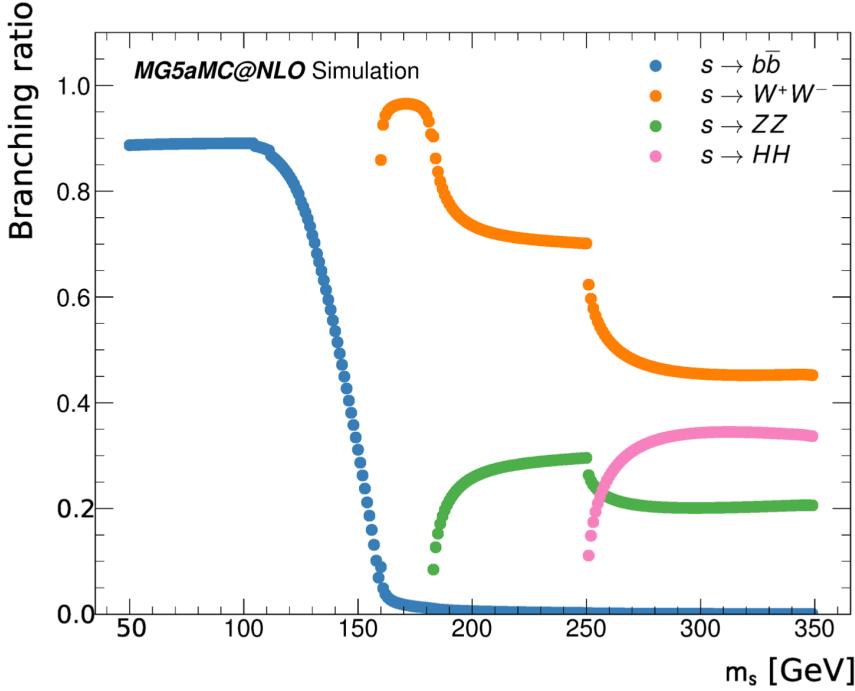


Figure 5: Predicted dark higgs decay branching ratios as a function of dark higgs mass. Source: [17]

169 There is no a-priori reason to expect the dark higgs mass m_s to necessarily lie below 150 GeV.
 170 Therefore, it's important to probe the model in the higher m_s range by analyzing the $s \rightarrow WW$
 171 decay mode. The proposed thesis work will analyze the latter decay mode using the full 139 fb^{-1}
 172 ATLAS run 2 dataset, focusing on the semileptonic final state $WW \rightarrow q\bar{q}\ell\nu$.

173 3 Introduction to the LHC and the ATLAS detector

174 The Large Hadron Collider (LHC) [19] is a circular proton-proton collider which resides in a 27 km
 175 tunnel near the European Organization for Nuclear Research (CERN). Superconducting magnets
 176 are used to accelerate counter-rotating bunched proton beams to high energy, and direct the beams
 177 into head-on collisions at four interaction points around the ring. The collisions take place at a
 178 world-leading centre of mass energy of up to 13 TeV.

179 Each interaction point is surrounded by a detector, which measures the energetic debris of
 180 particles produced by the high energy collisions to perform precision measurements of the SM and
 181 search for new physics. ATLAS (A Toroidal LHC ApparatuS) [20] is one of two multi-purpose
 182 detectors at the LHC, designed to record and study a wide range of physics processes resulting
 183 from the collisions.

184 The ATLAS detector, shown schematically in figure 6, provides full 4π coverage around the
 185 interaction point, with the exception of the beam pipe. It consists of several layers of sub-detectors,
 186 each of which is specialized for recording certain kinematic information and particle types. The
 187 sub-detectors are described in some detail below.

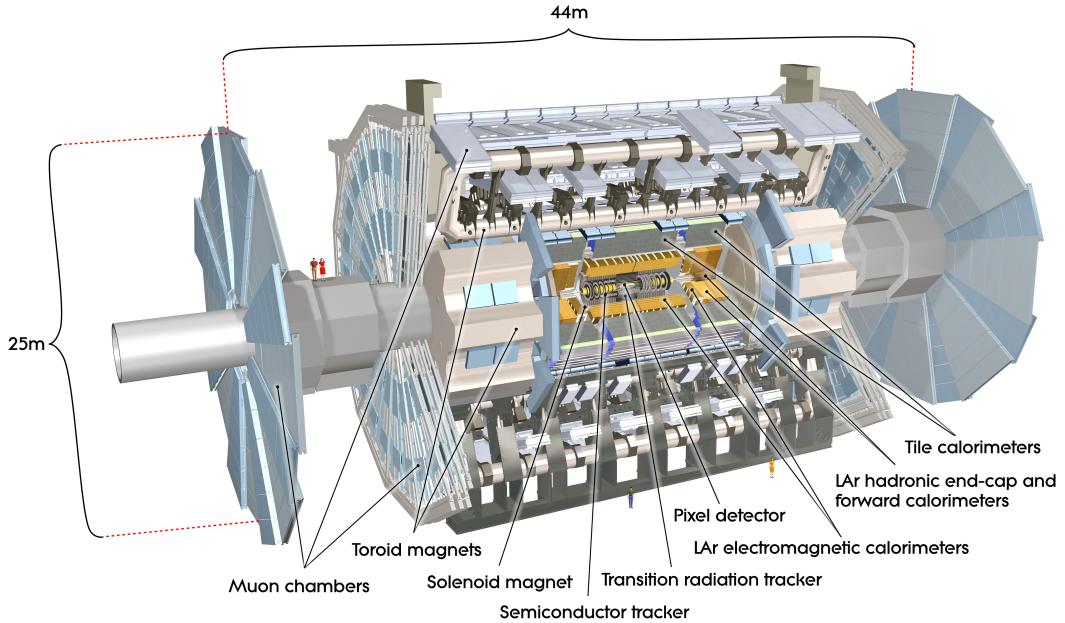


Figure 6: Schematic diagram of the ATLAS detector. Source: [20]

3.1 The Inner Detector

The inner detector (ID), located nearest the beam pipe, is specialized for charged particle tracking. It is immersed in a 2T magnetic field oriented parallel to the beam pipe, which bends the trajectories (“tracks”) of electrically charged particles as they pass through the field. Three distinct but complementary tracking technologies are employed along with pattern recognition tools to map the trajectories of charged particles passing through the ID. These reconstructed tracks are a critical component of vertex reconstruction, and the degree of bending and direction of the bent tracks at the production vertex provide information about the momentum, charge, and identity of the charged particles that produced them.

3.2 The Calorimeter

The calorimeter is designed to measure the energy of all particles which pass through it by initiating “showers” and fully absorbing their energy. The only particles which cannot be absorbed by the calorimeter are muons and neutrinos, which pass through without showering. The calorimeter is divided into two sub-detectors, the electromagnetic and hadronic calorimeters. Both are “sampling calorimeters”, which means they are comprised of repeated layers of dense absorbing material with “sampling” layers in between. The sampling layers track the location of the shower and record a small fraction of its energy, to which a calibration factor can be applied to infer the full shower energy.

206 **3.2.1 Electromagnetic Calorimeter**

207 The electromagnetic (EM) calorimeter forms the inner calorimeter layer, and is designed to fully
208 absorb and measure the energies of electrons and photons. Energy is deposited in the lead absorbing
209 layers in the form of EM showers [21], in which the initial electron or photon interacts with
210 the absorbing material to produce a cascade of photon radiation and electron pair production. The
211 sampling layers are filled with liquid argon (LAr), with sensors to measure the ionization produced
212 by charged particles passing through the LAr [22].

213 **3.2.2 Hadronic Calorimeter**

214 The hadronic calorimeter surrounds the EM calorimeter, and is designed to fully absorb and measure
215 hadronic showers, also known as “jets”, initiated by hadrons, which can make it through the
216 EM calorimeter due to their relatively long interaction length [20]. Unlike the EM showers de-
217 scribed above which proceed exclusively via electromagnetic interactions, hadronic showers pro-
218 ceed via both the strong and EM interactions, and as a result are in general more variable and less
219 localized.

220 The hadronic calorimeter is comprised of a tile calorimeter which encircles the EM calorimeter
221 barrel, and a LAr calorimeter with copper and tungsten absorbers in the end-cap region which
222 encloses the two ends of the barrel. The tile calorimeter uses steel as the absorber material and
223 scintillators read out by photomultiplier tubes (PMTs) in the sampling layers.

224 **3.3 The Muon Spectrometer**

225 The muon spectrometer [20] surrounding the calorimeter is specialized for tracking muons and
226 measuring their momentum. It employs the same principle used in the inner detector of applying
227 a strong magnetic field and measuring the resulting bent trajectories of the electrically charged
228 muons passing through to infer their momentum.

229 The magnetic field is generated by rectangular superconducting “toroid magnets” arranged
230 azimuthally in radial planes around the beam axis, which set up a toroidal field concentric to the
231 beam axis. In the region containing the strong field established by the toroid magnets, muon tracks
232 are recorded by three cylindrical layers of muon tracking chambers in the barrel region and three
233 layers of chambers arranged in wheels perpendicular to the beam axis in the end-cap region.

234 **4 Monte Carlo Simulation of Signal and Background**

235 To search for evidence of new physics in the ATLAS data, it is necessary to model the number
236 of SM “background” events that are expected in the selected data (see Sections 5.1 and 7 for
237 discussions of data selection criteria used for the search), as well as their distributions in the many
238 observables, such as E_T^{miss} , that can be computed from the raw data. Only the SM processes which
239 are expected to represent statistically significant backgrounds in the selected data are considered in
240 the analysis. The data can then be compared with the SM background model to search for evidence
241 of an excess of events above the SM background which may be consistent with a hypothesized
242 signal model.

243 Signal and SM background models are implemented by sophisticated Monte Carlo (MC) sim-
 244 ulations of both the passage of the final-state particles through the ATLAS detector and of the
 245 physical production mechanisms.

246 For a given process, for example the Z+jets background shown in figure 9a, “truth-level” in-
 247 formation for each MC event is first obtained from a random proton-proton collision by simulating
 248 the physical production mechanism for the process. The set of simulated final state particles, along
 249 with their kinematic information, are collectively known as the “truth-level” event. Truth-level
 250 events can subsequently be passed through a simulation of the ATLAS detector to model how they
 251 would actually be measured by the detector at “reconstruction-level”.

252 The kinematic distributions of reconstruction-level events, scaled by the integrated LHC beam
 253 luminosity and production cross section for the process, provide a statistical model of how the
 254 given process would appear in the ATLAS data.

255 5 Ongoing Search for the Dark Higgs Model in the Hadronic 256 Decay Channel

257 This section briefly summarizes the “mono-s(WW)” search for the dark higgs model in the fully-
 258 hadronic final state channel, which is currently undergoing review within the ATLAS collaboration.
 259 MC simulated samples were generated for the dark higgs model over the grid of dark higgs and Z’
 260 boson masses shown in figure 7.

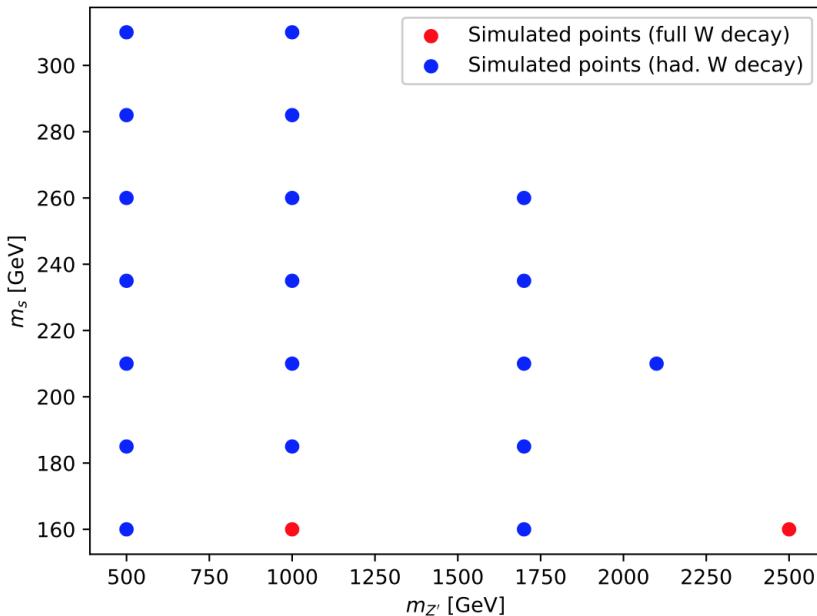


Figure 7: Grid of signal models generated using MC for the hadronic mono-s(WW) analysis at various dark higgs and Z’ boson masses. Red points are simulated with all three possible WW decay channels ($WW \rightarrow qqqq$, $WW \rightarrow qq\ell\nu$ and $WW \rightarrow \ell\nu\ell\nu$) and blue points are simulated exclusively with fully hadronic WW decays ($WW \rightarrow qqqq$) to save computing time.

5.1 Signal Region Selection

Selection cuts are applied to the data and MC samples in the “signal region” (SR) with the aim of maximizing the sensitivity of the data to the dark higgs signal of interest in the fully hadronic decay channel. To avoid biasing the selection based on features in the data, the SR is ‘blinded’, which means that the cuts are developed and optimized using the signal and background MC alone.

5.1.1 Pre-selection

Prior to cut optimization, loose pre-selection cuts are applied to broadly capture the region of interest for the search. As these cuts are not expected to change during subsequent optimization, they can be applied during an initial stage of data skimming to produce relatively small data files, with which further studies and optimization can be performed with reasonably fast data transfers and processing times. The pre-selection cuts used by the hadronic analysis are listed in the first column of Table 1, which is taken from the internal support note for the hadronic analysis.

The first three cuts and the final cut listed in Table 1 - 0 baseline lepton (ℓ), $E_T^{\text{miss}} > 200 \text{ GeV}$, τ veto and $N(\text{small-}R \text{ jets}) \geq 2$ - define the basic selection for a fully hadronic final state with high E_T^{miss} .

A veto placed on variable radius (VR) track jets tagged as having been produced from b quarks (i.e. “b-tagged”) is designed to reduce the “ $t\bar{t}$ ” background of SM top quark pair production shown in figure 9, as top quarks decay almost exclusively to b quarks. A lower bound of 15 is placed on the object based E_T^{miss} significance S with the aim of removing events with high E_T^{miss} originating from poor detector energy resolution. The E_T^{miss} significance is defined as [23]:

$$S = \frac{E_T^{\text{miss}}}{\sigma_L \sqrt{1 - \rho_{LT}^2}} \quad (2)$$

where σ_L^2 is the measured variance of the E_T^{miss} in the “longitudinal” direction parallel to the \vec{E}_T^{miss} vector, and ρ_{LT} is the correlation factor of momentum measurements in the directions parallel and transverse to the \vec{E}_T^{miss} .

Lastly, a lower bound of 0.35 is placed on the angular separation in the plane transverse to the beam line between each of the leading three small-radius jets and the \vec{E}_T^{miss} to select for the desired topology of DM recoiling against jets. Note that some events may contain only two small-radius jets, in which case the lower bound on $\Delta\phi_{\text{jets}, E_T^{\text{miss}}}$ is applied only to these two jets.

Table 1: The pre-selection which is required in the signal and control regions.

0 ℓ SR	1 μ control region	2 ℓ control region
0 baseline ℓ $E_T^{\text{miss}} > 200 \text{ GeV}$ τ veto	0 baseline e, 1 baseline+signal μ $E_T^{\text{miss,no}\mu} > 200 \text{ GeV}$ τ veto	2 baseline+signal ee / opposite sign $\mu^+\mu^-$ $p_T^{\ell\ell} > 200 \text{ GeV}$ τ veto
0 b-tagged VR track jets in event VR track jet ΔR overlap veto $\Delta\phi_{\text{jets}_{1,2,3} E_T^{\text{miss}}} > 0.35$ $S > 15$	0 b-tagged VR track jets in event VR track jet ΔR overlap veto $\Delta\phi_{\text{jets}_{1,2,3} E_T^{\text{miss}}} > 0.35$ $S > 15$	0 b-tagged VR track jets in event VR track jet ΔR overlap veto -
$N(\text{small-}R \text{ jets}) \geq 2$	$N(\text{small-}R \text{ jets}) \geq 2$	$N(\text{small-}R \text{ jets}) \geq 2$

288 **5.1.2 Kinematic Categories**

289 In addition to the pre-selection cuts described in Section 5.1.1, further selections are optimized sep-
 290 arately in the three kinematic categories illustrated in figure 8, known as “merged”, “intermediate”
 291 and “resolved”.

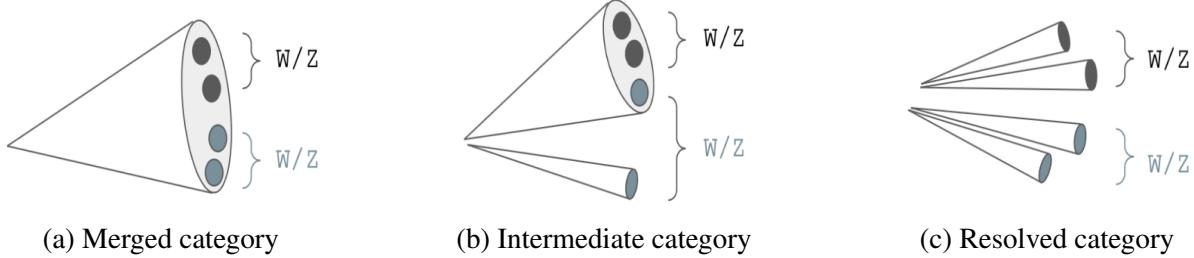


Figure 8: Illustration of the kinematic categories used for event selection optimization in the analysis of the mono-s(WW) hadronic channel. Source: Adapted from a figure shown by P. Gadow in an ATLAS internal approval presentation.

292 The “merged” category represents a regime in which the hadronic final state is sufficiently
 293 boosted in the transverse plane that the quark hadronizations identified as having originated from
 294 the hadronic WW decays cannot be individually resolved, and are instead reconstructed as a sin-
 295 gle large-radius jet. In the “resolved” category, the boost is sufficiently small that all four quark
 296 hadronizations identified as having originated from W decays can be resolved into individual jets.
 297 In between these two extremes is the “intermediate” category, which is characterized by two or
 298 three of the hadronized quarks being reconstructed together as a large-radius (large-R) jet and the
 299 others being resolved into individual jets.

300 **5.2 Control Regions**

301 Control regions (CRs) are designed to provide data-driven normalization constraints for the dom-
 302 inant Z+jets and W+jets SM background processes, shown in figure 9. The CR is chosen by
 303 selecting a region of phase space that is kinematically similar to the SR, and which contains a
 304 high purity of the background process to be normalized. The CR must be orthogonal to the SR to
 305 avoid unblinding any part of the SR. To minimize the possibility of any signal events in the data
 306 affecting the normalization of the SM background process, it is also desirable to select a CR in
 307 which negligible contamination of the MC signal events is observed. Data and MC distributions
 308 are compared in the intermediate kinematic category of the two CRs in figure 10.

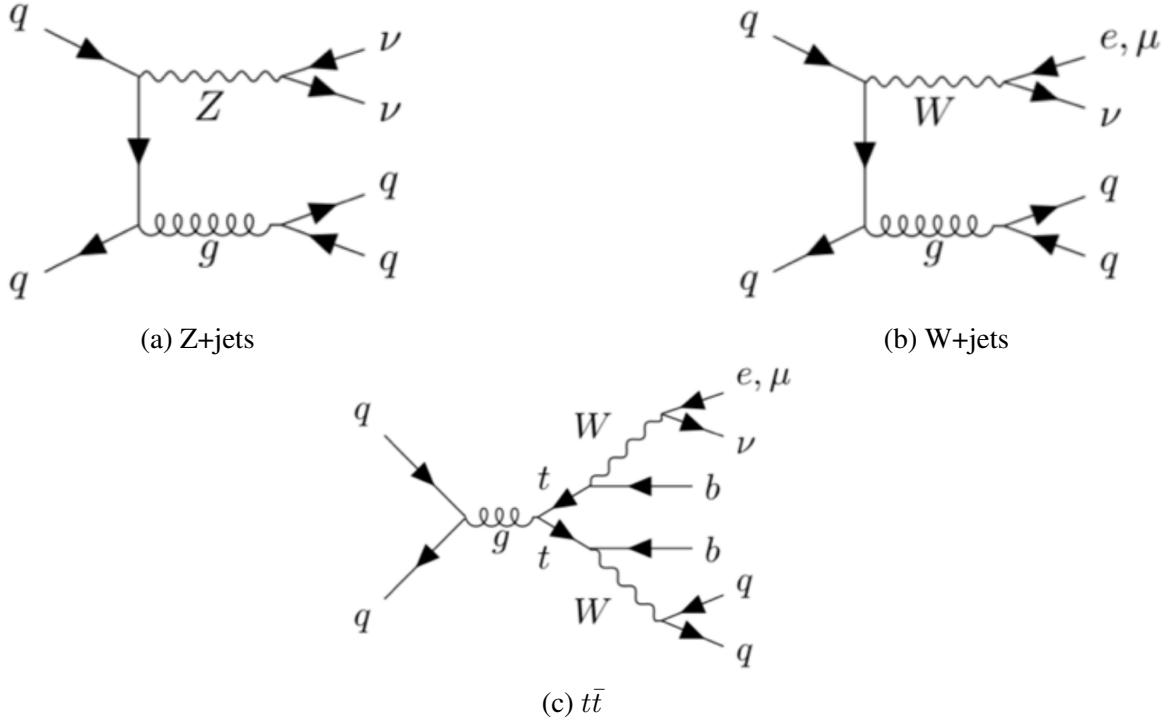


Figure 9: Dominant SM background processes for the mono-s(WW) search. The Z+jets background is only dominant in the hadronic channel, as it is significantly reduced by the 1-lepton requirement in the semileptonic channel.

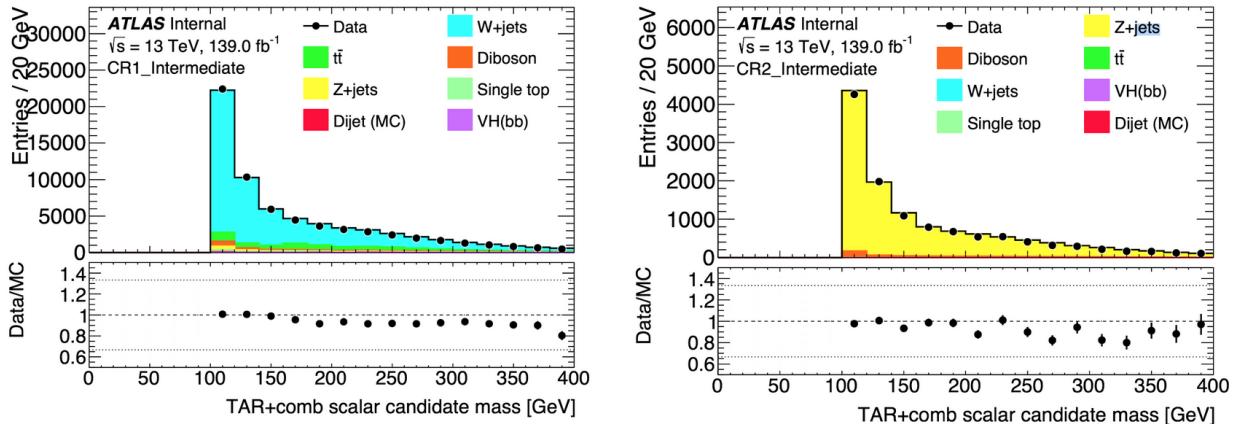


Figure 10: Data-MC comparisons of reconstructed dark higgs candidate mass distributions in the intermediate kinematic category of the 1-lepton (left) and 2-lepton (right) CRs

309 5.2.1 1 Lepton CR

310 The 1 lepton CR is identical to the SR, except that the lepton veto is replaced by a requirement of
 311 one signal muon. This CR is designed to obtain a purified sample of the SM W+jets background
 312 in which the E_T^{miss} arises from a neutrino produced in the leptonic decay of a W boson, and thereby

313 provide a constraint on the normalization of this background process. In the fully hadronic SR, the
314 lepton in this process is not properly reconstructed, and escapes detection.

315 5.2.2 2 Lepton CR

316 The 2 lepton CR is also largely identical to the SR. The primary difference is the replacement
317 of the lepton veto in the SR by a requirement of two signal muons or electrons in the CR. This
318 CR is designed to obtain a purified sample of the Z+jets background in the case where the Z
319 decays leptonically ($Z \rightarrow \ell\ell$). This sample is used to obtain a normalization factor for the Z+jets
320 background in the SR, where the Z decays are predominantly hadronic ($Z \rightarrow qq$), by making use of
321 existing high-precision measurements of the relative production cross sections of these two Z decay
322 channels. In addition to the two-lepton requirement, the E_T^{miss} significance cut is reversed in the 2
323 lepton CR and the $\Delta\phi_{\text{jets}_{1,2,3}, E_T^{\text{miss}}}$ cut is removed to boost the contribution from QCD background.

324 5.3 Systematic Uncertainties

325 In addition to the statistical uncertainty arising from the limited number of events present in the
326 data and simulated in the MC, systematic uncertainties arising from both theoretical and experi-
327 mental sources can affect the normalization and shape of observables used in the analysis. These
328 systematic uncertainties are evaluated as variations in the weights associated with MC generated
329 events. The weight variations are produced by varying MC input parameters from their nominal
330 values by their respective systematic uncertainty.

331 5.3.1 Experimental systematics

332 Experimental systematic uncertainties arise from various aspects of modelling the ATLAS detector
333 and LHC machine, and are provided by the relevant ATLAS performance groups. These include
334 the limited precision of our knowledge of the LHC beam luminosity, uncertainties associated with
335 lepton and muon reconstruction and identification, and jet energy resolution. Uncertainties associ-
336 ated with the reconstruction of the E_T^{miss} are also considered.

337 5.3.2 Theoretical Modelling Uncertainties

338 Theoretical modelling uncertainties associated with MC event generation are evaluated both for
339 signal and for the dominant Z+jets and W+jets SM background processes. Variations arising from
340 the modelling of the parton density functions (PDFs) [24] of the colliding protons are considered,
341 along with the renormalization and factorization scales [25] used to model QCD processes.

342 For the W+jets and Z+jets background processes, which are nominally modelled with the
343 SHERPA generator [26], the variation arising from generating the events with an alternative event
344 generator called PYTHIA 8 [27] is also considered. The alternative generator uses a different al-
345 gorithm to model parton showering, with all other settings in the alternative generator kept the
346 same.

347 **5.4 Sensitivity Estimates**

348 To estimate the sensitivity of the search in the hadronic channel, the data is binned in E_T^{miss} and
 349 dark higgs candidate mass. Mock binned data is generated for each signal point to be statistically
 350 consistent with the simulated SM background. The SM background and signal model are then fit
 351 to this mock background-only data for events passing the event selection criteria. The W+jets and
 352 Z+jets background shapes are obtained from MC, and their normalization factors from the 1-lepton
 353 and 2-lepton CRs, respectively.

354 The fit is used to obtain a CLs value [28] for each signal point, defined as:

$$\text{CLs} = \frac{p_\mu}{1 - p_b} \quad (3)$$

355 where p_μ is the p-value of the background+signal hypothesis and p_b is the p-value of the background-
 356 only hypothesis.

357 The CLs can also be expressed as a “significance” by treating the CLs as a p-value and com-
 358 puting the corresponding standard deviation from the null background-only hypothesis. Signal
 359 points with CLs value below 0.05, or significance above 1.96, could be excluded for the hypoth-
 360 esized production cross section with 95% confidence if the data is statistically consistent with the
 361 background.

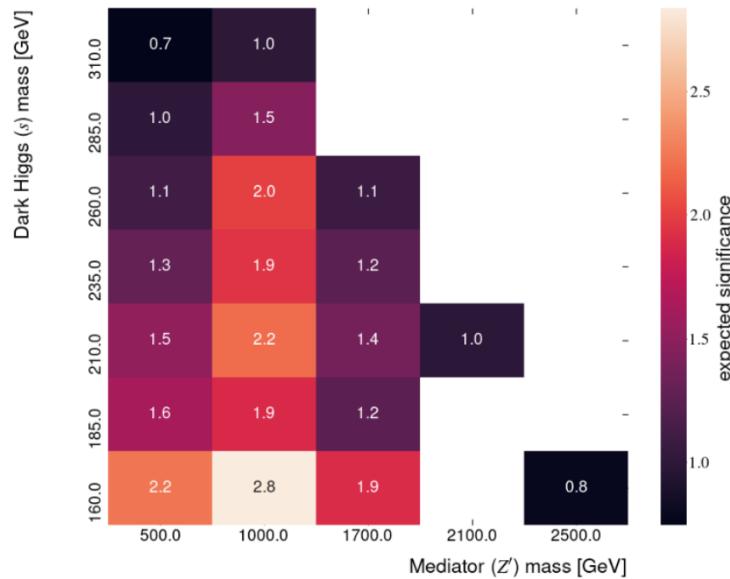


Figure 11: Sensitivity estimate for hadronic channel, expressed in terms of significance. The fit includes all analysis regions and systematic uncertainties.

362 **6 Search for the Dark Higgs Model in the Semileptonic Decay
 363 Channel**

364 The proposed thesis will focus on the search for the dark higgs model in the semileptonic $WW \rightarrow$
 365 $qq\ell\nu$ decay channel. This section summarizes the current status of ongoing analysis work in this

366 channel.

367 6.1 Signal Sample Generation

368 6.1.1 Parameter Choices

369 MC signal samples for the dark higgs model with $s \rightarrow WW$ decay are generated over a range of
370 $m(s)$ and $m(Z')$ for which the analysis is expected to be sensitive to the model. There are several
371 other free parameters in the dark higgs model which need to be fixed to perform the scan. The
372 parameter choices are identical to those used for the re-interpreted mono-H(bb) search described
373 in Section 2.4.1:

- 374 • Coupling g_q of the Z' boson to quarks: 1
- 375 • Coupling g_χ of the Z' boson to DM: 0.25
- 376 • DM mass $m_\chi = 200$ GeV
- 377 • Mixing angle θ between the SM and dark sector higgs bosons: 0.01

378 The values of g_q , g_χ and m_χ , listed above are standard choices used by many other LHC
379 searches. The use of standard parameter choices facilitates comparisons and combinations of
380 search results between different analyses. The value of θ was chosen from the approximate range
381 of 0.01-0.1 suggested by the authors of the phenomenology paper [1] in which the model is intro-
382 duced. It is discussed in the paper that the precise value of the mixing angle is not important for the
383 analysis. It need only be sufficiently large that the dark higgs will decay promptly and sufficiently
384 small as to avoid constraints from existing measurements of the SM Higgs couplings [29].

385 6.1.2 Semileptonic Signal Grid

386 A signal grid has been generated for the mono-s(WW) signal with semileptonic WW decay,
387 presently at the same points that were generated with hadronic decay for the hadronic analysis
388 described in Section 5. There are additionally two signal points generated with all possible WW
389 decay modes (i.e. ‘inclusive’).

390 Work is ongoing to re-generate the grid with a more recent release of the ATLAS simulation
391 software, and extend it as shown in Fig. 12 to allow for a more complete coverage of the phase
392 space region to which the analysis could be sensitive.

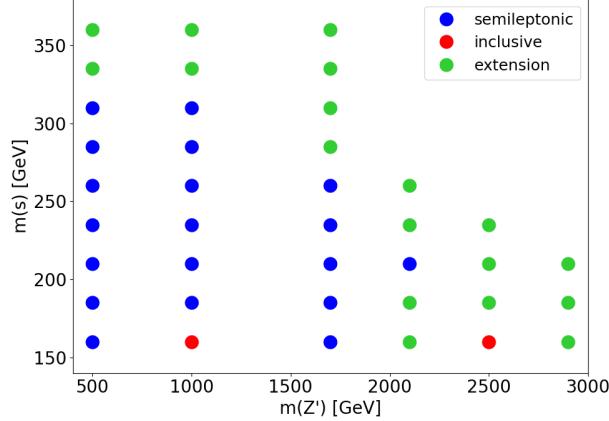


Figure 12: Grid of signal points with respect to dark higgs (s) and Z' boson masses

393 6.2 Event Selection for the Semileptonic Channel

394 This section describes the current event selection used to optimize the sensitivity of the analysis to
 395 the dark higgs signal with semileptonic WW decay. The selection is in the process of being refined,
 396 so many details presented here are preliminary and will likely evolve as the ongoing refinements
 397 progress.

398 6.2.1 Pre-selection

399 As was done for the hadronic analysis, several pre-selection cuts are applied to broadly define the
 400 channel prior to further cut optimization. The current pre-selection cuts are listed below.

- 401 • Event passes E_T^{miss} trigger
- 402 • 1 signal muon or 1 signal electron
- 403 • $E_T^{\text{miss}} > 150 \text{ GeV}$

404 In addition to the current pre-selection cuts, the following two cuts are common to both kine-
 405 matic categories of the signal region (see Section 6.2.2), and may become listed as pre-selection in
 406 the future:

- 407 • Veto on b-tagged jets
- 408 • Lepton- E_T^{miss} transverse mass $m_T(\ell, E_T^{\text{miss}}) > 150 \text{ GeV}$

409 The transverse mass $m_T(E_T^{\text{miss}}, \ell)$ between the lepton and E_T^{miss} is considered in the semilep-
 410 tonic channel, because it is sensitive to the presence of additional E_T^{miss} beyond that arising from
 411 the neutrino in the semileptonic W decay. It is computed in the ATLAS data and MC as:

$$412 m_T(\ell, E_T^{\text{miss}}) = \sqrt{2p_{T,\ell}E_T^{\text{miss}}(1 - \cos\theta_{\ell,E_T^{\text{miss}}})} \quad (4)$$

412 which comes from the full transverse mass definition [30]

$$m_{T, \text{full}}^2(\ell, E_T^{\text{miss}}) = (E_{T,\ell}^2 + E_{T,E_T^{\text{miss}}}^2 - (p_{T,\ell}^2 + p_{T,E_T^{\text{miss}}}^2)) \quad (5)$$

$$= m_\ell^2 + m_{E_T^{\text{miss}}}^2 + 2E_{T,\ell}, E_{T,E_T^{\text{miss}}} (1 - \cos \theta_{\ell, E_T^{\text{miss}}})$$

under the assumptions that the masses associated with the lepton and E_T^{miss} are negligibly small compared with their momenta. The assumption of negligible lepton mass is in general justified given the energy of LHC collisions. The assumption of negligible mass associated with E_T^{miss} is justified if the true E_T^{miss} arises only from the neutrino in the semileptonic W decay, as it would in the leading SM backgrounds. In the signal model, however, there is additional mass associated with the E_T^{miss} arising from DM production. The result, shown in figure 13 after applying the pre-selection cuts, is that the bulk of the SM background has $m_T(\ell, E_T^{\text{miss}})$ below the W mass peak, but the signal distribution tends to be peaked closer to ~ 250 GeV.

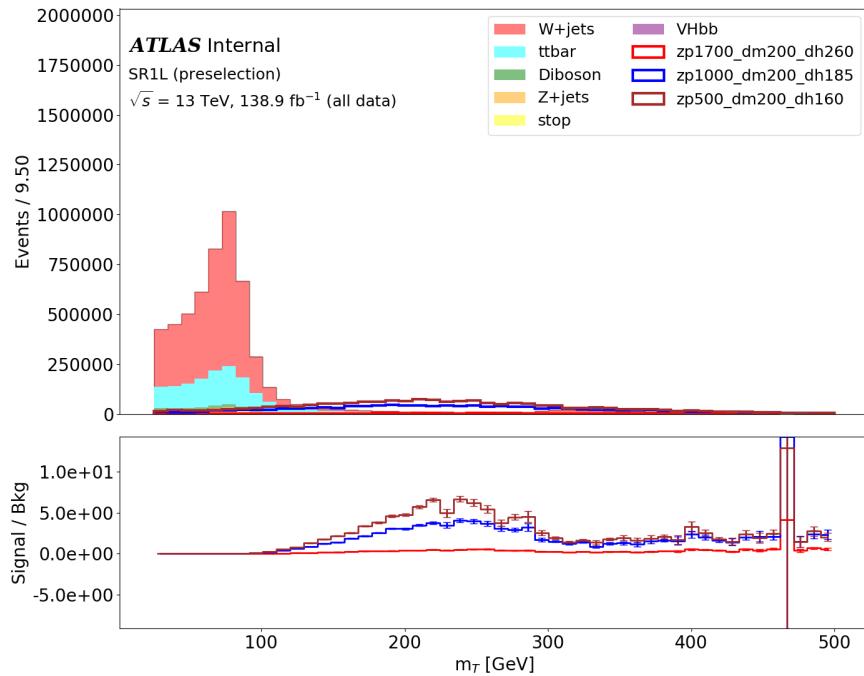


Figure 13: Transverse mass distribution for SM background and several signal points with pre-selection cuts

6.2.2 Kinematic Categories

Further cuts are optimized within the two kinematic categories illustrated in figure 14. As in the hadronic analysis, the kinematic category definitions aim to characterize the degree to which the leading hadronic jets in the event are boosted. The “merged” category aims to capture the highly boosted regime, in which the final state is sufficiently boosted in the transverse plane that the leading hadronic showers are reconstructed as a single large-radius jet. The “resolved” category captures the regime in which all hadron showers have a low enough boost as to be individually reconstructed as small-radius jets. In both categories, there is a possibility of overlap between

429 the leptonic and hadronic decay products. Studies are ongoing to improve the algorithm used for
 430 disentangling the final-state lepton from jets for events in which such overlap occurs - see Section
 431 8.2 for details.

432 The intermediate category used in the hadronic analysis, characterized by a combination of
 433 large- and small-radius jets, is not considered for the semileptonic channel because the semilep-
 434 tonic WW decay in the signal model produces only two final-state quarks.

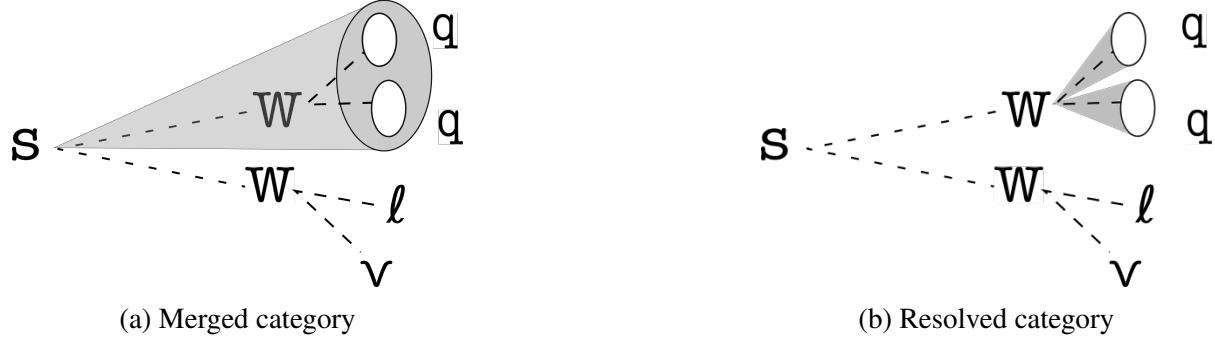


Figure 14: Kinematic categories for the semileptonic WW decay channel

435 The current merged and resolved category cut definitions, listed and discussed in tables 2 and
 436 3 respectively, have been developed and optimized by two other members of the analysis team.
 437 The E_T^{miss} distributions for each category are shown in figure 15. There is ongoing work within the
 438 analysis team to investigate the benefit of tuning the selection further by optimizing the selections
 439 within the $WW \rightarrow qqe\nu$ and $WW \rightarrow qq\mu\nu$ decay channels separately.

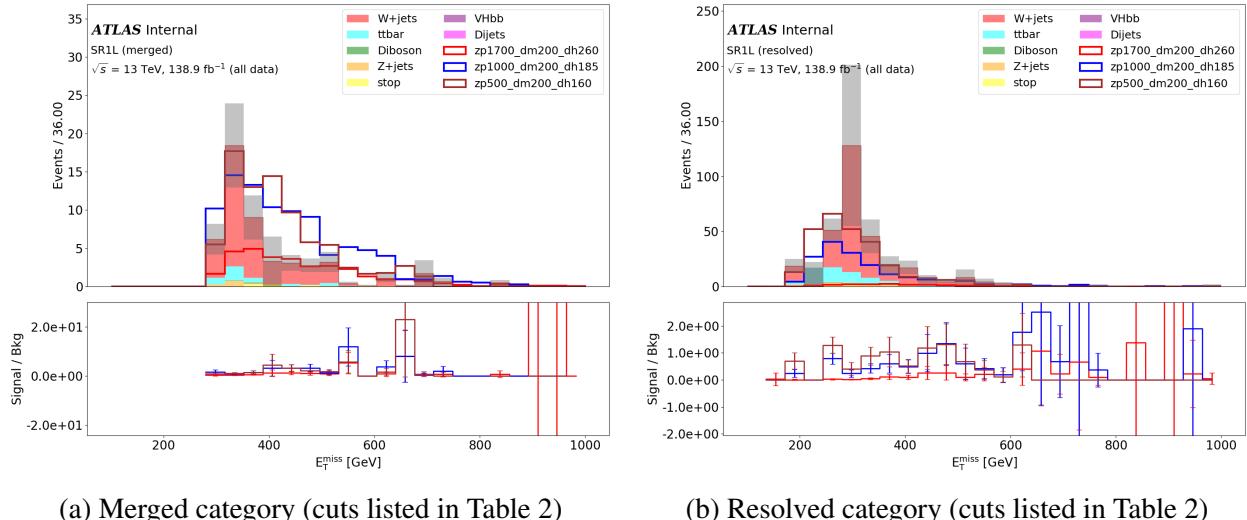


Figure 15: E_T^{miss} distributions in the SR in the two kinematic categories developed for the semileptonic WW decay channel. Signals are shown with amplitudes multiplied by a factor of 2. Grey bands represent statistical uncertainty.

Table 2: Current data selection cuts in the **merged** category of the semileptonic decay channel

Selection	Discussion
Pre-selection	Listed in Section 6.2.1
At least 1 large-radius jet	-
$E_T^{\text{miss}} > 300 \text{ GeV}$	-
Reconstructed mass of large-radius jet $\in [60 \text{ GeV}, 100 \text{ GeV}]$	Increases the likelihood that the selected large-R jet originates from a W boson.
E_T^{miss} significance > 15	-
$\Delta R(\ell, \text{large-R jets}) < 1.6$	$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ is the angular distance between the lepton and the large-radius jet, where $\Delta\phi$ is the angular separation in the transverse plane and pseudorapidity η is a quantity related to the angle θ of a particle relative to the beam axis: $\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]$. Pseudorapidity is used rather than θ because differences $\Delta\eta$ in pseudorapidity are invariant under Lorentz boosts. An upper bound is placed on the angular distance to select for the boosted topology of interest, in which all decay products of the dark higgs boson are boosted in approximately the same direction.
D_2 of leading large-R jet < 1.4	The D_2 variable [31] uses energy correlation functions to give a measure of the likelihood that a jet contains only two spatially separated energy sources. This allows the semileptonic signal channel - in which only two dominant energy sources (i.e. “prongs”) would be expected due to the $W \rightarrow qq$ decay - to be separated from the QCD background which may contain many such prongs.

Table 3: Current data selection cuts in the **resolved** category of the semileptonic decay channel

Selection	Discussion
Pre-selection	Listed in Section 6.2.1
At least 2 small-radius jets	-
Reconstructed mass of the W candidate $\in [60 \text{ GeV}, 100 \text{ GeV}]$	The W candidate is selected by considering all combinations of small-radius jets, and identifying the pair whose reconstructed mass is closest to the W boson mass of 80.4 GeV.
E_T^{miss} significance > 13	-
$m_T(\ell, E_T^{\text{miss}}) > 200 \text{ GeV}$	-
W candidate $p_T > 100 \text{ GeV}$	The lower bound on the W candidate p_T increases the likelihood that the two jets used to reconstruct the W boson actually originated from the decay of a boosted parent particle.
$\Delta R(W, \ell) < 1$	An upper bound on the angular distance between the W boson candidate and the lepton selects for the topology of interest in which one of the boosted W bosons decays leptonically, such that in the lab frame, the lepton is expected to be measured in approximately the same direction as the reconstructed W .

440 **6.2.3 W+jets Control Region**

441 The W+jets CR is defined with the aim of providing a data-driven estimate of the normalization of
 442 the W+jets background, illustrated in figure 9b, in the SR. The exact definition of this CR is still
 443 being studied and optimized, but at its simplest, it would reverse only the transverse mass cut:

$$m_T(\ell, E_T^{\text{miss}}) > 150 \text{ GeV} \rightarrow 50 \text{ GeV} \leq m_T(\ell, E_T^{\text{miss}}) \leq 150 \text{ GeV}$$

444 This reversal makes the W+jets CR orthogonal to the SR, and aims to boost the statistics and
 445 purity of the W+jets background without drastically changing the kinematics of the region. Table
 446 shows that the $m_T(\ell, E_T^{\text{miss}})$ cut reversal improves the W+jets purity with very low signal con-
 447 tamination, but only improves the W+jets statistics in the resolved category. The signal fraction in
 448 table 4 is defined as:

$$\text{signal fraction} = 100\% \times \frac{\sum \text{signal events}}{\sum \text{background events}} \quad (6)$$

Table 4: Comparison of W+jets CR performance measures between SR and preliminary W+jets CR. Signal fraction is averaged over all signal points, and uncertainty is the standard deviation among all points.

Region	W+jets Statistics	W+jets Purity	Signal Fraction
Merged SR	59.1	65.2%	$31\% \pm 11\%$
Resolved SR	337	56.8%	$13\% \pm 12\%$
Merged W+jets CR	53.8	78.5%	$0.49\% \pm 0.30\%$
Resolved W+jets CR	704	83.4%	$0.082\% \pm 0.080\%$

449 Additional cut reversals are being considered to further boost the statistics and purity of the
 450 W+jets background in this CR. However, careful study is needed to determine whether the region
 451 remains sufficiently representative of the SR with these additional cut reversals to provide a reliable
 452 normalization factor for the W+jets background in the SR.

453 **6.2.4 $t\bar{t}$ Control Region**

454 An additional CR is being considered, but has yet to be studied, to constrain the second-to-most
 455 dominant $t\bar{t}$ background. At its simplest, this CR would simply reverse the b-jet veto, which is
 456 primarily intended to reduce the $t\bar{t}$ background in the SR:

$$N(\text{b-tagged jets}) < 1 \rightarrow N(\text{b-tagged jets}) \geq 1$$

457 **7 Sensitivity Studies for the Semileptonic channel**

458 Preliminary sensitivity studies have been performed for the semileptonic decay channel using the
 459 HistFitter statistical analysis framework [32]. The approach is analogous to that described in Sec-
 460 tion 5.4 for the hadronic channel sensitivity, where the data is binned in several variables, and the

$150 < E_T^{\text{miss}} < 300$	$300 < E_T^{\text{miss}} < 450$	$E_T^{\text{miss}} > 450$
---------------------------------	---------------------------------	---------------------------

Table 5: Binning in E_T^{miss} [GeV] used for the resolved category

$\Delta R(\ell, j)$	< 0.5	< 0.5	≥ 0.5 and < 1.6	≥ 0.5 and < 1.6
\mathcal{S}	> 15 and ≤ 22.5	> 22.5	> 15 and ≤ 22.5	> 22.5

Table 6: Binning in ΔR and \mathcal{S} used for the merged category

461 binned signal+background MC is compared with mock data generated to be consistent with the
 462 MC-simulated SM background. The resulting CLs significances representing the exclusion power
 463 of each signal point can be compared between the hadronic and semileptonic WW decay channels
 464 to obtain a rough comparison of their respective sensitivities.

465 7.1 Fit Setup

466 This section describes the details of the fit setup for obtaining exclusion limits in the semileptonic
 467 channel. The fit has been designed such that it can be performed either within the merged and
 468 resolved categories separately, or using the full SR in which the two categories are combined.

469 7.1.1 Binning

470 Currently, the resolved and merged categories are binned in different variables for the fit. The
 471 resolved category uses the three E_T^{miss} bins shown in table 5 and the merged category uses four 2D
 472 bins in $\Delta R(\ell, j)$ and E_T^{miss} significance (\mathcal{S}) shown in table 6. These binning choices may be further
 473 optimized and/or amalgamated in the future.

474 7.1.2 W+jets Background Normalization

475 As with the hadronic fit, the shapes of all backgrounds are fully determined from the MC. A data-
 476 driven normalization factor for the W+jets background in the SR is determined by matching the
 477 statistics of MC and mock data in the preliminary W+jets CR described in Section 6.2.3, which is
 478 currently unbinned. All other backgrounds are currently normalized by scaling the MC amplitudes
 479 to the integrated luminosity.

480 7.1.3 Systematic Uncertainty

481 As the systematic uncertainties on signal and background have not yet been evaluated for the
 482 semileptonic channel, rough estimated systematics of 30% on overall background and 15% on
 483 signal are applied based on the final reported impacts of the systematics considered for the hadronic
 484 channel.

485 **7.2 Preliminary Sensitivity Estimates**

486 **7.2.1 Category-Specific**

487 Figure 16 shows the expected significance as a function of m_S and $m_{Z'}$ for the separate fits to the
 488 resolved and merged categories. Interpolation is performed with respect to m_S and $m_{Z'}$ to obtain
 489 both the colour map and the “exclusion contour” along $CLs = 0.05$. Based on these expected limits,
 490 it is anticipated that the sensitivity of the analysis will be driven largely by the merged category,
 491 particularly in the high- m_S region.

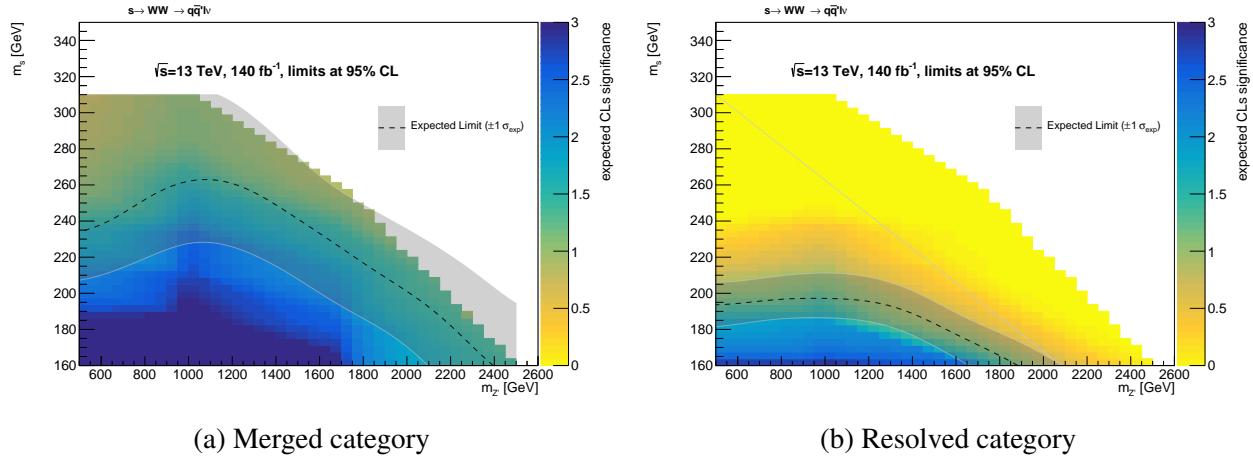


Figure 16: CLs significance as a function of m_S and $m_{Z'}$ within the individual categories

492 **7.2.2 Combined SR**

493 Figure 17 shows the sensitivity estimate for the combined merged+resolved SR, both with and
 494 without the signal grid interpolation and CLs contour. The significance grid on the left is designed
 495 to match the format in which the sensitiviy estimates are presented for the hadronic channel in
 496 figure 11 to allow for a more direct comparison. This preliminary comparison suggests that the
 497 semileptonic decay channel could offer comparable or even superior sensitivity compared with the
 498 hadronic channel.

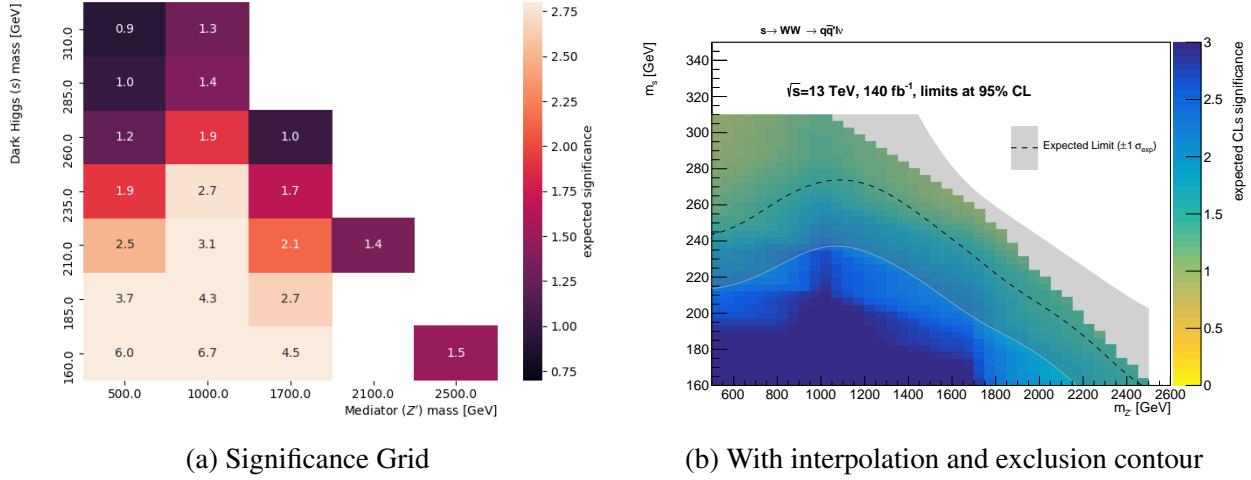


Figure 17: CLs significances for the combined SR

499 8 Remaining Work and Outlook

500 The analysis of the dark higgs signal model in the semileptonic WW decay channel continues
 501 to be developed and optimized. Preliminary sensitivity estimates suggest that this decay channel
 502 could offer valuable sensitivity to the model in addition to the hadronic channel currently nearing
 503 completion. The proposed thesis work will involve developing the analysis in the semileptonic
 504 channel in close collaboration with several colleagues here at UVic and at LMU in Munich.

505 In addition to leading the UVic effort on this analysis, I share the role of “analysis contact”
 506 with a colleague at LMU. The primary purpose of the analysis contact role is to liaise on behalf of
 507 the analysis group with conveners and other members of the wider analysis groups in ATLAS to
 508 which the analysis belongs.

509 The major remaining tasks anticipated to complete the analysis are listed below in approximate
 510 order of priority. Table 7 summarizes the approximate timeline anticipated for completion of the
 511 remaining work, subsequent stages of approval within the ATLAS collaboration and publication
 512 of results.

513 8.1 Electron-Muon Differences

514 The algorithm presently applied to remove overlap between leptons and large-R jets appears to
 515 cause significant differences in jet reconstruction between semileptonic decays with a muon (i.e.
 516 ‘muon channel’) vs. an electron (i.e. ‘electron channel’) in the final state. This results in systematic
 517 differences in various large-R jet observables between the two channels, which cause substantially
 518 more events in the muon channel to pass the existing selections than in the electron channel. This
 519 asymmetry between the lepton and muon channels is both concerning from a physical perspective
 520 - events belonging to the lepton and muon channels should in principle produce statistically con-
 521 sistent distributions of observables with the exception of the observed lepton flavour - and it also
 522 reduces the statistics available to the analysis in the electron channel.

523 Two approaches are being pursued to address these differences in large-R jet observables be-
524 tween the two lepton channels.

525 The first approach is to move from the current “LCTopo” jet reconstruction algorithm to track-
526 assisted reclustered (TAR) jets [33]. This will allow for the implementation of a more sophisticated
527 jet-lepton overlap removal algorithm currently being developed and studied by another physics
528 analysis, which uses tracking information to remove tracks associated with the lepton, as well
529 as any small-radius sub-jets overlapping with the lepton jet, then re-builds the TAR jet with the
530 remaining small-radius sub-jets.

531 The second approach is to split the electron and muon channels for cut optimization, and sepa-
532 rately tune any cuts to which the two channels exhibit different responses. If this approach is taken
533 in conjunction with the first approach, any lingering differences in TAR jet observables between
534 the two channels with the improved overlap removal algorithm can be flexibly accommodated by
535 tuning cuts associated with these observables separately to minimize any associated loss of signal
536 statistics.

537 **8.2 Signal Discrimination Variables**

538 Studies are ongoing to identify or develop variables which could offer useful discrimination be-
539 tween signal and background in the semileptonic SR. These include jet substructure variables to
540 help distinguish large-R jets produced by a boosted $W \rightarrow qq$ decay from eg. the QCD quark
541 pair production in the W+jets background, as well as approaches to computing a reconstructed
542 dark higgs mass that account for the presence of neutrino-induced E_T^{miss} . Once these studies are
543 complete, a set of selection variables will be chosen which exhibit the best discrimination between
544 signal and background. The expected signal significances will be evaluated for various ranges of
545 cut combinations on the chosen variables within each SR (eg. the merged category of the muon
546 channel) to determine the optimal cut placements. The expected signal significance is defined as
547 [34]:

$$Z = \sqrt{2 \left[\ln \left(\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right) - \frac{b^2}{\sigma_b^2} \ln \left(1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right) \right]} \quad (7)$$

548 where s is the expected number of signal events, b is the expected number of background events
549 and σ_b is the statistical uncertainty on the expected number of background events.

550 **8.3 Optimize the W+jets and $t\bar{t}$ CRs**

551 Once the signal selection is tuned and finalized, the CRs can be studied, and the effects of any cut
552 reversals in addition to the basic definitions presented in Sections 6.2.3 and 6.2.4 can be studied
553 to optimize the performance of the CRs. The performance of the CR is ultimately determined
554 by its ability to isolate the background process of interest with sufficiently high statistics, low
555 signal contamination and kinematic similarity to the SR to offer a reliable scaling factor for the
556 background of interest in the SR. Data-MC comparisons with the CR-provided scaling factor can
557 be studied in validation regions chosen explicitly to have kinematics similar to the SR to test the
558 validity of the assumption that the scale factor determined in the CR is applicable to the SR.

559 **8.4 Experimental and Modeling Systematics**

560 Once the SR and CR selections are optimized, experimental and modeling systematics will be
561 added to the MC data in the form of associated variations on the nominal event weights, as de-
562 scribed in Section 5.3. Many of the same systematics applied in the hadronic analysis and de-
563 scribed in Sections 5.3.1 and 5.3.2 will likely be applied in the semileptonic channel, but the sizes
564 and relative impacts of some systematics will differ.

565 **8.5 Finalize the Fit with Full Systematics**

566 With the analysis regions finalized, the fit described in Section 7 will be updated with the final
567 region definitions. The systematics will be added to the fit as uncertainties on the expectation
568 value of their associated nuisance parameters in the likelihood function [32] used for the fit.

569 **8.6 Unblinding and Publication**

570 When the analysis cuts are finalized and the fitting framework is in place, the analysts will request
571 approval from the conveners of its “Exotics” and “JDM” (jets+DM) physics group and sub-group,
572 respectively, to “unblind” the SR. This means that rather than fitting to mock data generated to be
573 statistically consistent with the MC in the SR, the fit will be performed with the actual ATLAS
574 data in the SR to search for an above-background excess. If no excess is found, the observed CLs
575 values can be used to set final exclusion limits based on the data.

576 Following unblinding, the aim is to publish the results as an open-access article in a physics
577 journal.

Table 7: Tentative Timeline for Remaining Work

Goal	Target Date	Discussion
Finalize variables and regions for selection	September 2020	There is nothing blocking progress to this end, but time is required to develop and study new variables and finalize the choices of analysis regions within which to optimize.
Finalize optimized SR and CR selections	November 2020	We are potentially dependent upon the timeline of ongoing work in another analysis towards developing an improved lepton-jet overlap removal (OR) algorithm using tracking information. While a preliminary version exists and has been implemented in our analysis code, final selection optimization cannot take place until the improved OR algorithm is implemented, if we wish to make use of it. The stated timeline for the improved algorithm is several months.
Add experimental and modeling systematics to MC and fit	December 2021s	We will likely be able to make use of most or all of the existing infrastructure in the code framework used by the hadronic channel for adding the systematic variations to the MC. However, the technical details of implementing them in the fit will differ due to the different fitting frameworks used by the two analyses.
Request editorial board	January 2021	A request is made for an editorial board (EB) when the analysis strategy is in place and documented in a supporting note. The EB is usually formed 1-1.5 weeks after a request is made.
Request subgroup approval	February 2021	A request for subgroup approval typically takes place \sim 4 weeks after obtaining an EB.
Unblind the SR	April 2021	Once the analysis is approved by the subgroup, the SR can be unblinded. It typically takes \sim 9 weeks to obtain subgroup approval after it is requested, as modifications may need to be made to the analysis strategy in response to feedback received during the approval process.
Request group approval of analysis and paper draft	June 2021	Group approval is typically requested \sim 7 weeks after the analysis obtains subgroup approval. The group approves both the analysis and the paper draft in one step.
Circulate the paper draft to the ATLAS collaboration	August 2021	It typically takes \sim 6 weeks to receive group approval. Once received, the paper is circulated to the entire ATLAS collaboration for comment.
Publish results	Fall 2021	The anticipated timeline for publication is very approximate, and may depend on the details of the results to be included.

578 References

- 579 [1] Michael Duerr, Alexander Grohsjean, Felix Kahlhoefer, Bjoern Penning, Kai Schmidt-
 580 Hoberg, and Christian Schwanenberger. Hunting the dark higgs. *Journal of High Energy*
 581 *Physics*, 2017(4), Apr 2017.
- 582 [2] Jacob J. Ethier and Emanuele R. Nocera. Parton distributions in nucleons and nuclei. *Annual*
 583 *Review of Nuclear and Particle Science*, 70(1):null, 2020.
- 584 [3] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018) and 2019 update.

- 585 [4] N. et al. Aghanim. Planck 2018 results. VI. Cosmological parameters. 7 2018.
- 586 [5] J. L. Feng. Dark matter candidates from particle physics and methods of detection. *Annual*
587 *Review of Astronomy and Astrophysics*, 1(48):495–545, 2010.
- 588 [6] Marc Schumann. Direct detection of WIMP dark matter: concepts and status. *Journal of*
589 *Physics G: Nuclear and Particle Physics*, 46(10):103003, aug 2019.
- 590 [7] Teresa Marrodan Undagoitia and Ludwig Rauch. Dark matter direct-detection experiments.
591 *J. Phys. G*, 43(1):013001, 2016.
- 592 [8] Marco Cirelli. Indirect searches for dark matter. *Pramana*, 79(5):1021–1043, Oct 2012.
- 593 [9] Jan Conrad and Olaf Reimer. Indirect dark matter searches in gamma- and cosmic rays.
594 *Nature Physics*, 13:224–231, 03 2017.
- 595 [10] Antonio Boveia and Caterina Doglioni. Dark matter searches at colliders. *Annual Review of*
596 *Nuclear and Particle Science*, 68(1):429–459, 2018.
- 597 [11] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Phys.*
598 *Rept.*, 267:195–373, 1996.
- 599 [12] O. Buchmueller, Matthew J. Dolan, and Christopher McCabe. Beyond Effective Field Theory
600 for Dark Matter Searches at the LHC. *JHEP*, 01:025, 2014.
- 601 [13] Daniel Abercrombie, Nural Akchurin, Ece Akilli, Juan Alcaraz Maestre, Brandon Allen, Bar-
602 Barbara Alvarez Gonzalez, Jeremy Andrea, Alexandre Arbey, Georges Azuelos, Patrizia Azzi,
603 and et al. Dark matter benchmark models for early lhc run-2 searches: Report of the atlas/cms
604 dark matter forum. *Physics of the Dark Universe*, 27:100371, Jan 2020.
- 605 [14] Csaba Csaki. The minimal supersymmetric standard model (mssm). *Modern Physics Letters*
606 A, 11(08):599–613, Mar 1996.
- 607 [15] Stefano Profumo, Leonardo Giani, and Oliver F. Piattella. An introduction to particle dark
608 matter, 2019.
- 609 [16] Search for Dark Matter Produced in Association with a Higgs Boson decaying to $b\bar{b}$ at $\sqrt{s} =$
610 13 TeV with the ATLAS Detector using 79.8 fb^{-1} of proton-proton collision data. Technical
611 Report ATLAS-CONF-2018-039, CERN, Geneva, Jul 2018.
- 612 [17] RECAST framework reinterpretation of an ATLAS Dark Matter Search constraining a model
613 of a dark Higgs boson decaying to two b -quarks. Technical Report ATL-PHYS-PUB-2019-
614 032, CERN, Geneva, Aug 2019.
- 615 [18] Kyle Cranmer and Itay Yavin. Recast - extending the impact of existing analyses. *Journal of*
616 *High Energy Physics*, 2011(4), Apr 2011.
- 617 [19] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001–
618 S08001, aug 2008.

- 619 [20] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003,
 620 2008.
- 621 [21] Lev Davydovitch Landau, G. Rumer, and Piotr Leonidovich Kapitza. The cascade theory of
 622 electronic showers. *Proceedings of the Royal Society of London. Series A. Mathematical and*
 623 *Physical Sciences*, 166(925):213–228, 1938.
- 624 [22] Huaqiao Zhang and. The ATLAS liquid argon calorimeter: Overview and performance.
 625 *Journal of Physics: Conference Series*, 293:012044, apr 2011.
- 626 [23] Object-based missing transverse momentum significance in the ATLAS detector. Technical
 627 Report ATLAS-CONF-2018-038, CERN, Geneva, Jul 2018.
- 628 [24] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018) and 2019 update.
 629 *18. Structure Functions.*
- 630 [25] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018) and 2019 update.
 631 *9. Quantum Chromodynamics.*
- 632 [26] T Gleisberg, S Hoche, F Krauss, M Schonherr, S Schumann, F Siegert, and J Winter. Event
 633 generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(02):007–007, feb 2009.
- 634 [27] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1.
 635 *Computer Physics Communications*, 178(11):852–867, Jun 2008.
- 636 [28] P.A. Zyla et al. (Particle Data Group), to be published in Prog. Theor. Exp. Phys. 2020,
 637 083C01 (2020). *40. Statistics.*
- 638 [29] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, O. S. AbouZeid, N. L.
 639 Abraham, H. Abramowicz, and et al. Measurements of the higgs boson production and decay
 640 rates and constraints on its couplings from a combined atlas and cms analysis of the lhc pp
 641 collision data at $\sqrt{s} = 7$ and 8 tev. *Journal of High Energy Physics*, 2016(8), Aug 2016.
- 642 [30] P.A. Zyla et al. (Particle Data Group), to be published in Prog. Theor. Exp. Phys. 2020,
 643 083C01 (2020). *47. Kinematics.*
- 644 [31] Simone Marzani, Gregory Soyez, and Michael Spannowsky. Looking inside jets. *Lecture*
 645 *Notes in Physics*, 2019.
- 646 [32] M. Baak, G. J. Besjes, D. Côté, A. Koutsman, J. Lorenz, and D. Short. HistFitter software
 647 framework for statistical data analysis. *Eur. Phys. J. C*, 75:153, 2015.
- 648 [33] Track assisted techniques for jet substructure. Technical Report ATL-PHYS-PUB-2018-012,
 649 CERN, Geneva, Jul 2018.
- 650 [34] Glen Cowan. Discovery sensitivity for a counting experiment with background uncertainty,
 651 2012.