# Constraining Cosmology and Galaxy Evolution with Highly Multiplexed Spectroscopy

Ruari Mackenzie

A Thesis presented for the degree of
Doctor of Philosophy


Department of Physics
Durham University
United Kingdom

August 2018

# Constraining Cosmology and Galaxy Evolution with Highly Multiplexed Spectroscopy

Ruari Mackenzie

Submitted for the degree of Doctor of Philosophy

August 2018

**Abstract:** The scientific questions addressed in this thesis divide into fundamental cosmology and galaxy evolution, although divergent in theme these are connected by the use of high multiplex spectroscopy. Specifically, we employ spectroscopic galaxy redshift surveys to probe the connection between galaxies and the Cosmic Microwave background, and to link galaxies to neutral gas reservoirs which contain fuel for star-formation. We present the following observational investigations:

- A spectroscopic redshift survey of ∼7000 galaxies in the core of the Cosmic Microwave Cold Spot, with the aim of following up the claim of a supervoid capable of imprinting the Cold Spot via the Integrated Sachs-Wolfe effect. Using 2dF+AAOmega we have surveyed 66 deg$^2$ out to $z < 0.5$, using a $i \leq 19.2$ magnitude limited selection based on imaging from the VST-ATLAS Survey. We confirm the presence of a void at $z = 0.14$ with a scale radius of 119h$^{-1}$ Mpc and a density contrast of $\delta_m = -0.34$, however it is deeper and much less extended than previously claimed. The ΛCDM ISW imprint of this void is only -6.25 $\mu$K, compared to the Cold Spot's central temperature of ∼150 $\mu$K . We detect further under-densities at $z = 0.26$ and 0.30, although these contribute even less to the temperature decrement than the larger $z = 0.14$

void. When combined with previous data we can exclude a large contribution to the Cold Spot from a void at $z < 1$, at higher redshift the ISW effect is less significant. Similarity between our control field and the Cold Spot survey places constraints on any non-standard attempt to link the voids to the CMB Cold Spot.

- An integral field survey of six $z \simeq 3.5$ Damped Lyman $\alpha$ systems with MUSE, to search for host galaxies and to study their environments. Our survey did not preselect DLAs based on absorption properties, on the contrary of current leading DLA host searches that select only high metallicity systems. MUSE has revealed that three of the six fields contain bright Lyman $\alpha$ emitters (LAEs) with a few hundred km $s^{-1}$ of the absorption redshift, detecting associated galaxies out to larger impact parameters than was previously possible. In one field (J0255+0048) we discovered an extended Ly$\alpha$ nebula in close proximity to the DLA, extending up to 50 physical kpc from the DLA location. The object has a striking correspondence between absorption and emission components, which may indicate the structure is an ongoing merger with tidal debris at the DLA location. Our most metal-poor system (J1220+0921) was revealed with MUSE to be embedded in a dense filament-like structure, which extends across the full MUSE field of view (380 kpc). We have compared the distribution of galaxies around the DLAs to models and simulations, using the EAGLE hydrodynamical simulation and a model based on the GALICS semi-analytic prescription. Using these mock data we have shown our observations are consistent with both EAGLE and simple halo prescription put forward to explain the strong clustering of DLAs. Furthermore we show how an expanded version of our survey may provide useful constraints on the small scale clustering of DLAs and the link between neutral gas reservoirs and star-formation.

# Declaration

The work in this thesis is based on research carried out at by the author at the Centre for Extragalactic Astronomy, Department of Physics, Durham University, UK, under the supervision of Prof. Tom Shanks and Dr Michele Fumagalli. No part of this thesis has been submitted elsewhere for any other degree or qualification.

The results and analysis presented in chapter 2 have been published as the following article:

- *Evidence against a supervoid causing the CMB Cold Spot*, Mackenzie, R.; Shanks, T; Bremer, M. N.; Cai, Y.-C.; Gunawardhana, M. L. P.; Kovács, A.; Norberg, P.; Szapudi, I., 2017, MNRAS, 470, 2328

The author of the thesis was primarily responsible for all aspects of the work.

The results and analysis presented in chapter 3 have been published as the following article:

- *Witnessing galaxy assembly in an extended $z \simeq 3$ structure*, Fumagalli, M.; Mackenzie, R.; Trayford, J.; Theuns, T.; Cantalupo, S.; Christensen, L.; Fynbo, J. P. U.; Møller, P.; O'Meara, J.; Prochaska, J. X.; Rafelski, M.; Shanks, T., 2017, MNRAS, 471, 3686

The author of the thesis was primarily responsible for the data reduction and analysis but did not lead the generation of the text.

Chapter 4 is in preparation for submission to MNRAS, the author was primarily responsible for all aspects of the research and text.

# Acknowledgements

I would like to thank my parents, Hugh and Sandra, for supporting me through many years of education. Thanks to my sister, Lucy, for encouraging me from an early age.

Many thanks to my supervisors, Tom Shanks and Michele Fumagalli. Tom has always been quick to share his wisdom and healthy scepticism. I thank Michele for his endless support and unconditional positivity. I am grateful to my many colleagues in Durham and elsewhere, who have shared their knowledge and time with me over the past four years.

I am deeply grateful to my friends in the Department and beyond, life in Durham would not have been the same.

*Dedicated to*

my father, Alexander Hugh Mackenzie,

*and*

my family and friends.

# Contents

**Abstract** **iii**

**1 Introduction** **1**

1.1 Overview . . . . . 1

1.2 Cosmology . . . . . 2

1.2.1 The Cosmic Microwave Background . . . . . 7

1.2.2 Standard Cosmology: ΛCDM . . . . . 9

1.2.3 The Integrated Sachs-Wolfe Effect . . . . . 11

1.3 Gas and Galaxies . . . . . 13

1.3.1 Galaxy Clustering . . . . . 14

1.3.2 Damped Lyman $\alpha$ Systems . . . . . 16

1.3.3 Numerical Simulations . . . . . 21

1.4 Galaxy Redshift Surveys . . . . . 23

1.4.1 Thesis Outline . . . . . 26

**2 A Redshift Survey of the CMB Cold Spot** **29**

2.1 Introduction . . . . . 30

2.2 Survey and Data Reduction . . . . . 34

2.3 Results . . . . . 37

2.3.1 Redshift Distributions . . . . . 37
2.3.2 Void Model . . . . . 38
2.3.3 Perturbation Fitting in the Cold Spot . . . . . 40
2.3.4 Perturbation Fitting in GAMA G23 . . . . . 43
2.4 Discussion . . . . . 45
2.4.1 The reality of the $z = 0.42$ void . . . . . 46
2.4.2 Photo-$z$ and spectroscopic $n(z)$ . . . . . 49
2.4.3 A coherent SGC galaxy distribution? . . . . . 51
2.4.4 Origin of The CMB Cold Spot . . . . . 54
2.5 Conclusions . . . . . 57

**3 Galaxy assembly at $z \approx 3.25$ in an extended structure 59**

3.1 Introduction . . . . . 60
3.2 Data acquisition and reduction . . . . . 63
3.3 Analysis of the imaging and spectroscopic data . . . . . 64
3.3.1 The physical properties of the $z \approx 3.25$ DLA . . . . . 64
3.3.2 Search for galaxy counterparts . . . . . 68
3.3.3 Properties of the detected galaxies . . . . . 72
3.4 Nature of the extended emission . . . . . 75
3.4.1 Photoionization from galaxy G . . . . . 76
3.4.2 Contribution from *in situ* star formation . . . . . 78
3.5 Comparison with the EAGLE simulations . . . . . 80
3.5.1 The EAGLE simulations and radiative transfer post-processing 83
3.5.2 Statistical analysis of DLAs associated to galaxies in EAGLE 85
3.5.3 Close analogue of the observed system . . . . . 88
3.6 Summary and Conclusions . . . . . 89

**4 A MUSE Integral Field Survey of DLAs at $z \simeq 3.5$ 93**

4.1 Introduction . . . . . 94

4.2 Observations and data reduction . . . . . 99

4.2.1 Sample selection . . . . . 99

4.2.2 MUSE observations and data reduction . . . . . 102

4.2.3 Absorption line spectroscopy . . . . . 103

4.3 Search for Galaxy Associations . . . . . 106

4.3.1 Search for Ly$\alpha$ emitters . . . . . 108

4.3.2 Testing the robustness of LAE identifications . . . . . 110

4.3.3 Identification of continuum objects . . . . . 113

4.4 Properties of the high-confidence associations . . . . . 115

4.4.1 Notes on individual fields . . . . . 115

4.4.2 Continuum counterparts of the LAEs . . . . . 120

4.5 Description of models and simulations . . . . . 120

4.5.1 The EAGLE simulations . . . . . 121

4.5.2 The GALICS semi-analytic model . . . . . 126

4.5.3 A halo prescription for DLAs . . . . . 127

4.6 Discussion . . . . . 133

4.6.1 Detection rates and comparison with previous studies . . . 133

4.6.2 Constraints on the DLA host halo mass . . . . . 137

4.6.3 Forecasts for future searches . . . . . 144

4.7 Summary and Conclusions . . . . . 145

**5 Summary and Conclusions** **147**

5.1 The CMB Cold Spot . . . . . 147

5.2 Gas and Galaxies . . . . . 151

5.3 Future Outlook and Concluding Remarks . . . . . 155

**A Chapter 4 Additional Material** **157**

A.1 Additional tests on the robustness of the LAE sample . . . . . 157

A.2 Notes on the fields with lower significance detections . . . . . 161

A.3 Continuum Spectra of LBGs . . . . . 162

A.4 Properties of the candidate LAEs . . . . . 163

# Chapter 1

# Introduction

## 1.1 Overview

This thesis reviews the results of observational projects to probe two long standing questions in cosmology and galaxy evolution. The scientific topics of these investigations are different, but they are linked by the common application of high multiplex spectroscopy. Specially we will address the following questions:

- Is the Cosmic Microwave Background Cold Spot caused by a supervoid? Is this evidence of a departure from standard cosmology?
- How are neutral gas reservoirs linked to galaxies at high redshift?

In Chapter 1, the background knowledge of the field will be reviewed. The introduction has been divided into themes of galaxy evolution and cosmology, but research chapters will draw from both sections. First we introduce basic relativistic cosmology and demonstrate a key concept, that one can measure distance to remote galaxies via spectroscopy and measuring redshift. We then describe the Cosmic Microwave Background (CMB) which will be relevant to Chapter 2. The CMB underpins the current cosmological model which is outlined next. We introduce the Integrated Sachs-Wolfe effect which links the CMB to galaxies in the modern universe. This

effect is the basis for Chapter 2. In the section we move onto topics relating to galaxy formation and evolution. We first introduce the concept of galaxy clustering, which is a major topic of Chapter 4. We will then detail the history and properties of Damped Lyman $\alpha$ absorption systems which are the subject of Chapters 3 and 4. Chapters 3 and 4 employ numerical simulations to compare observation with theory, and these are briefly reviewed. Lastly we briefly review the observational basis of highly multiplexed spectroscopy.

## 1.2 Cosmology

Physical cosmology is the study of the structure, origin and evolution of the Universe. The current understanding of cosmology is the foundation on which extragalactic astronomy is based. It can be argued that physical cosmology emerged as an observational science in the early $20^{\text{th}}$ century when the first spectra of spiral nebulae were obtained. Slipher (1913) was the first to observe a shift of the spectral lines measured from the Andromeda Nebula (M31) indicating a velocity towards the observer of 300 km $\text{s}^{-1}$, measured spectroscopically via the Doppler effect. A sample of spiral nebulae revealed that Andromeda was an outlier, most observed objects had large recession velocities of hundreds of km $\text{s}^{-1}$ (Slipher, 1915). At this time it was controversial as to whether the spiral nebulae, now known as galaxies, were beyond the Milky Way or part of its extent. In the 1920's the distances to the brightest spiral nebulae were measured (Hubble, 1925), using Cepheid Variable stars to measure absolute distance via the Period-Luminosity relation (Leavitt & Pickering, 1912). Hubble (1929) combined Cepheid measurements with radial velocities to present the first convincing evidence of a relationship between recession velocity ($v_r$) and distance ($D$), now known as Hubble's Law (Eqn. 1.2.1).

$$v_r = \text{H}_0 \times D \tag{1.2.1}$$

$\text{H}_0$ is known as the Hubble Constant, its value has long been controversial but is

accepted to be around 70 km $s^{-1}$ $Mpc^{-1}$ (Planck Collaboration et al. 2016b; Riess et al. 2016, also written as 100 $h$ km $s^{-1}$ $Mpc^{-1}$). Hubble's Law is now understood in the context of a universe undergoing metric expansion, such a cosmology is described by the Friedmann–Lemaître–Robertson–Walker (FLRW) metric (Eqn. 1.2.2). The FLRW metric is a solution to General Relativity (GR), the implication of which was derived independently before Hubble's 1929 paper (Friedmann, 1922; Lemaître, 1927) and is given by

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a(t)^2\Big[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2\Big], \tag{1.2.2}$$

where $c$ is the speed of light, $r$, $\theta$ and $\phi$ are co-moving spherical coordinates, $ds$ is the line element, $k$ is the curvature and $a(t)$ is the scale factor. The FLRW metric describes a homogeneous, isotopic universe in accordance with the Cosmological Principle, but allows for a non-static universe. The scale factor describes the expansion history of the universe, the present day value is defined as $a(t_0) = 1$. Curvature describes to the geometry of the universe and it depends on the density and Hubble constant.

If one considers a small separation ($kr^2 << 1$) the geometry is Euclidean and the radial distance to a position $r$ is given by $D(t) = a(t)r$. The recession velocity, due to expansion, is therefore given by $\mathrm{d}D(t)/dt$ (Eqn. 1.2.3).

$$v_r = \frac{\mathrm{d}a(t)}{\mathrm{d}t}r = \dot{a}r = \frac{\dot{a}}{a}D \tag{1.2.3}$$

Which is Hubble's Law where $\mathrm{H}(t) = \frac{\dot{a}}{a}$, with $\mathrm{H}_0$ the present day value of the Hubble parameter. More generally, it can demonstrated that photons propagating in FLRW space can have their frequencies modified by changes in the scale factor. In GR, light follows null geodesics with $\mathrm{d}s = 0$ and the metric reduces to $c\mathrm{d}t^2 = a(t)^2\frac{\mathrm{d}r^2}{1-kr^2}$. A useful quantity to define is the co-moving distance $D_c$ which is time independent.

$$D_c = \int_0^{r_1} \frac{\mathrm{d}r}{(1-kr^2)^{1/2}} = \int_{t_e}^{t_o} \frac{c\mathrm{d}t}{a(t)} \tag{1.2.4}$$

This form of the metric can be separated into time and $r$ dependency, from this it can be shown that for a propagating photon the ratio of observed and emitted wavelength is related to the ratio of the scale factors at emission and observation (Eqn. 1.2.5). Also defined is redshift ($z$), a dimensionless quantity which is used to label both distance and time.

$$\frac{\lambda_o}{\lambda_e} = \frac{a_o}{a_e} = \frac{1}{a_e} = (1+z) \tag{1.2.5}$$

Hence by measuring the wavelength shift of light from a distant object, the scale factor at the time of emission can be determined although in practice this relation is complicated by peculiar velocities. The above equations assumes that objects move only because of expansion (called the Hubble flow), while peculiar velocity means that a galaxy is moving with respect to the Hubble flow. This means that the observed shift ($z_{obs}$) will have a component due to cosmological expansion ($z_{cos}$) and the Doppler effect from a peculiar velocity ($v_{pec}$). In the case of low velocity ($v_{pec} << c$) the observed redshift can be calculated as $(1+z_{obs}) = (1+z_{cos})(1+v_{pec}/c)$. In order to relate the scale factor at emission to a distance, Eqn. 1.2.4 demonstrates that the evolution of the scale factor must be understood. This requires solving the Einstein field equations which relate space-time to the energy and matter content. Assuming the universe has a constant matter density $\rho$ and pressure $p$, one can obtain the Friedmann equations (Eqn. 1.2.6 and 1.2.7, Friedmann 1922).

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{1.2.6}$$

$$\mathrm{H}^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{1.2.7}$$

The Friedmann equations describe how the evolution of the scale factor is driven by gravity ($G$ is the gravitational constant), curvature, pressure and the Cosmological Constant ($\Lambda$). The Cosmological Constant arises in GR as constant of integration,

as such its value is not defined. The constant was first utilised by Einstein to balance the effect of gravity in order to reconcile the belief of a static universe with GR. Later, this solution was shown to be unstable and observational work indicated the universe was not static (Hubble, 1929).

To solve the Friedmann equations the density evolution must be understood, as components with different equations of state will change their density differently as the universe expands. In the case of matter, as the universe expands, the particle number will be conserved so the density will be inversely proportional to the volume (i.e. $\rho_M \propto a^{-3}$). For radiation, the energy density is also diluted by the volume, but there is an additional component due to redshift which decreases the energy of each photon ($\rho_\gamma \propto a^{-4}$). Generally the total density can be written as Eqn. 1.2.8, where the "0" indicates the present day density.

$$\rho(a) = \rho_{\gamma,0}a^{-4} + \rho_{M,0}a^{-3} \tag{1.2.8}$$

It is convenient at this point to recast the density in terms of the critical density, the density at which the universe is flat ($k = 0$) and is just expanding quickly enough to expand forever (analogous to an orbital escape speed). Using Eqn. 1.2.7, one can derive the critical density, $\rho_c$, given by

$$\rho_c = \frac{3\mathrm{H}_0^2}{8\pi G}. \tag{1.2.9}$$

Densities are often written as the ratio of the density to the critical density, $\Omega_0 = \rho_0/\rho_c$. Using this, Eqn. 1.2.7 can be simplified to the following.

$$E(z) = \frac{\mathrm{H}(z)}{\mathrm{H}_0} = \left[\Omega_{\gamma,0}(1+z)^4 + \Omega_{m,0}(1+z)^3 + \Omega_{k,0}(1+z)^2 + \Omega_{\Lambda,0}\right]^{1/2} \tag{1.2.10}$$

Where $\rho_k = -kc^2/\mathrm{H}_0^2$ and $\rho_\Lambda = \Lambda c^2/3\mathrm{H}_0^2$ have been introduced. With this parametrisation of the evolution of the Hubble parameter it is finally possible to solve for

the co-moving distance (Eqn. 1.2.4), using the Eqn. 1.2.3 to rewrite the integral over redshift.

$$D_c = \int_{t_e}^{t_o} \frac{c\mathrm{d}t}{a(t)} = \int_a^1 \frac{c\mathrm{d}a}{a(t)^2 \mathrm{H}(a)} = \int_0^z \frac{c\mathrm{d}z\prime}{(1+z\prime)E(z\prime)} \tag{1.2.11}$$

The co-moving distance to an object does not change as the universe expands, but it means that one 'co-moving meter' in the distant past was smaller than today. The co-moving distance is related to the physical (or proper) distance by $aD_c = D_p$, proper distances do not expand with the universe.

As well as computing the distance to a source, it is also necessary to convert the observed properties into ones which are independent of the redshift of a galaxy, such as the separation between two galaxies or a luminosity. The co-moving distance already described is purely along the radial direction, we therefore define the angular diameter distance, $D_A = l/\theta$, where $\theta$ is the observed angular size of an object with (physical) size $l$. The angular diameter distance is non-trivial in the case of non-zero curvature, in the case of $k = 0$ it simplifies to:

$$D_A = \frac{D_c}{(1+z)} \tag{1.2.12}$$

Analogously to the angular diameter distance we define a luminosity distance $D_L$, such that $F = L/(4\pi D_L^2)$, for an object of rest-frame luminosity $L$ and an observed flux $F$. Note that the luminosity and flux are bolometric, integrated over all wavelengths. The luminosity distance can be expressed in terms of the co-moving distance:

$$D_L = (1+z)D_c \tag{1.2.13}$$

Eqn. 1.2.11 now allows for the co-moving distance to an object to be calculated from the observed redshift, given some cosmological parameters. The angular diameter distance defined by Eqn. 1.2.12 allows for separations on the sky to be converted

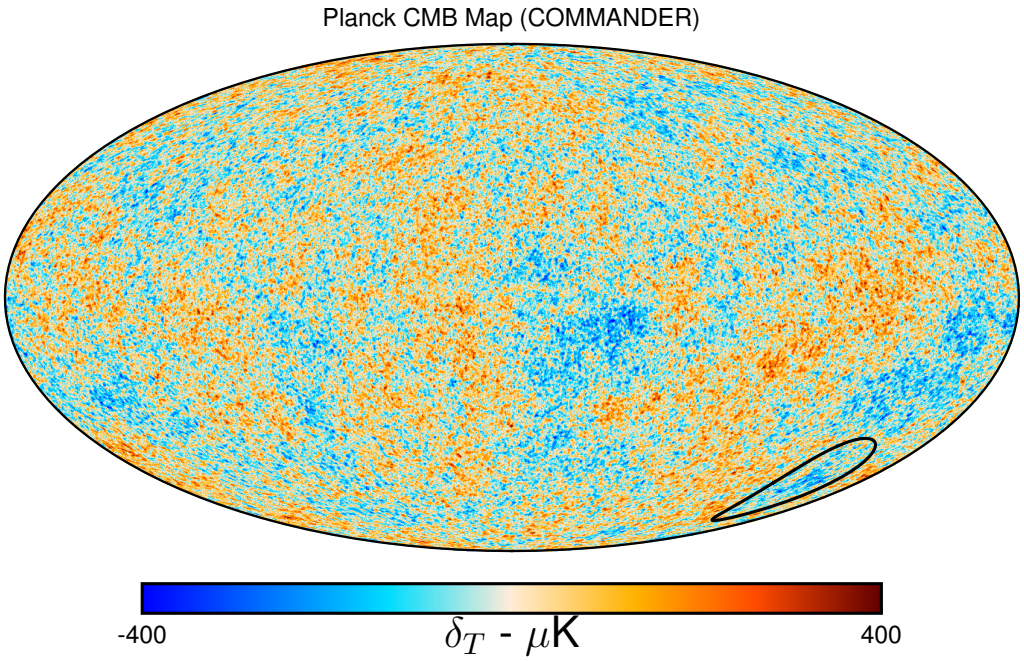


Figure 1.1: A map of the temperature anisotropy of the CMB as measured by the Planck mission (Planck Collaboration et al., 2016a). The map uses the COMMANDER foreground subtraction and is plotted in Galactic coordinates. The monopole and dipole have been subtracted. The black outline in the lower right shows a 15° radius around the famous Cold Spot feature, the circle is heavily distorted by the Mollweide projection.

into distances when combined with the redshift. This is tremendously powerful as it enables spectroscopic surveys to map the distribution of galaxies in three dimension, calculated from the redshifts and coordinates on the sky. First however the cosmological parameters which describe the content of the universe in Eqn. 1.2.10 have to be understood.

### 1.2.1 The Cosmic Microwave Background

After the discovery of the relationship between redshift and distance (Hubble, 1929), the next major body of evidence for expanding universe was the discovery of the Cosmic Microwave Background, which is also one of the richest probes of cosmological parameters. The CMB was originally discovered accidentally by Penzias & Wilson (1965), who reported an excess antenna temperature in their radio telescope of 3.5±1 K. The pair originally believed the excess to be interference and attempted to

eliminate it, but upon sharing their work with astronomers they realised the potential impact of their measurement. The CMB was proposed to be the relic radiation of a hot Big Bang, competing cosmologies proposed other hypotheses which would later be ruled out.

In a hot Big Bang scenario matter in the early universe was hot and totally ionised, the universe was rapidly expanding and so matter cooled with expansion. In this fully ionised state, free electrons effectively scatter any propagating light, and so light cannot travel far. Eventually the cooling universe reaches a temperature ($\sim 3000$ K) when the protons and light nuclei could recombine with electrons to form atoms, this is the epoch of recombination. The universe then becomes transparent to (non-ionising) radiation and light can propagate over cosmological distances for the first time. As this light was emitted at an early epoch, it is observed at long wavelengths as the CMB, peaking at 160 GHz. It was realised that in this scenario the initial perturbations (which would later collapse to form structure) would result in fluctuations in the CMB temperature across the sky, albeit with a small amplitude ($\Delta T/T \leqslant 10^{-3}$, e.g. Peebles & Yu 1970). By measuring these fluctuations, one can constrain structure formation, cosmology and the properties of the initial perturbations.

Early ground based and balloon studies of the CMB anisotropies lead to the detection of the dipole (Conklin, 1969; Henry, 1971), the dipole is largely caused by the motion of the observer with respect to the CMB rest-frame (Planck Collaboration et al., 2014f). The launch of the Cosmic Background Explorer (COBE) mission represented a huge leap forward in capability, when the Differential Microwave Radiometer (DMR) detected fluctuations at the $10^{-5}$ level on scales larger than $7^{\circ}$ (Smoot et al., 1992). The Far Infrared Absolute Spectrophotometer (FIRAS), another COBE instrument confirmed the nature of the CMB and demonstrated that the spectrum was a near-perfect blackbody as expected from a hot Big Bang (Mather et al., 1994; Fixsen et al., 1996). After COBE, the focus of CMB research became resolving the CMB with sufficient resolution to measure the baryon oscillations in

the power-spectrum which have tremendous power to constrain cosmology. Baryon Acoustic Oscillations are created by sound waves propagating in the photon-baryon fluid in the early universe, travelling out from collapsing perturbations. These sound waves can travel a fixed distance before the universe cools, imprinted a characteristic scale on the CMB and matter power-spectrum. In particular the Wilkinson Microwave Anisotropy Probe (WMAP, e.g. Spergel et al. 2007) and Planck missions (Planck Collaboration et al. 2016a) produced high resolution CMB maps of the full sky with extreme fidelity. Fig. 1.1 shows a foreground cleaned map from Planck. From these maps, cosmological parameters can be tightly constrained by fitting the power spectrum of the CMB, i.e. the map decomposed into amplitudes of spherical harmonics (e.g. Planck Collaboration et al. 2016b). Furthermore, these datasets have enabled the systematic study of secondary anisotropies in the CMB, fluctuations in the CMB which were produced long after the CMB was emitted. Some examples of secondary anisotropies are the Sunyaev–Zel'dovich effect (Birkinshaw et al., 1984), gravitational lensing (Smith et al., 2007) and the Integrated Sachs-Wolfe effect (e.g. Planck Collaboration et al. 2016d).

### 1.2.2 Standard Cosmology: ΛCDM

The standard cosmological model, Λ Cold Dark Matter (ΛCDM), is the convergence of observation and theoretical cosmology over the past few decades. The strength of standard cosmology is in its ability to match many independent cosmological probes simultaneously. We review the development of two pillars of the ΛCDM, dark matter and the cosmological constant, as the expanding universe has already been introduced in Sect. 1.2.

The origin of the hypothesis of dark matter is unclear, with some citing work by Lord Kelvin in the late $19^{\text{th}}$ century, and others work that was done on stellar velocities near the Sun (Oort, 1932; Kapteyn, 1922). The first clear evidence came from studies of the dynamics of galaxy clusters, it was shown the velocity dispersion

of galaxies in the Coma Cluster is too high if the mass of the cluster was only based on the stars (Zwicky, 1933). By applying virial theorem, Zwicky estimated the mass in the cluster would have to exceed the luminous matter by a factor of 400. On smaller scales, the technique of measuring rotation curves took time to mature, with early studies of M31 indicating the mass-to-light ratio increased with radius (Babcock, 1939). In the 1970's and early 1980's, observational and theoretical work precipitated into the dark matter hypothesis. The early studies of M31 and the Milky Way were expanded to other galaxies, which also showed a characteristic flattening in the rotation curve indicating unseen mass at large radii (Rogstad & Shostak, 1972; Rubin et al., 1980). In parallel, theoretical work argued that in order to explain the stability of spiral galaxies very massive galactic halos were required, which was consistent with the work from rotation curve analyses (Ostriker & Peebles, 1973). Theorists proposed Cold Dark Matter as a model for large scale structure, where cold means that the particles have negligible thermal velocities in the early universe (Blumenthal et al., 1984). At this time there was no reason to believe this matter was not baryonic, many believed it to be compact objects which were not very luminous. In the 1990's, extensive gravitational microlensing searches were conducted to find these compact halo objects, ruling out compact object dark matter over a wide range of masses (Alcock et al., 2000). Evidence for collisionless dark matter began to arise, with measurements of galaxy clustering from redshift surveys estimating $\Omega_m \simeq 0.3$ (Peacock et al., 2001), while primordial nucleosynthesis constraining the density of baryonic matter to be much lower (Copi et al., 1995). This picture of non-baryonic dark matter has been strengthen by measurements of the CMB (e.g. Spergel et al. 2007; Planck Collaboration et al. 2016b) and colliding galaxy clusters (Clowe et al., 2006). Although alternative models to dark matter exist, none yet have the predictive power or observational success of CDM.

Within ΛCDM, the dominant contribution to the energy density at the present time is the cosmological constant, $\Lambda$. $\Lambda$ was originally introduced by Einstein into GR to construct solutions for a static universe, however after evidence emerged for a non-

static universe, the constant fell into disuse before being reintroduced for a different purpose. Perhaps the first observational evidence for $\Lambda$ came from galaxy clustering at large scales, which pointed to $\Omega_m \simeq 0.3$ (Efstathiou et al., 1990), implying $\Lambda > 0$ for a flat universe. Further evidence for $\Lambda$ came from observations of the acceleration of the universe, exploiting type Ia supernovae as standardisable candles to measure luminosity distance (Perlmutter et al., 1999; Riess et al., 1998). Results from the CMB and from Baryon Acoustic Oscillations point to a flat universe and therefore a non-zero cosmological constant in the context of $\Lambda$CDM (Eisenstein et al., 2005; Balbi et al., 2001; Spergel et al., 2007). However there exists tension between particle physics and the value of $\Lambda$ inferred from observations (Weinberg, 1989; Carroll, 2001). For this reason the effect attributed to $\Lambda$ in standard cosmology is more generally referred to as dark energy. There are many alternative models such as modifications to general relativity or additional forces and fields, for example $f(\mathrm{R})$ gravity (Sotiriou & Faraoni, 2010) and Quintessence (Tsujikawa, 2013). These alternatives can be tested in a series of methods, for example with growth of large-scale structure with time (e.g. Pezzotta et al. 2017) and the Integrated Sachs-Wolfe effect (e.g. Barreira et al., 2014a).

In summary, standard cosmology consists of a flat universe ($\Omega = 1$) dominated by dark energy at late times ($\Omega_\Lambda \simeq 0.7$), with about 5 times more cold dark matter than baryonic matter ($\Omega_m \simeq 0.3$, $\Omega_b \simeq 0.05$).

### 1.2.3 The Integrated Sachs-Wolfe Effect

The Sachs-Wolfe effect is a key mechanism in the formation of the anisotropies of the CMB (Sachs & Wolfe, 1967) and it is the shifting of CMB photons in energy as they traverse gravitational potentials. GR states that photons climbing out of a gravitational well will be redshifted, and conversely blueshifted if moving into one. The net contributions of the Sachs-Wolfe effect on the CMB is typically broken down into two components, the integrated Sachs-Wolfe (ISW) and the ordinary

effect (sometimes just called the SW effect). The ordinary SW effect is relevant at very early times, shortly after CMB was emitted. As the CMB photons propagate away from the surface of last scattering they experience gravitational redshift if the photons are emitted from an over-dense region, and similarly be blue-shifted for an under-dense one. The ISW effect on the other hand is caused by the structure along the line of sight between the observer and the surface of last scattering.

As a photon moves into the gravitational field, e.g. of a galaxy super-cluster, it will experience boost in frequency. As it climbs back up the potential on its way out, one would expect the redshifted to exactly cancel the previous blueshift, as is indeed the case in GR for a Schwarzschild metric. However, if the potential of the cluster is modified as the photon crosses it, the shifts will not necessarily cancel. In accelerating cosmologies the cluster is stretched as the CMB light passes through it, therefore its potential is weakened over time. Therefore the CMB light which has passed through an over-density will be at a slightly higher temperature than the mean. Alternatively for an under-dense region, such as a void, the CMB will have a slightly lower temperature. The change in CMB temperature due to the ISW effect is related to the change in gravitational potential with (conformal) time ($\dot{\Phi}$), integrated from the surface of last scattering to the observer:

$$\delta_T(\mathbf{n}) = \frac{\Delta T}{T}(\mathbf{n}) = 2\int_{t_{LS}}^{t_0} \dot{\Phi}(\mathbf{n}, t)dt, \tag{1.2.14}$$

where $\delta_T$ is the temperature is the ISW temperature anisotropy, measured at the current time ($t_0$) in direction $\mathbf{n}$. $T$ is the mean CMB temperature and $t_{LS}$ is the time of decoupling. It can be shown that the on linear scales, the potential is $\Phi \propto a^2\rho D$, where $D$ is the linear growth factor. In a matter dominated universe $D \propto a$ and $\rho_M \propto a^{-3}$, therefore $a^2\rho D$ is constant. The ISW effect is negligible when the universe is matter dominated. The Rees-Sciama effect is the non-linear extension of the ISW effect, which can be non-zero in matter dominated universes but with a lower amplitude than the ISW effect (Rees & Sciama, 1968). Within standard cosmology

there have been two epochs where the ISW effect was non-negligible, immediately after matter-radiation decoupling when photons had a noticeable contribution to the energy density (dubbed the early-time ISW effect), and at $z < 1$ where the universe has begun to accelerate under the increasing impact of the cosmological constant (late-time ISW effect). The late-time ISW effect is thus a dynamical test of the evolution of large scale structure, and is therefore sensitive to dark energy. Some alternative models of the accelerated expansion of the universe predict different ISW effects (Beck et al., 2018; Barreira et al., 2014a).

The CMB is most notably studied through its angular power spectrum, however the contribution of the late-time ISW effect is small in comparison to the primary CMB anisotropies. The contribution is also largest at large scales ($l < 10$), which is where measurements suffer from cosmic variance the most. A better method to detect the ISW effect is to cross-correlate CMB fluctuations with tracers of the large scale structure (Crittenden & Turok, 1996). This technique has been successfully applied to galaxy and quasar samples (Fosalba et al., 2003; Cabré et al., 2006; Sawangwit et al., 2010a), radio catalogues and even maps reconstructed from CMB lensing (Planck Collaboration et al., 2016d). Combined analyses have reported detection significances of $\sim 4.4\sigma$ (Giannantonio et al., 2012), but most searches provide marginal results. An alternative method is stacking the CMB on the position of large scale structures, voids and super-clusters (Granett et al., 2008).

The ISW is basis of the research presented in Chapter 2. We have tested claims that a large void caused a feature on the CMB called the Cold Spot, via the ISW effect (Szapudi et al., 2015).

## 1.3 Gas and Galaxies

Cold Dark Matter in numerical simulations of structure formation ubiquitously forms spheroidal halos with a self-similar density profile (Navarro et al., 1996). Galaxy

formation in ΛCDM proceeds as gas is accreted into the collapsed dark matter halo, over time the gas cools sufficiently to fragment and form stars (White & Rees, 1978)

In Chapters 3 and 4 we will use absorption line systems, specifically Damped Lyman $\alpha$ systems, to study the connection between neutral gas and galaxies at high redshift. We first review the basics of galaxy clustering, which is a major technique in Chapter 4. We will then summarise the properties of DLAs and the history of the field, this also builds on the discussion of galaxy clustering. Chapter 4 makes extensive use of galaxy formation simulations, which are introduced in the last subsection.

### 1.3.1 Galaxy Clustering

Galaxies form in dark matter halos in standard cosmology, but as the dark matter properties of galaxies cannot be measured directly they must be inferred from other observations. Galaxy clustering offers an observable quantity which can be related to the halos masses off the galaxy population, via a property called bias. First, we introduce a dimensionless form of the density field $\delta$, given by

$$\delta(x) = \frac{\rho(x)}{\langle\rho\rangle} - 1, \tag{1.3.1}$$

where $\rho(x)$ and $\langle\rho\rangle$ are the matter density and position $x$ and the mean density respectively. Bias means that the overdensities and underdensities of galaxies ($\delta_g = n_g/\langle n_g\rangle - 1$) does not correspond to the same overdensity in the total matter field (dominated by dark matter). The galaxy density is related to the matter density via the bias, $b$.

$$\delta_g = b\delta_m \tag{1.3.2}$$

At large scales and for mild perturbations the bias can be treated as independent of the fluctuation scale, but for more extreme cases numerical simulations are required. The physical origin for bias is illustrated in Fig. 1.2, which shows an example density

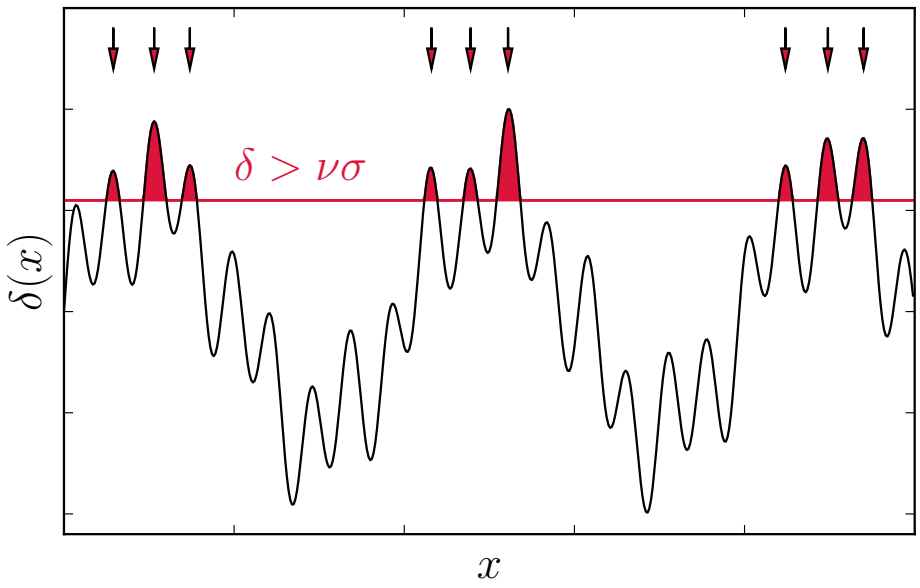


Figure 1.2: An demonstration of the principle behind biased traces of the density field, such as galaxies or halos. The figure shows a one dimensional illustrative density field, constructed out of some sinusoidal perturbations. If halos first begin to collapse only above a certain density threshold ($\nu\sigma$), then naturally the halos will be biased which means that the peaks above the threshold (red arrows and filled) do not linearly trace the density field.

field with the peaks marked. Massive dark matter halos will collapse earlier from the high the peaks of this field. Peaks in the density field are more likely to reach the threshold for collapse if within a large scale over-density. This description was put forward for galaxy cluster formation (Kaiser, 1984) but was generalised to smaller halos and galaxies (Cole & Kaiser, 1989). For halos, the bias is related to the halo mass (e.g. Sheth & Tormen 1999), so by measuring the bias of a population of galaxies one can compute the halo mass of the population (weighted by bias).

The statistics of a density field can be parametrised by the correlation function, $\xi(r)$, defined by $\xi(r) = \langle \rho(x)\rho(x+r) \rangle$. Observationally the density field is sampled with galaxies, the correlation function is estimated using galaxy pair counts through different estimators (e.g. Landy & Szalay 1993). For galaxies or halos, the correlation function is the excess probability of finding two galaxies separated by a distance $r$ over the random expectation. By Eqn. 1.3.2 it is easy to see how the correlation function for matter relates to that of a galaxy population.

$$\xi_g(r) = b^2 \xi_m(r) \tag{1.3.3}$$

Hence by measuring the correlation function for a sample of galaxies one can estimate the linear bias, and the halo mass. The matter correlation function ($\xi_m(r)$) can be calculated from linear theory, by evolving an initial power spectrum of perturbations derived for analyses of the CMB. On large scales, where the evolution is insensitive to non-linear evolution, the growth of perturbations over time is parametrised by the linear growth factor, $D(t)$, such that

$$\delta(x,t) = D(t)\delta(x,0), \tag{1.3.4}$$

Where $\delta(x,t)$ is the density field at position $x$ and time t, $\delta(x,0)$ is the initial perturbation.

### 1.3.2 Damped Lyman $\alpha$ Systems

The process by which gas is accreted into galaxies is difficult to probe at high redshift, and one way to study this gas is by using absorption line systems in the spectra of background quasars. Fig. 1.3 shows a section of the Ly$\alpha$ forest in a $z = 4.17$ quasar. The Ly$\alpha$ transition has a strong oscillator strength so absorption is a sensitive probe of neutral gas. The types of absorption line systems observed can be divided up by their column density, which is the areal number density of absorbing atoms along the line of sight. Most of the lines seen in Fig. 1.3 are produced by forest absorbers, with column densities $N_{\rm H\,I} < 10^{17}$ cm$^{-2}$. Above a column density of $10^{17}$ cm$^{-2}$ neutral hydrogen becomes optically thick to radiation with energies above 13.6 eV, i.e. ionising Lyman continuum radiation. Absorption line systems in the range of $10^{17} < (N_{\rm H\,I}$ / cm$^{-2}) < 2 \times 10^{20}$ produce characteristic break in the quasar spectrum at 912 Å (absorber rest frame), these are known as Lyman Limit Systems (LLS). Damped Lyman $\alpha$ Systems (DLAs) are the highest column density absorbers,

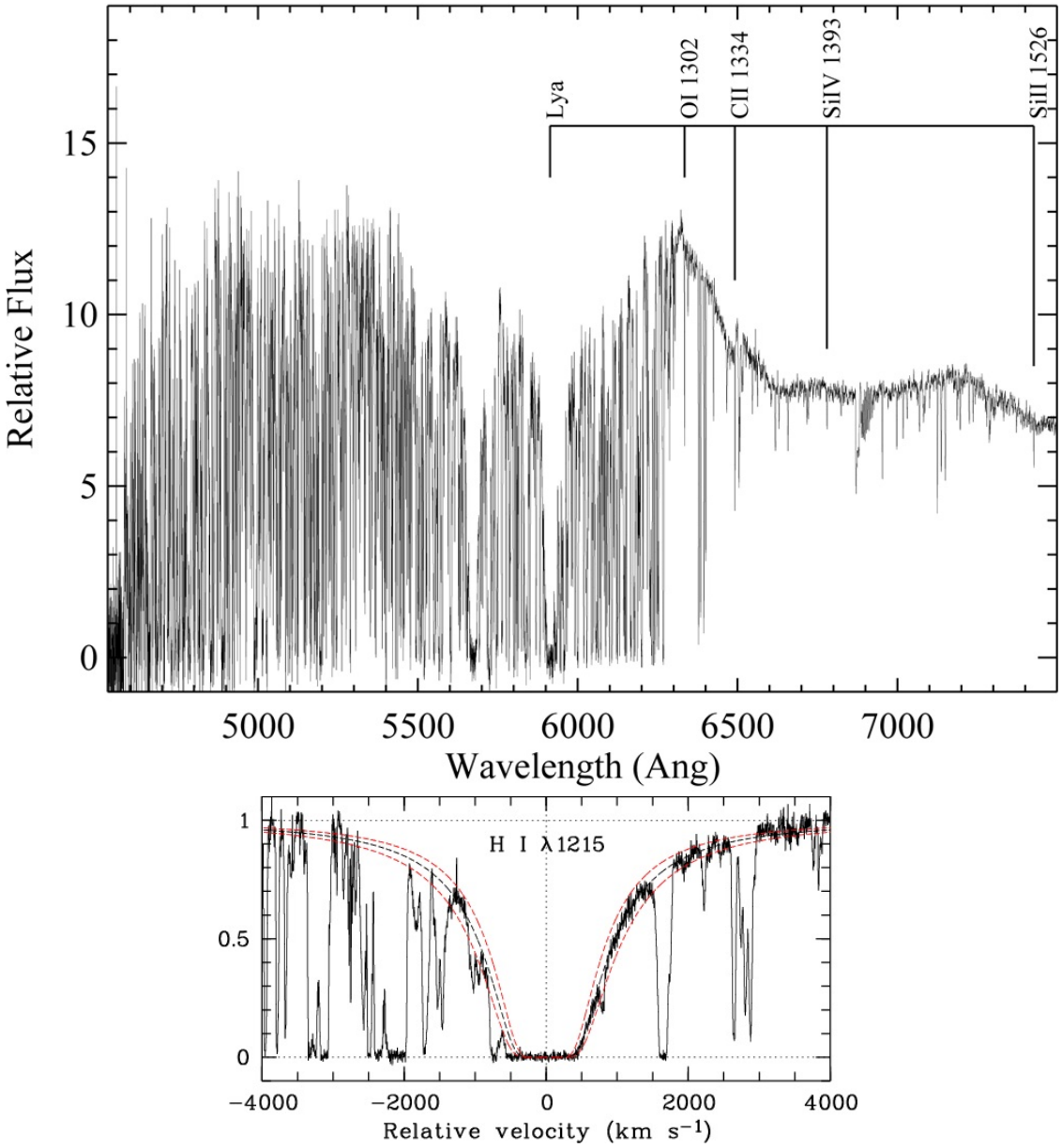


Figure 1.3: **Top:** The Ly$\alpha$ forest observed in QSO PSS0209+0517 with the transitions of a DLA at $z = 3.864$ marked, a second DLA can be seen at $z = 3.667$ (Wolfe et al., 2005). **Lower:** The Ly$\alpha$ line of a DLA in the spectrum of Q2343+125 with $z = 2.431$ showing the width of damped absorption in velocity space, with a Voigt profile fit to the absorption (Noterdaeme et al., 2007).

defined as $N_{\rm H\,I} \geq 2 \times 10^{20}$ cm$^{-2}$. DLAs are characterised by their extremely broad absorption trough, this is caused by the Lorentzian wings of the naturally broadened absorption profile, which come to be strong at very high column density. DLAs are interesting as a probe of galaxy evolution because most of the hydrogen is neutral, as opposed to lower column absorbers which have more ionisation (Viegas, 1995). DLAs are self-shielded from the ultraviolet background, and are hence the most plausible systems to fuel star formation at high redshift. Understanding how DLAs are (or are not) linked to galaxies carries information about the neutral gas budget of high redshift galaxies (Peroux, 2002).

DLAs first began to be detected in the 1970's, their extremely broad absorption lines meant they could be identified even at low spectral resolution (Beaver et al., 1972; Carswell et al., 1975). Follow-up of these systems using 21 cm absorption against radio loud quasars confirmed the H I was cool and had now velocity dispersion (Wolfe & Davis, 1979). At this time $z > 2$ galaxies had not been directly confirmed (only quasars) and so the opportunity to study high redshift galaxies in detail was appealing. The first survey for DLAs at $z > 2$ identified 15 systems (Wolfe, 1986) and it was suggested that DLAs were H I disks and the progenitors of modern galaxies. Metal absorption lines associated with the Ly$\alpha$ were shown to be dominated by the low-ionisation ions (low-ions), indicating these metals are not highly ionised and may trace the HI. Furthermore, because of the low ionisation correction needed the elemental abundances and metallicities of DLAs can be measured precisely (Viegas, 1995). The HIRES high dispersion echelle spectrograph on Keck allowed for the velocity structure of the low-ions to be resolved (Wolfe et al., 1994). A model of DLA kinematics based on the hypothesis that they were thick rotating H I disks was able to match the velocity profiles observed in the low-ions (Prochaska & Wolfe, 1998). Haehnelt et al. (1998) put forward a competing explanation for DLAs, as protogalactic clumps which were able to explain the velocity widths via rotation, infall and merging. The Sloan Digital Sky Survey with a high multiplex spectrograph identified thousands of DLAs in absorption in a systematic spectroscopic survey

(Prochaska & Herbert-Fort, 2004), allowing for precision measurements of the column density distribution (Noterdaeme et al., 2012) and providing a large sample for follow-up.

Early in the study of DLAs, it was believed they were the interstellar medium of high redshift galaxies, as there are multiple lines of evidence which suggest a link between DLAs and star formation. Firstly, DLAs are the major reservoirs for neutral hydrogen over most of the history of the universe (Wolfe, 1986; Lanzetta et al., 1995), H I is the fuel that becomes $H_2$ to form stars. DLAs are also enriched in metallicity above the typical IGM, as only ∼10-13% of the IGM volume is enriched to [M/H]≥ 3[1](D'Odorico et al., 2016) while DLAs at the same redshift have an average metallicity of [M/H]= −1.1 (Rafelski et al., 2012). Furthermore the DLA population evolves to higher metallicity with cosmic time, as would be expected if DLAs were polluted from ongoing star formation (Rafelski et al., 2012). There is also indirect evidence of heating in DLAs from measuring the column density of C II* (Wolfe et al., 2003), which due to the same electron state which gives rise to the 158 $\mu$m [C II] line, an important cooling line in the ISM (Wright et al., 1991). These results show that some of the DLA population appear to have excess heating above the UV background which could be due to star formation. The clustering bias of DLAs has also been measured by the Baryon Oscillation Spectroscopic Survey (BOSS), which found $b = 2.0 \pm 0.2$ (Pérez-Ràfols et al., 2018b). This bias would correspond to a halo mass of $5.7 \times 10^{11}$ $M_\odot$ , indicating that at least some fraction of the DLA population must be associated with massive halos (and therefore galaxies). In Chapter 4 we will revisit the bias with our observations.

The search to identify DLA host galaxies direct began soon after the systems were detected in absorption (e.g. Foltz et al. 1986). Many searches targeted Ly$\alpha$ emission from the host galaxy, as this had be benefit that in the deep trough of the DLA absorption the quasar is suppressed completely allowing for faint emission to be

[1]Here metallicity is defined as the log abundance with respect to the solar abundance pattern (Asplund et al., 2009). [M/H]=0 means the ratio of metals to hydrogen is the same as solar, [M/H]= −3 means the metal to hydrogen ratio is a thousand times less than solar.

detected. Slit spectroscopy had the advantage of greater sensitivity (e.g. Foltz et al. 1986), while narrowband imaging could detect sources further from the quasar (e.g. Smith et al. 1989). Most searches for Ly$\alpha$ bright galaxies found no confirmed detections, but Møller & Warren (1993) detected the first candidate with this technique (however for a DLA close to the quasar redshift). Imaging with broadband filters can reach faint galaxies, however in the case of DLAs the bright quasar ($m_r$ ∼18-20) makes the detection of faint $z > 2$ galaxies ($m_r$ ∼>24) at small impact parameters difficult. The excellent spatial resolution of the Hubble space telescope was exploited to search for DLA hosts (e.g. Colbert & Malkan 2002). Warren et al. (2001) detected many counterparts in deep NICMOS imaging, however few were confirmed spectroscopically (Møller et al., 2002). One method to deal with the bright quasar is the *Shot in the Dark* technique, which allows for star formation to be measured in broadband filters at the position of the DLA (O'Meara et al., 2006). The technique uses targeted sightlines that have both a DLA and a higher-redshift LLS or DLA, whereby the higher redshift system acts a blocking filter for the quasar below its Lyman limit. Therefore, one can image the sightlines below this limit where the quasar is suppressed and potentially detect the far ultraviolet of the lower redshift DLA host galaxy (Fumagalli et al., 2010). A survey of 32 quasar fields using this technique with deep HST WFC3/UVIS and Keck imaging failed to detect star formation at the DLA, but placed stringent limits (mean $\mathrm{SFR}_{UV} < 0.09$ $\mathrm{M}_{\odot}\mathrm{yr}^{-1}$ in 2 kpc aperture, Fumagalli et al. 2015). Attempts to use this data to identify hosts away from the DLA sightlines however suffered from confusion (Fumagalli et al., 2015). From the limited detections at high redshift evidence for a relation between the DLA metallicity and the host galaxy luminosity emerged (Møller et al., 2004). On this basis, a highly successful X-Shooter programme detected 8 candidate host galaxies, 7 with Ly$\alpha$ emission by targeting high metallicity systems (Krogager et al., 2017). Detections at submillimeter wavelengths with ALMA (also selected in metallicity) have begun to probe the obscured population of DLA host galaxies (Neeleman et al., 2017). This targeting of high metallicity DLAs is successful in finding host galaxies

but leaves the more metal-poor (i.e. the majority) of DLAs relatively unexplored. In Chapter 3 and 4 we present results from a MUSE survey of DLAs which did not preselect in metallicity.

### 1.3.3 Numerical Simulations

It is essential when studying galaxy evolution to have a means by which to compare observations and theory. On small scales, structure formation cannot be described by simple linear theory, however the evolution can be solved numerically. In practice this is done by solving the forces in a given situation and applying them to the simulated matter in small time steps. The earliest simulations of structure formation considered only the affect of gravity, simulating a universe of collisionless dark matter (e.g. Davis et al. 1985). Simulations typically solve for the dynamics of dark matter using N-body computations, whereby particles make up the dark matter density field with the force on each particle being the sum of gravitational forces from all other particles. In these calculations, gravity is softened at small scales, to prevent arbitrarily large forces at small particle separations (e.g. Aarseth 1963). For cosmological simulations in order to minimise the effect of having a finite volume, periodic boundary conditions are commonly used, as this enables simulations of a limited volume within a homogeneous universe but means that clustering properties on scales comparable to the simulation size are not captured.

Dark matter only simulations have simulated larger cosmological volumes as both computational power and numerical efficiently has increased, reaching volumes of tens of $Gpc^3$ (e.g. Springel et al. 2005; Angulo et al. 2012). While dark matter only simulations have many uses in cosmology, they do not capture the baryon component of galaxy evolution, but there are multiple methods of linking these simulations with realistic galaxies. The most basic method is abundance matching, which assigns galaxies to subhalos by matching by abundance between simulated dark matter halos and observed galaxy populations (Conroy et al., 2006; Moster et al., 2010).

Another more advanced method is using Halo Occupation Distributions (HODs), which can be used to model galaxy clustering (Berlind & Weinberg, 2002). Once the HOD parameters are fit to observations, the scheme can then be applied to populate simulated halos with galaxies (e.g. Smith et al. 2017). Although these methods can produce realistic distributions of galaxies they do not simulate the process of galaxy formation, semi-analytic models (SAMs) on the other hand post-process dark matter only simulations and calculate galaxy properties based on analytic descriptions of galaxy evolution (White & Frenk, 1991). SAMs can reproduce a wide range of observations, but they have many parameters contained within the model which are calibrated to observations (e.g. Lacey et al. 2016).

Methods of linking dark matter only simulations to galaxies are however indirect, as some processes must be simulated more directly to understand their impact on galaxy formation. Hydrodynamical simulations allow for some of the gas physics in galaxies to be simulated alongside the gravitational forces, such as heating, cooling, star formation and the production of metals. The hydrodynamics are usually simulated either through smoothed particle hydrodynamics (SPH, e.g. Springel et al. 2001a) or mesh based codes (e.g. Springel 2010). In either methodology, there is a minimum scale which is resolved in the simulation which usually depends on the local density. Because of this, simulations of galaxy formation do not currently resolve the scales on which star formation occurs, the process of forming stars is therefore approximated analytically using a subgrid model. In the case of star formation, the empirical star formation law is typically applied (Kennicutt, 1998), and a cosmological hydrodynamical simulation can contain several subgrid models some of which require calibration (Crain et al., 2015). Hydrodynamical simulations are much more computationally expensive than pure dark matter simulations and are therefore limited to smaller volumes. A key concept in current galaxy formation simulations (and many SAMs) is feedback, which is a set of astrophysical processes which can limit the growth of a galaxy. Feedback was proposed to match the observed galaxy luminosity function, as more basic galaxy formation prescriptions produce too many low and

very high mass galaxies (e.g. Silk, 2011). In the case of low mass galaxies, feedback is driven by the supernovae which occur after an episode of star formation, with the energy injected by the supernovae heating the surrounding gas and suppressing further star formation (Dekel & Silk, 1986). For massive galaxies, this suppression arises from the Active Galactic Nuclei (AGN) (Di Matteo et al., 2005).

Galaxy formation simulations enable observations to be put in context with current theory. Hydrodynamical simulations are of particular use as they can be compared to absorption line studies of galaxies, studying the absorption properties of galaxies is one way of testing models like feedback. In Chapters 3 and 4 we will draw comparisons to a hydrodynamical simulation and a semi-analytic model.

## 1.4 Galaxy Redshift Surveys

To best study galaxies, it is first necessary to determine the distance to them as this allows for galaxies to be positioned in three dimensional space, and also for observed properties (e.g. flux) to be converted into intrinsic ones (e.g. luminosity). As established in Sect. 1.2, the cosmological redshift of a galaxy can be converted into a distance, enabling distance estimates even to high redshifts. A limitation of this is that a galaxy's redshift will have a component due to cosmological redshift and one due to the peculiar velocity of the galaxy, via the Doppler effect. Although this reduces the precision with which distance can be determined typical peculiar velocities are small, on the order of a few 100 km $s^{-1}$. In large galaxy clusters with very high halo masses peculiar velocities can rise to 1000's of km $s^{-1}$.

The highest accuracy redshifts are measured spectroscopically, using a spectrograph to disperse the light from an object into a spectrum and measuring the wavelength of characteristic spectral features. Fig. 1.4 shows some template galaxy spectra, constructed from stacked spectra from SDSS (Dobos et al., 2012) and Bielby et al. (2013). Characteristic features are emission lines (e.g. Ly$\alpha$ H$\alpha$, [O II]), absorption lines (e.g. Ca II H&K lines) and large spectral features such as breaks (e.g. the

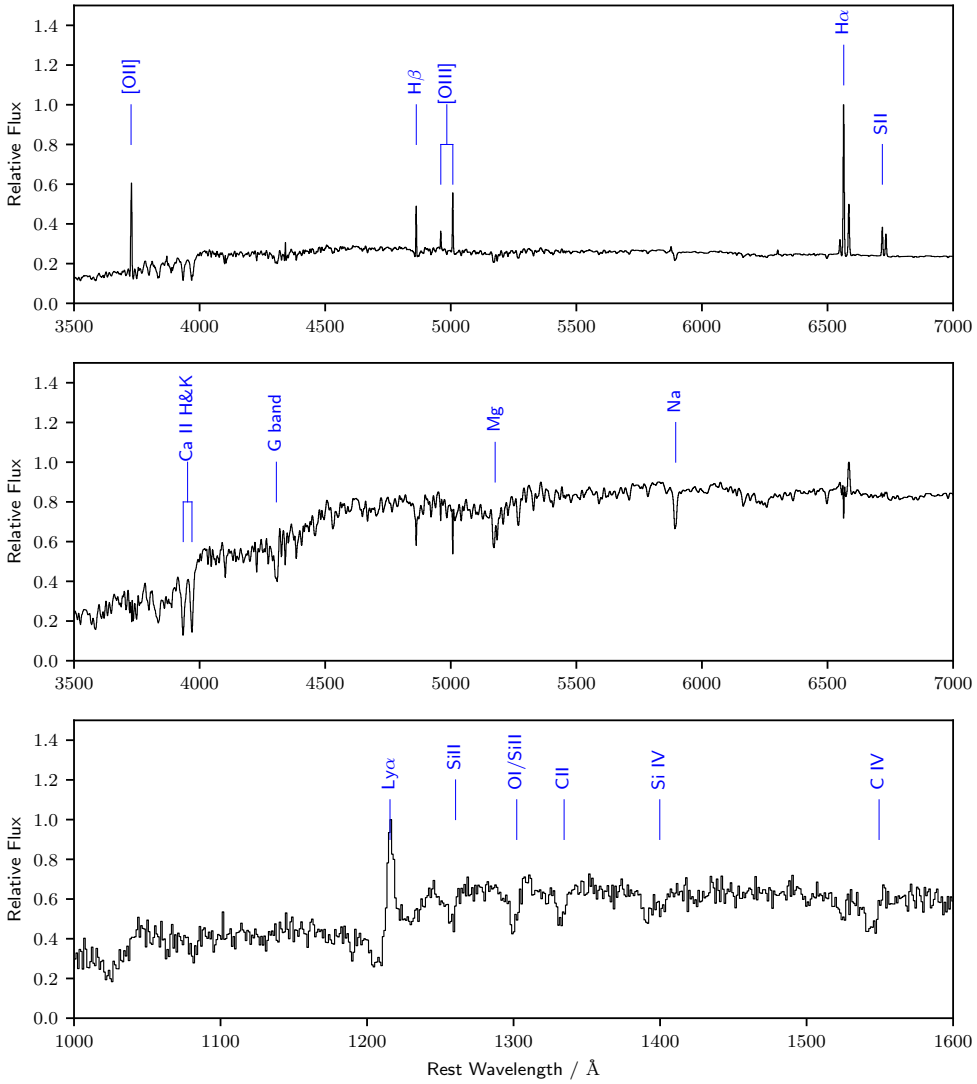


Figure 1.4: Some example template spectra of low and high redshift galaxies, plotted in rest wavelength. Strong emission and absorption lines have been marked. From top to bottom the galaxies are: a low redshift star-forming galaxy, a passive galaxy and a high redshift Lyman Break Galaxy.

Lyman and 4000 Å breaks). Emission lines can be detected in low signal-to-noise ratio spectra but not all galaxies have strong emission lines. A key factor to obtaining a successful redshift is having sufficient wavelength coverage to catch features at a wide range of possible redshifts. It is possible to obtain lower precision redshifts without spectroscopy, photometric redshifts are based on estimates from the flux of an object in different filters. Multiple photometry points can be fit with templates to estimate the redshift, another method trains neutral networks on photometry of galaxies with known redshifts. With photometric redshifts the distance uncertainties are large, even a good photometric redshift has an error tens of Mpc at $z = 1$, for this reason spectroscopic redshift are used in cases where precision is required.

The first large-scale redshift surveys used single object spectrographs, obtaining just tens of redshifts per night (Geller & Huchra, 1989). The Las Campanas Redshift Survey was the first major survey to use a Multi-Object Spectrograph (MOS), which could observe over 100 galaxies simultaneously using a fibre fed MOS (Shectman et al., 1996). Fibre fed systems position optical fibres in the focal plane of the telescope, the fibres relay the light to a spectrograph with the fibre ends places at the slit. This design lead to a revolution in wide field redshift surveys with the 2 degree Field Galaxy Redshift Survey (2dFGRS, Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS, York et al. 2000), which enabled the detection of Baryon Acoustic Oscillations (Cole et al., 2005; Eisenstein et al., 2001) among other science goals. Fibre-fed spectrographs enable wide field of view but suffer from limitations in subtracting the sky emission. A multi-slit MOS is an alternative approach which places a mask with slits cut into it at the focal plane, and the spectrograph optics then disperse the slits with a grating or prism. A multi-slit MOS has a better ability to estimate the sky emission as each slit spans the object and samples the sky on either side. Multi-slit spectrographs coupled with 8-10 meter telescopes have conducted large and deep surveys (e.g. Le Fèvre et al. 2005), and opened up the high redshift universe (Steidel et al., 1996). Fibres and slits both have advantages but both are limited in how densely they can sample objects, a third method is

integral field spectroscopy which allows for a modest field to be completely sampled. An integral field unit (IFU) reformats the image at the focal plane of a telescope into a slit, either with a mirror slicer or fibres, which is then fed into a spectrograph. With an IFU, spectra can be obtained for all objects within the field of view, pre-selection of targets is not required as with multi-object spectroscopy. This allows faint emission line galaxies to be detected over a wide range in redshift, which would not be possible with narrow or broadband selection (Inami et al., 2017). IFUs can also isolate extended emission and probe kinematics (e.g. Leclercq et al. 2017).

A recent advance in high multiplex spectroscopy has been the application of industrial replication, instruments are now being built with several cloned spectrographs. Notably, this was applied to the Multi Unit Spectroscopic Explorer (MUSE; Bacon et al., 2010), a large IFU fitted to the Very Large Telescope. MUSE boasts good spatial sampling (0.2 arcseconds) and a wide wavelength coverage ($3650 - 9300$ Å) over a 1 square arcminute field of view. This combination of performance is unprecedented, it has been made possible by replicating 24 spectrograph units which tile the field of view.

High multiplex spectroscopy has many applications and is a powerful tool in studying galaxies and cosmology. IFUs offer full sampling but are currently limited to modest fields of view. Multi-slit and multi-fibre spectrograph offer wider fields of view but require pre-selection of spectroscopic targets.

### 1.4.1 Thesis Outline

Now that we have introduced the relevant background of the field we present the scientific projects that make up this thesis. The three topics presented are:

- A redshift survey towards the CMB Cold Spot to test claims of divergence from standard cosmology.
- A detailed analysis of a gas-rich structure at high redshift.

- An integral field survey of Damped Lyman $\alpha$ systems at $z \simeq 3.5$, comparing observations with competing models.

These bodies of research are linked by application of highly multiplexed spectroscopy to construct a three dimensional view of galaxies and large scale structure, enabling tests of models on different scales. In Chapter 6 will summarise the conclusions of this thesis.

# Chapter 2

# A Redshift Survey of the CMB Cold Spot

We report the results of the 2dF-VST ATLAS Cold Spot galaxy redshift survey (2CSz) based on imaging from VST ATLAS and spectroscopy from 2dF AAOmega over the core of the CMB Cold Spot. We sparsely surveyed the inner 5° radius of the Cold Spot to a limit of $i_{AB} \leq 19.2$, sampling $\sim 7000$ galaxies at $z < 0.5$. We have found voids at $z = 0.14$, 0.26 and 0.30 but they are interspersed with small over-densities and the scale of these voids is insufficient to explain the Cold Spot through the ΛCDM ISW effect. Combining with previous data out to $z \sim 1$, we conclude that the CMB Cold Spot could not have been imprinted by a void confined to the inner core of the Cold Spot. Additionally we find that our 'control' field GAMA G23 shows a similarity in its galaxy redshift distribution to the Cold Spot. Since the GAMA G23 line-of-sight shows no evidence of a CMB temperature decrement we conclude that the Cold Spot may have a primordial origin rather than being due to line-of-sight effects.

## 2.1 Introduction

The Cosmic Microwave Background (CMB) provides the earliest snapshot of the evolution of the Universe. Detailed observations of its structures by the Wilkinson Microwave Anisotropy Probe (WMAP) and *Planck* missions have shown a universe broadly in concordance with the ΛCDM paradigm. There remain a few anomalies which have been a source of tension with standard cosmology and one such example is the CMB Cold Spot (Vielva et al., 2004). The CMB Cold Spot is an $\sim 5^\circ$ radius, -150 $\mu$K feature in the CMB in the Southern Hemisphere which represents a departure arising in between $< 0.2\%$ (Cruz et al., 2005) to $< 1-2\%$ (Planck Collaboration et al., 2016c) Gaussian simulations. It consists of a cold $5^\circ$ radius core surrounded by a less extreme $10^\circ$ radius halo. The Cold Spot is also surrounded by a high temperature ring which is important for its original detection using a Spherical Mexican Hat Wavelet (SMHW).

A number of proposals have been put forward to explain the Cold Spot, including a non-Gaussian feature (Vielva et al., 2004), an artefact of inflation (Cruz et al., 2005), the axis of rotation of the universe (Jaffe et al., 2005) and the imprint of a supervoid via the Integrated Sachs-Wolfe (ISW) effect (Inoue & Silk, 2006). The ISW effect (Sachs & Wolfe, 1967) occurs in accelerating cosmologies due to the decay of gravitational potentials over time. There is tentative statistical evidence to support the existence of the ISW effect from the cross-correlation of large-scale structure with the CMB, typically up to $3\sigma$ with single tracers and $4-4.5\sigma$ in some combined analyses (e.g. Cabré et al., 2006, Ho et al., 2008, Giannantonio et al., 2008, Sawangwit et al., 2010b, Giannantonio et al., 2012, Planck Collaboration et al., 2016d). The ISW effect must be measured statistically as the primary anisotropy dominates on most scales. It has been hypothesized that a very large void at $z < 1$ could imprint itself on the CMB and explain the Cold Spot in part (e.g. Inoue & Silk, 2006), however the ability of this to explain the Cold Spot has been disputed (e.g. Nadathur et al., 2014). The argument, prior to a detection of such a void,

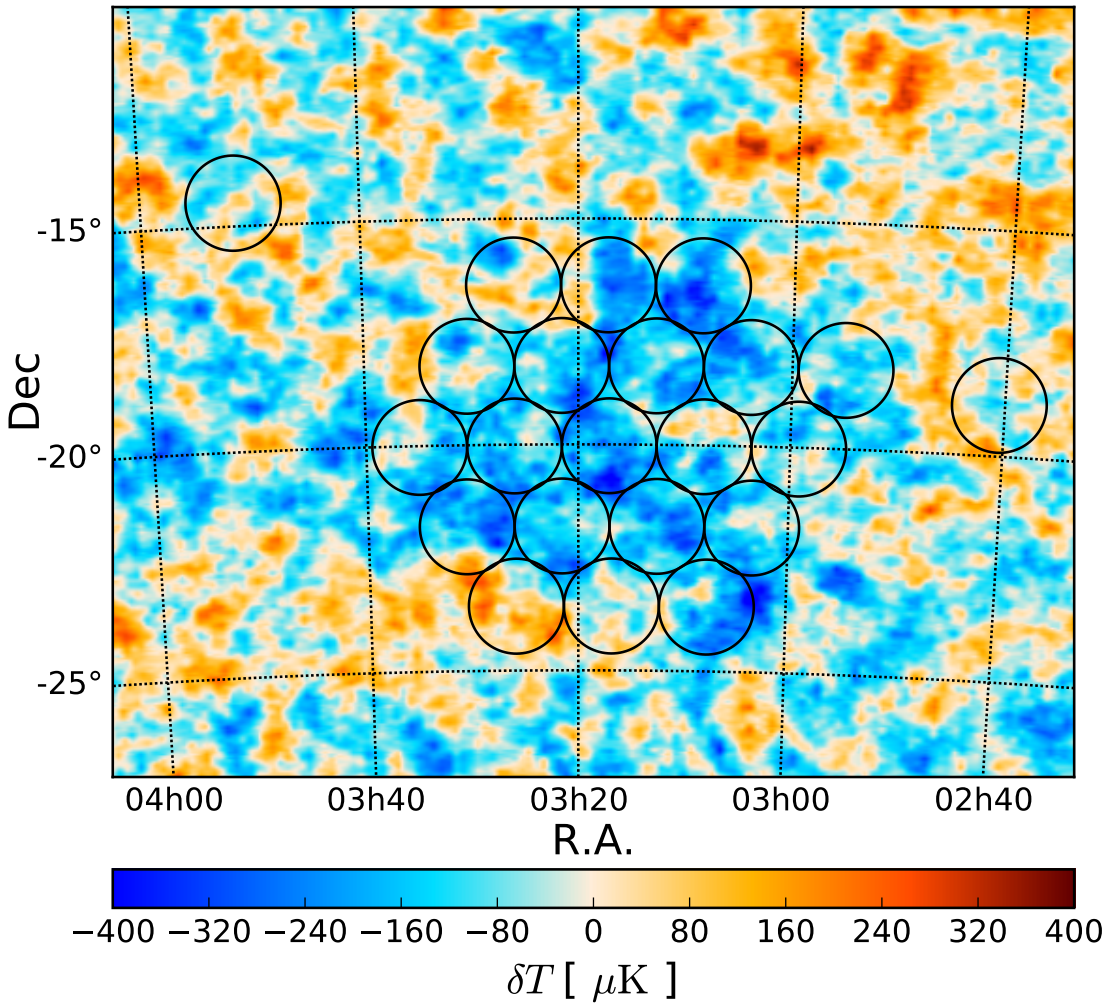


Figure 2.1: The 2CSz survey geometry: Superimposed on the Planck SMICA map of the CMB Cold Spot are circles representing the 22 3 deg$^2$ galaxy redshift fields observed using AAT 2dF+AAOmega. 20 of these fields lie within a 5° radius of the Cold Spot centre.

was that the probability of any void occurring in ΛCDM was much lower than the probability of the Cold Spot arising from primordial Gaussian fluctuations.

The significance of the Cold Spot as an anomaly has been widely discussed. The main problem is to quantify the amount of *a posteriori* selection in the originally claimed 0.2% significance of Cruz et al. (2005). In particular, Zhang & Huterer (2010) pointed out that the use of top-hat or Gaussian kernels provided much lower significance for the Cold Spot than the original SMHW kernel and Bennett et al. (2011) emphasised this viewpoint in their review. Vielva (2010) argued that as long as the original Cold Spot detection was 'blind' and the SMHW kernel well-motivated in a search for non-Gaussian features then this 'look elsewhere' effect in terms of kernels was less relevant. Zhao (2013), Gurzadyan et al. (2014) and Planck Collaboration et al. (2016c) tried a related approach to address the Cold Spot significance and chose the coldest pixels in CMB simulations to look at the small-scale statistics within the surrounding pixels. In the version of Planck Collaboration et al. (2016c), it was found that the temperature profile of the Cold Spot was poorly described by the simulations with $< 1 - 2\%$ having a higher $\chi^2$ compared to the mean than the data. Here we shall essentially adopt this approach, now following Nadathur et al. (2014) and Naidoo et al. (2016), and ultimately test how much any foreground void that is found can reduce this 1-2% significance assuming the original SMHW kernel.

Motivated by theoretical discussion there have been many attempts to detect a void associated with the CMB Cold Spot. Rudnick et al. (2007) searched NVSS radio sources and claimed to find a lower density of objects in the Cold Spot region but this was disputed by Smith & Huterer (2010). Granett et al. (2010) used 7 CFHT MegaCam fields to make a photo-$z$ survey for large under-densities. They found no evidence of a void $0.5 < z < 0.9$ but their data were consistent with a low-$z$ void. This was in line with Francis & Peacock (2010) who found evidence for an under-density in 2MASS in the Cold Spot direction but the ISW imprint was $\sim 5\%$ of the CMB Cold Spot temperature decrement. Bremer et al. (2010) used VLT VIMOS to make a $21.9 < i_{AB} < 23.2$ galaxy redshift survey in 6 small sub-fields

of the Cold Spot area. The total area covered was 0.37deg$^2$ and the redshift range covered was $0.35 < z < 1$. Using VVDS (Le Fèvre et al., 2005) data as control fields, Bremer et al. (2010) found no evidence for anomalously large voids in the Cold Spot sightline. At lower redshifts, Szapudi et al. (2015) using a Pan-STARRS, 2MASS and WISE combined catalogue, constructed photometric redshifts and detected a 220h$^{-1}$Mpc radius supervoid with a central density contrast, $\delta_m \sim -0.14$, spanning $z \approx 0.15 - 0.25$. However, this supervoid would not explain the entirety of the CMB Cold Spot as a ΛCDM ISW effect. The authors argued that the alignment of the Cold Spot and the supervoid could be evidence of a causal link due to some mechanism beyond standard cosmology. It has been argued that there is evidence for voids showing an ISW-like effect above the standard prediction (e.g. Granett et al., 2008) but at marginal significance and other analyses have found results consistent with standard cosmology (e.g. Nadathur & Crittenden, 2016, Hotchkiss et al., 2015). Kovács & García-Bellido (2016) extended this work to include photometric redshifts from 2MASS (2MPZ) and spectroscopic redshifts from 6dFGS. Using these datasets it was claimed that the under-density detected by Szapudi et al. (2015) extends along the line of sight back to $z \sim 0$ with a void radius of up to 500 h$^{-1}$Mpc. The void was suggested to be elongated in the redshift direction and had a smaller radius of 195h$^{-1}$Mpc in the angular direction. Even with these larger estimates of the $z \approx 0.15$ void's scale the Cold Spot temperature may only be partly explained by the ΛCDM ISW effect. But significant uncertainties remain in the void parameters due to the nature of photometric redshifts, and in order to test claims of divergence from ΛCDM, the parameters of the supervoid must be better determined. The sightline must also be unique in order to explain the uniqueness of the Cold Spot in the CMB.

We have therefore carried out the 2dF-VST ATLAS Cold Spot Redshift Survey (2CSz) over the inner 5° radius core of the Cold Spot in order to test the detection made by Szapudi et al. (2015) and, if the supervoid were confirmed, to measure its parameters to assess any tension with ΛCDM. Throughout the paper we use Planck 2015 cosmological parameters (Planck Collaboration et al., 2016b), with $H_0 = 100$h

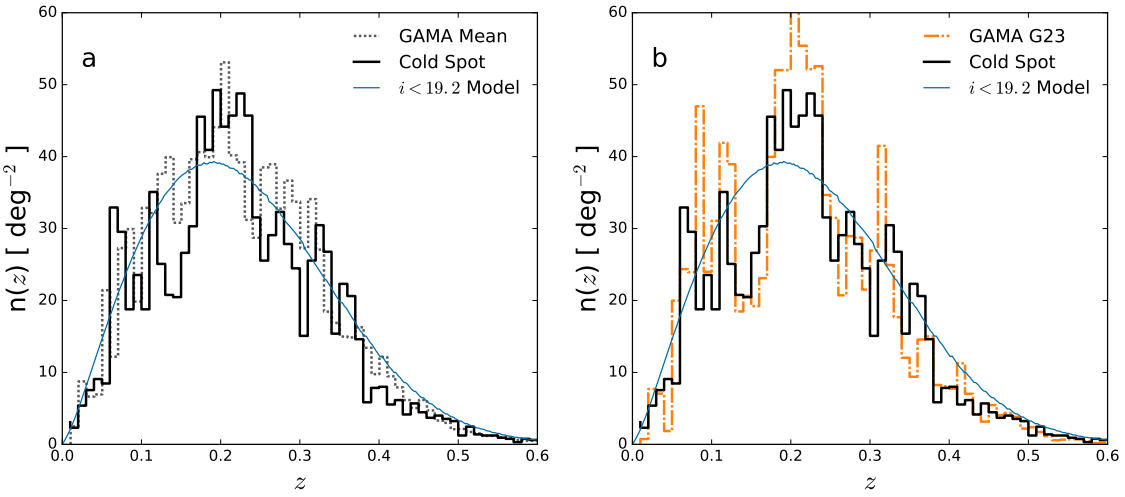


Figure 2.2: (a) The galaxy redshift distribution of the 2CSz (black). Also shown is the $n(z)$ from the average of the 4 GAMA fields at R.A. $\sim$ 9h, 12h, 15h and 23h (G23) at the same $i_{AB} < 19.2$ limit (grey dotted) and the homogeneous model of (Metcalfe et al., 2001) (blue). (b) The galaxy redshift distribution of the 2CSz (black). Also shown is the $n(z)$ from the GAMA G23 field, at the same $i_{AB} < 19.2$ limit (yellow dot-dashed), and the same homogeneous model as in (a) (blue).

kms$^{-1}$Mpc$^{-1}$, $h$=0.677, $\Omega_{M,0}$=0.307, $T_{cmb,0}$=2.725 K.

## 2.2 Survey and Data Reduction

The first goal of 2CSz was to probe the supervoid of Szapudi et al. (2015) with spectroscopic precision. We therefore targeted the inner 5° radius with 20 contiguous 2dF fields (see Fig. 1). A further 2 fields were targeted at larger radii in the sightlines of two $z \sim 0.5$ quasars, which, in other work, will be used with HST COS spectra to probe the void structure in the Lyman $\alpha$ forest as well as in the galaxy distribution. In all fields, 2dF galaxies were sampled at a rate of $\sim$110 deg$^{-2}$ to a limit of $i_{AB} < 19.2$. The survey was selected analogously to the GAMA G23 survey[1], but sub-sampled to the number density matched to a single 2dF pointing per field ($\sim$ 1/8 sampling). This provided us with a highly complete control field.

[1] The Galaxy And Mass Assembly (GAMA) survey (Driver et al., 2009, Driver et al., 2011) includes 3 Equatorial fields at R.A. $\sim$ 9hrs, 12hrs and 15hrs, each covering about 60 deg$^2$, highly spectroscopically complete to $r_{AB} < 19.8$. There is also one SGC field (G23) covering 50 deg$^2$ similarly complete to $i_{AB} = 19.2$ (Liske et al., 2015).

The imaging basis for this spectroscopic survey was the VLT Survey Telescope (VST) ATLAS (Shanks et al., 2015), an ongoing ∼4,700deg$^2$ *ugriz* survey at high galactic latitude over the two sub-areas in the North and South Galactic Caps (NGC and SGC respectively), the latter of which includes the Cold Spot region. VST ATLAS reaches an $i$ band $5\sigma$ depth of 22.0 AB mag for point sources and has a median $i$ band seeing of $0.81''$, allowing clean star-galaxy classification to our magnitude limit. The main selection criterion was to select extended sources with $i_{Kron,AB} \leq 19.2$ where $Kron$ indicates a pseudo-total magnitude with the usual definition. Additional quality control cuts were applied to the data to ensure the removal of stars and spurious objects from the galaxy catalogue. Although the extended source classification removes most stars, an additional star-galaxy cut was applied ($i_{Kron,AB} - i_{ap3,AB} < 0.1 \times i_{Kron,AB} - 1.87$) where $i_{ap3,AB}$ denotes the magnitude corresponding to the flux within a $2''$ diameter aperture (c.f. Fig. 22 of Shanks et al., 2015). To reject spurious objects (e.g. ghosts around bright stars), sources without $z$ band detections were rejected, as were objects near Tycho-2 stars at radii calibrated to VST ghosts. Additionally, a cut of SKYRMS $\leq 0.2$ ADU was applied to the RMS of the sky measurement for each source in the catalogue to remove further artefacts. These cuts were validated with GAMA G23. All magnitudes were corrected for Galactic extinction (Planck Collaboration et al., 2014b).

The spectroscopic survey was completed in 22 2dF fields with 20 covering the inner 5° radius of the Cold Spot. The survey footprint is shown in Fig. 2.1. 2dF covers a 3 deg$^2$ area with approximately 392 fibres, ∼ 25 of which were used as sky fibres. The number density of selected galaxies was 722 deg$^{-2}$, further randomly sampled down to ∼ 200deg$^{-2}$ in order to provide sufficient targets to utilise all fibres. Many targets cannot be observed due to limitations in positioning of the fibres to avoid fibre collisions and to limit fibre crossings. This down-sampled target list was finally supplied to the 2dF fibre allocation system *Configure*.

The spectroscopic observations were carried out in visitor mode on $16^{th}$, $17^{th}$ and $18^{th}$ of November 2015, during grey (Moon phase) conditions with typical seeing of

$\sim 2.0''$. We observed using AAOmega with the 580V and 385R gratings and the 5700Å dichroic. This gives a resolution of $R \sim 1300$ between 3700Å and 8800Å. Each field was observed with $3 \times 15$ minute exposures; flats and an arc frame were also taken with each plate configuration. Fields observed at high airmass at the beginning and end of the night had additional 15 minute exposures where possible. Dark and bias frames were taken during the day before and after each night.

Spectroscopic observations were reduced and combined using the $2dFdr$ pipeline (Croom et al., 2004, Sharp & Parkinson, 2010). The data were corrected with the fibre flat and median sky subtracted. Dark frames were not ultimately used as on inspection they did not improve the data quality. The sky correction parameters used were throughput calibration using sky lines, iterative sky subtraction, telluric absorption correction and PCA after normal sky subtraction. The resulting reduced spectra were then redshifted manually using the package *runz* (Colless et al., 2001). Redshifts were ranked in quality from 5 (Template quality), 4 (Excellent), 3 (Good), 2 (Possible) and 1 (Unknown redshift). Only redshifts of quality 3 or greater were used in the final science catalogue. Typically excellent quality redshifts had multiple strong spectra features (e.g. H$\alpha$, [OII] and Ca II K and H lines) and good redshifts contained at least one unambiguous feature. Overall the redshift success rate was approximately 89% ranging from 71% to 97%; typically the success rate is a strong function of the phase and position of the Moon.

With an 89% success rate, incompleteness will have only a small effect if redshift failures are random rather than systematic and we modeled this with GAMA G23. To test what effect magnitude dependent completeness could have on these results we measured the completeness with magnitude for our survey, finding that completeness is $\sim 96\%$ for $i_{AB} \leq 18.2$ and decreases to $\sim 82\%$ for $18.7 < i_{AB} \leq 19.2$. This magnitude dependent completeness will bias the $n(z)$ towards the redshift distribution of the brighter galaxies. To estimate the effect this has on the $n(z)$ we weight the GAMA G23 $n(z)$ with the completeness as a function of magnitude from 2CSz. Taking the ratio of the weighted and unweighted $n(z)$ we obtain the completeness

| | |
|---|---|
| Main Selection | $i \leq 19.2$ |
| Area | 66 deg$^2$ |
| Number of Galaxies* | 6879 |
| Completeness | 89% |
| Redshift Range | $0.0 \leq z \leq 0.5$ |
| Galactic coordinates $(l, b)$ | (209, -57) |

* Galaxies with redshift quality $\geq 3$

Table 2.1: 2CSz survey parameters.

fraction as a function of redshift, $f(z) \approx 0.95 - 0.232z$ for $z < 0.45$. This linear modulation of the $n(z)$ does not significantly affect the results but this analysis assumes that redshift failures depend only on the magnitude of the object and not the redshift. We do not apply a correction to the data as we do not believe this assumption holds (see Section 2.4.1).

## 2.3 Results

### 2.3.1 Redshift Distributions

The 2CSz redshift distribution of the ∼6879 quality $> 2$ galaxies is shown in Fig. 2.2(a), along with the mean GAMA redshift distribution and a homogeneous model (Metcalfe et al., 2001). Comparison with the homogeneous model allows for under and over-densities to be identified. Due to the sub-sampling of the spectroscopic survey we normalised the $n(z)$'s to the galaxy number magnitude counts in the Cold Spot and G23 regions using an ATLAS $iz$ band-merged catalogue. We found that the 75deg$^2$ Cold Spot area was $16 \pm 3\%$ under-dense relative to the $\sim$ 1000deg$^2$ around G23. We also found that the Cold Spot had a $7.4 \pm 0.7\%$ number density deficit relative to a similarly large $\sim$ 1000deg$^2$ region surrounding the Cold Spot whereas the G23 galaxy count was consistent with the SGC average over its full $\sim$ 2600deg$^2$ area. Both the SGC number count and the mean galaxy density averaged over the 4 complete GAMA fields are in good agreement with the homogeneous model. To allow comparison with G23 we chose to normalise the Cold Spot observed $n(z)$ by

7.4% lower in total counts than both the homogeneous model and the G23 observed $n(z)$ and this is what is shown in Fig. 2. Ignoring the large scale gradient like this is certainly correct if it is a data artefact. But there is also a case to be made for it even if it is real since the Cold Spot is essentially a small-scale, $\sim 75\text{deg}^2$, feature rather than a $\sim 1000\text{deg}^2$ feature.

Here and throughout field-field errors are used. These are based on a (2dF) field size of $\sim 3$ deg$^2$.

The mean GAMA redshift distribution comes from the 4 GAMA fields, G23, G09, G12 and G15 selected with $i_{AB} \leq 19.2$. The latter three $r$-limited fields were checked to be reasonably complete at the $i_{AB} \leq 19.2$ limit for this analysis. The stacked GAMA redshift distribution fits well with the Metcalfe et al. (2001) homogeneous model for galaxies with $i_{AB} \leq 19.2$. Fig. 2.2(a) shows indications of inhomogeneity in the Cold Spot sightline where we see evidence of an under-density spanning $0.08 < z < 0.17$ and there is also a hint of a smaller under-density at $0.25 < z < 0.33$. This would be consistent with the Szapudi et al. (2015) supervoid but we also see evidence for an over-density at $0.17 < z < 0.25$, apparently in conflict with the previous claim that the supervoid was centred in this range. Given the photometric redshift error, there may be no real contradiction between the datasets but their single void model does appear inconsistent with our spectroscopic data (see Section 2.4.2). Another under-density is seen at $0.37 < z < 0.47$ but systematic errors, such as spectroscopic incompleteness, become more important at this point (see Section 2.4.1).

### 2.3.2 Void Model

In order to obtain the parameters of an under-density and determine its ISW imprint a void profile must be selected and fit to the redshift distribution. Some previous work has used simple top-hat void models as the measured profile was dominated by photo-$z$ error. In the case of our well sampled spectroscopic survey the structure

of the void is important to the fitting and allows us to estimate the ISW imprint of any void. Following Kovács & García-Bellido (2016), we have chosen the ΛLTB void profile described by a Gaussian potential (i.e. $\alpha = 0$ in Finelli et al., 2016, eq. 2.3.1) which will allow us to use the analytic expression for the ISW temperature profile given by these authors. This compensated void profile is described by eq. 2.3.1,

$$\delta_m(r) = \delta_0 \; g(a) \left(1 - \frac{2}{3}\frac{r^2}{r_0^2}\right) \exp\left(-\frac{r^2}{r_0^2}\right), \tag{2.3.1}$$

where $\delta_m(r)$ is the matter density contrast at radius $r$ from the void centre, $\delta_0$ is the matter density contrast at the void centre, $g(a)$ is the growth factor at scale factor, $a$, and $r_0$ is the void radius. As shown by Finelli et al. (2016) the ISW imprint of a void described by eq. 2.3.1 can be calculated using eq. 2.3.2,

$$\begin{aligned}\frac{\delta T}{T}(\theta) \approx \frac{3\sqrt{\pi}}{22}\frac{H(z_0)\Omega_\Lambda F_4(-\Omega_\Lambda/\Omega_M(1+z_0)^3)}{H_0(1+z_0)^4 F_1(-\Omega_\Lambda/\Omega_M)}\times \\ \left(1 + \mathrm{erf}\left(\frac{z_0}{H(z_0)r_0}\right)\right)\delta_0 (H_0 r_0)^3 \exp\left[-\frac{r^2(z_0)}{r_0^2}\theta^2\right],\end{aligned} \tag{2.3.2}$$

where $\frac{\delta T}{T}(\theta)$ is the ISW temperature imprint at angle $\theta$ away from the centre of the void and $z_0$ is the central redshift of the void. $F_1$ and $F_4$ are described by eq. 2.3.3 and 2.3.4 respectively where ${}_2F_1$ is the hypergeometric function,

$$F_1 = \; {}_2F_1 \left[1, \frac{1}{3}, \frac{11}{6}, \frac{-\Omega_\Lambda a^3}{\Omega_M}\right], \tag{2.3.3}$$

$$F_4 = \; {}_2F_1 \left[2, \frac{4}{3}, \frac{17}{6}, \frac{-\Omega_\Lambda a^3}{\Omega_M}\right]. \tag{2.3.4}$$

Finelli et al. (2016) also give an expression for the Rees-Sciama effect (Rees & Sciama, 1968), the second order ISW effect. As the Rees-Sciama effect is sub-dominant to the ISW effect at the scale of the CS at low redshift in the standard cosmology (Cai et al., 2010) , we will neglect it's contribution in our calculations.

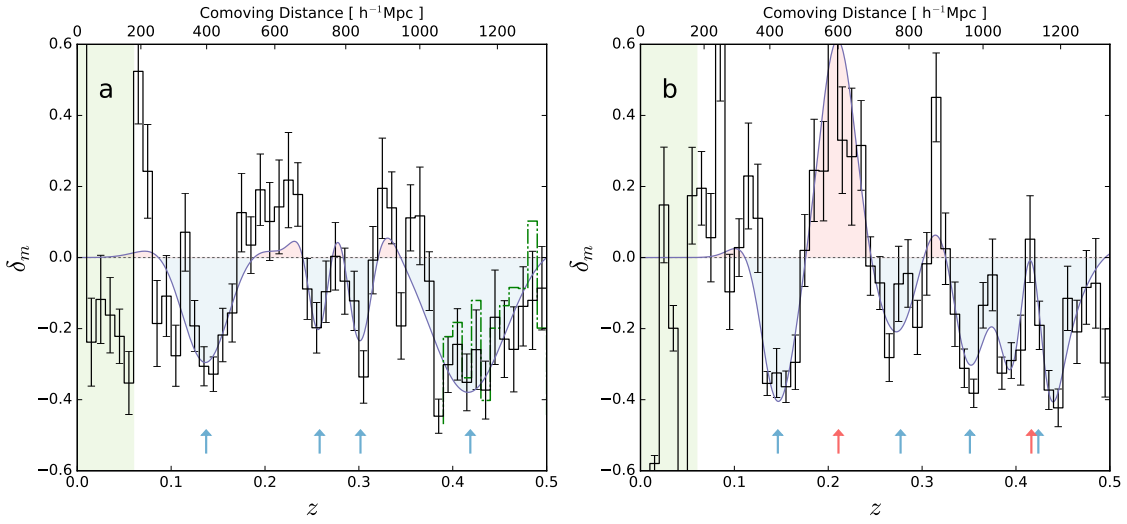


Figure 2.3: (a) The matter density contrast for 2CSz (black histogram), the best-fit void models (dark blue) and the 'Local Hole' extent (green), modelled under-densities are filled in blue and overdensities in red. The dashed line shows the result at $z > 0.38$ when only 2dF fields with $> 90\%$ redshift success rate are used. Arrows indicate the centre of each fitted under-density (blue) and over-density (red). (b) The mass density contrast for the GAMA G23 with symbols as in (a).

### 2.3.3 Perturbation Fitting in the Cold Spot

In order to estimate the ISW imprint of the observed inhomogeneities we have fitted the redshift distribution with compensated perturbations with the profile given by eq. 2.3.1. Although our spectroscopic survey has 3D information we pursue this 1D analysis to mimic the void finding used in past photo-$z$ analyses, so the same large under-densities are selected. In order to do this it was first necessary to transform the $n(z)$ to the matter density contrast, $\delta_m(z)$, done by first converting to the galaxy density contrast, $\delta_g(z)$, and then dividing by the galaxy bias, $b_g$. These transformations are shown in eq. 2.3.5 and 2.3.6,

$$\delta_g(z) = \frac{n(z)}{n_{model}(z)} - 1 \qquad (2.3.5)$$

$$\delta_m(z) = \frac{\delta_g(z)}{b_g} \qquad (2.3.6)$$

where $n_{model}(z)$ is the predicted redshift distribution from the homogeneous model (Metcalfe et al., 2001). Since the magnitude limits for the 2CSz and G23 galaxies are the same, the bias for both samples can be estimated from the GAMA G23

correlation function, obtaining a linear bias of $b_g = 1.35$. Although simplistic, this linear bias assumption is accurate enough for the large scales of interest here.

Fig. 2.3(a) shows the matter density contrast for the 2CSz survey, assuming field-field errors. A number of features can be seen in Fig. 2.3(a). At the lowest redshifts ($z < 0.06$) the 'Local Hole' can be seen as a $\sim 25\%$ under-density. This is well studied in the literature and seems to extend across the SGC (e.g Whitbourn & Shanks, 2014). At $z = 0.06$ there is an over-density separating the 'Local Hole' from a $\sim 40\%$ under-density which extends to $z = 0.17$. Another peak in the distribution is followed by two under-densities ($z = 0.23$, 0.25 and 0.3 respectively). Lastly there is a clear break at $z = 0.38$ and a $\sim 30\%$ under-density extending to $z = 0.5$ where it converges towards the homogeneous model. This feature may be due to redshift dependent incompleteness as we will discuss later (see Section 2.4.1).

In order to fit the redshift distributions in an unbiased way we have adopted an iterative fitting procedure that minimises the necessary complexity of any fit, quantified with the Akaike information criterion (AIC) (e.g. Porciani & Norberg, 2006). The AIC statistic takes into account the improvement in the fit of a more complex model but additionally penalizes it for this increased complexity. We use the AIC statistic specifically because it can be corrected in the case when the number of data points is not much larger than the number of parameters. We have fitted individual under-densities, $\delta_m(r)$, with 3D perturbations described by eq. 2.3.1, averaged over the 5 deg radius of 2CSz. In order to describe the features seen in Fig. 2.3 we model the line of sight $n(z)$ as a combination of perturbations. The fitting assumes the void is centred on the Cold Spot. The whole redshift range was fitted simultaneously, with the 'Local Hole', at $z \leq 0.0625$ excluded from the fit as it is not unique to the Cold Spot. We do not believe this will affect our results as there is a clear over-density, which appears to be a wall, separating the Local Hole and the lowest redshift void we consider. Our iterative method initially assumed N perturbations seeded with random parameters and fitted them to the data. Fitting was carried out with a Levenberg-Marquardt algorithm and quoted errors are standard errors

calculated from the covariance matrix. Iterating over new random values and fitting we converge on the best fit parameters for N perturbations. The best fits for each value of N were then compared via the corrected AIC statistic, the minimum of which gave the optimum fit and the relative likelihood allowed for other values of N to be rejected if significantly poorer. The corrected AIC statistic is given by eq. 2.3.7 where $k$ is the number of parameters being fit, $N_{data}$ is the number of data points and $\hat{L}$ is the maximised likelihood function,

$$\begin{aligned} AIC &= 2k - 2ln(\hat{L}) + \frac{2k(k+1)}{N_{data} - k - 1} \\ &= 2k + \chi^2 + \frac{2k(k+1)}{N_{data} - k - 1}. \end{aligned} \tag{2.3.7}$$

The second line of eq. 2.3.7 holds in the case of normally distributed residuals. The relative probability of one model over another with a greater AIC value is given by the Akaike weights (eq. 2.3.8) where $AIC_{min}$ is the minimum AIC, $\Delta AIC_i = AIC_i - AIC_{min}$ and $k_{max}$ is the maximum k considered,

$$w_i = \frac{e^{-\Delta AIC_i/2}}{\sum_{k=1}^{k_{max}} e^{-\Delta AIC_k/2}}. \tag{2.3.8}$$

Hence a $p = 0.05$ rejection of the weaker model corresponds to a $\Delta AIC \sim 6$ and we shall adopt $\Delta AIC = 6$ as a threshold for rejecting models over the best fit. More complex models were considered until one was rejected over a simpler model.

This analysis suffers from degeneracies in that we cannot discern the difference between two voids and a wide void with an interior, narrow over-density. For this reason, the fitting ranges of parameters were restricted in range to provide sensible fits. Specifically we restricted the void radius to be between 50 and 150 $h^{-1}$Mpc and the central density contrast was constrained to lie in the physical range, $\delta_0 \geq -1$. Parameters at the radius limits were individually re-fit. Fits were also rejected which had perturbations at the very edges of the fitting range, i.e. $z_0 < z_{min} + 0.01$ or $z_0 > z_{max} - 0.01$. Additionally the compensated profile we have adopted cannot

| $N$ | $k$ | $AIC_{min}$ | |
|---|---|---|---|
| | | Cold Spot | G23 |
| 1 | 3 | 248.85 | 441.97 |
| 2 | 6 | 147.17 | 240.14 |
| 3 | 9 | 131.65 | 197.66 |
| 4 | 12 | (123.91) | 154.64 |
| 5 | 15 | 125.18 | 151.07 |
| 6 | 18 | 132.33 | (141.45) |
| 7 | 21 | - | 149.86 |

Table 2.2: The minimum AIC values for each value of N perturbations for the Cold Spot and G23. $k$ is the number of free parameters. The minimum AIC values best fits for are shown in parenthesis.

describe sharp narrow under-densities as they are averaged out in the survey field; however the purpose of this analysis is to detect large voids and this places upper limits on ISW contributions. We have allowed over-densities to be fitted with the perturbation described by eq. 2.3.1 but as this profile was derived for voids the resulting $\delta T$ values should be treated with caution. The minimum AIC values for each value of N perturbations are shown in Table 2.2. The resulting best fits are shown in Fig. 2.3. For the Cold Spot, the iterative procedure selected N= 4 perturbations as the best fit (all under-densities) to give the fits summarized in Table 2.3. The AIC test does not strongly reject the N= 5 solution but we note the difference between the models is only in the fitting of the $z \sim 0.42$ void with one profile or two and the resulting total $\delta T$ differs by just $2.7\mu K$ which is not significant.

### 2.3.4 Perturbation Fitting in GAMA G23

As noted above, we originally planned to use GAMA G23 as a control field but analysis showed that even on 50 deg$^2$ scales there was sufficient sample variance to merit using a model which we validated with the stacked $i_{AB} \leq 19.2$ $n(z)$ from all four GAMA fields with a combined area of $\sim$240deg$^2$. Indeed, Fig. 2.2(b) shows that upon comparison the Cold Spot redshift distribution bears remarkable similarity with G23 in the under-densities at $z \sim 0.15$, 0.3 and 0.4. In particular, the significant

| | $z_0$ | $r_0$ ($h^{-1}$Mpc) | $\delta_0$ | $\delta T(\theta = 0)$ ($\mu$K) |
|---|---|---|---|---|
| Cold Spot | | | | |
| | 0.14±0.007 | 119±35 | -0.34±0.08 | -6.25±5.7 |
| | 0.26±0.004 | 50±13 | -0.87±0.12 | -1.02±0.8 |
| | 0.30±0.004 | 59±17 | -1.00+0.72 | -1.80±2.1 |
| | 0.42±0.008 | 168±33 | -0.62±0.16 | -22.6±14.7 |
| G23 | | | | |
| | 0.15±0.004 | 82±33 | -0.49±0.17 | -2.92±3.7 |
| | 0.21±0.006 | 88±21 | +0.89±0.35 | +6.09±5.1 |
| | 0.28±0.007 | 85±29 | -0.36±0.24 | -2.06±2.6 |
| | 0.35±0.006 | 74±22 | -1.00+0.10 | -3.40±3.1 |
| | 0.42±0.005 | 150±20 | -0.63±0.13 | -16.1±7.4 |
| | 0.42±0.002 | 50±5 | +4.16±1.6 | +3.96±2.0 |

Table 2.3: Best fit 3-D ΛLTB parameters for compensated perturbations (eq. 2.3.1) estimated from the Cold Spot and GAMA G23 density contrast profiles in Fig. 2.3. The central temperature decrement, $\delta T$, predicted from the ISW effect is also given.

under-density $0.35 < z < 0.5$ that occurs in both fields could point to a selection effect in the survey. However, the mean GAMA redshift distribution shown in Fig. 2.2(a) shows little evidence for this. It also raises the question of whether or not some of these features could be coherent between G23 and the Cold Spot. Certainly at the lowest redshifts of $z < 0.05$ the under-density is consistent with the 'Local Hole' which spans the SGC (Whitbourn & Shanks, 2014). In Section 2.4.3 we shall use the 2dF Galaxy Redshift Survey (2dFGRS, Colless et al., 2001), whose Southern Strip spans the SGC between GAMA G23 and the Cold Spot, to check if this apparent coherence is real or accidental.

Meanwhile, the $n(z)$ similarities open up the possibility of G23 still acting as a control field because it does not show a CMB Cold Spot. Therefore due to the similarities in the redshift distributions of G23 and 2CSz we have fitted the density contrast in the same way as the Cold Spot as shown in Fig. 2.3(b). The parameters of the best fit are summarised in Table 2.3 with N=6 perturbations selected (see Table 2.2), including 4 under-densities and 2 over-densities. We note that the highest redshift feature has been fitted with an under-density with an interior, narrower over-density

which together fit the two $z > 0.37$ under-densities seen in Fig. 2.3(b). The fitting procedure selects this over two under-densities because the under-densities are sharp and the density profile provides a poor fit individually. As we will discuss in Section 2.4.1 we believe these features are affected by systematics and therefore we did not re-fit them.

## 2.4 Discussion

We have detected three large under-densities along the CMB Cold Spot sightline, the largest with radius $r_0 = 119 \pm 35 \mathrm{h}^{-1}\mathrm{Mpc}$ centred at $z_0 = 0.14$ with a central density contrast of $\delta_0 = -0.34$. This supervoid is smaller but more under-dense than that proposed by Szapudi et al. (2015) which has $r_0 \sim 220 \mathrm{h}^{-1}\mathrm{Mpc}$ and $\delta_g = -0.25$. The Szapudi et al. (2015) void also has a higher central redshift at $z \sim 0.22$ and may include the other 2CSz voids at $z_0 = 0.26$ and $z_0 = 0.30$ (see Table 2.3), seen as a single supervoid due to the photo-$z$ errors. Kovács & García-Bellido (2016) drew upon additional datasets to suggest that the proposed supervoid extended back to zero redshift with radius $500 \mathrm{h}^{-1}\mathrm{Mpc}$ and with a smaller $195 \mathrm{h}^{-1}\mathrm{Mpc}$ radius in the angular direction. From eq. 2.3.2 we estimate the central temperature decrement due to our $z = 0.14$ void at $-6.25 \pm 5.7 \mu K$, small compared to some previous work (Kovács & García-Bellido (2016)), as expected due to the strong relationship between void radius and its ISW imprint. The combined ISW imprint of the three Cold Spot voids is $-9.1 \pm 6.1 \mu K$ and even adding the fourth questionable void this rises to just $-31.7 \pm 15.9 \mu K$. As we will discuss in section 2.4.1 we believe the $z = 0.42$ void is exaggerated by systematics. We also note that these estimates of the ISW imprint depend on the chosen void density profile used in the fitting process. Although the profile used here (eq. 2.3.1) is not unique it is at least representative of what previous studies have done and allows for direct comparison with literature (e.g. Kovács & García-Bellido, 2016, Finelli et al., 2016).

The strongest evidence against an ISW explanation for the Cold Spot that may

arise from our results is due to the similarity in the $n(z)$ between GAMA G23 and the Cold Spot. Despite this, G23 has no CMB Cold Spot. Indeed, the predicted central ISW decrement for G23 from summing the contributions in Table 2.3 above (excluding the features at $z > 0.4$) is $-3.6 \pm 7.5\mu K$, statistically consistent with the $-9.1 \pm 6.1\mu K$ predicted similarly for the Cold Spot. The predicted central ISW decrement for G23 is also consistent with that of the Cold Spot, even if no features in Table 2.3 are excluded. However, the CMB in the G23 sightline shows only a small central temperature decrement of $-15.4 \pm 0.3\mu$K, some $\sim 10\times$ lower than for the Cold Spot. Thus the similarity in the large-scale structure between G23 and the Cold Spot fields forms a further qualitative argument against foreground voids playing any significant role in explaining the Cold Spot. On this evidence alone the detected void cannot explain the CMB Cold Spot because a similar void in G23 has no such effect.

### 2.4.1 The reality of the $z = 0.42$ void

In the Cold Spot $n(z)$ an apparent, relatively strong, void can be seen at $0.37 < z < 0.5$ but we have already noted this is in a range where not only are the statistics poorer but where we know that magnitude dependent incompleteness becomes more important. The similarity of this feature with the $0.34 < z < 0.5$ under-densities in G23 suggests there may be some sort of selection effect or systematic which we will now investigate.

We therefore test the reality of this void in Fig. 2.4 where we compare the 2CSz $\delta_m(z)$ and the previous Bremer et al. (2010) VLT VIMOS $\delta_m(z)$ and see that an under-density at $z = 0.42$ may also be detected in that dataset, albeit at low $\sim 2\sigma$ significance. A lower bias of $b = 1$ has also been assumed here for the VIMOS $\delta_m(z)$ compared to $b = 1.35$ for 2CSz, on the grounds that the VIMOS galaxies are intrinsically fainter. This is consistent with results from the VVDS survey (Marinoni et al., 2005). We note that despite this apparent agreement the VLT VIMOS data

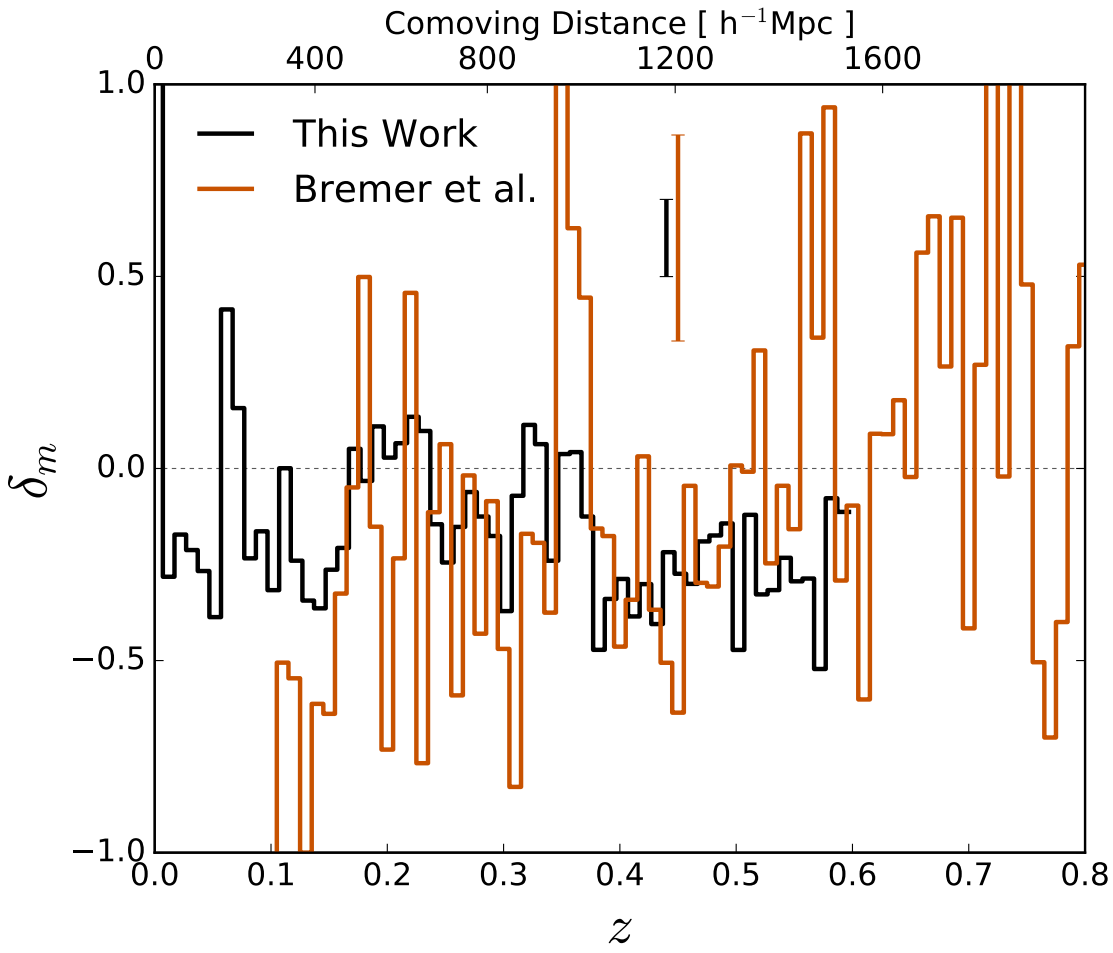


Figure 2.4: The 2CSz $\delta(z)$ (black) compared with VLT VIMOS (Bremer et al., 2010) $\delta(z)$ (orange) to test the reproducibility of the Cold Spot void at $z = 0.42$. Here a bias of $b = 1.35$ has been assumed for the 2dF $\delta(z)$ and a bias of $b = 1$ has been assumed for VLT VIMOS $\delta(z)$. Typical errors are plotted above the lines, Poisson errors are assumed for the VIMOS data.

probes a much smaller volume at this low redshift end and therefore would have large sample variance.

The absence of this feature from the mean GAMA $n(z)$ indicates that this feature cannot be intrinsic to the $i_{AB} < 19.2$ selection criteria. We instead suggest that it may be due to a systematic selection effect. Although the other GAMA fields are apparently unaffected by this systematic, this may be explained by the Cold Spot (and G23) data having slightly lower S/N due to somewhat shorter 2dF exposure times and redshift success rate viz. Cold Spot (45mins, 89.0%), G23 (30-50mins, 94.1%) vs. the other 3 Equatorial GAMA fields used here (50-60 mins, 98.5%). 2CSz was also conducted in gray time which will further reduce the S/N with respect to GAMA. The lower S/N ratio will increase spectroscopic incompleteness and we note that the 4000Å break and Ca II H and K absorption lines transition through the dichroic over this redshift range while the H$\alpha$ emission line also leaves the red arm of the spectrograph. It is possible that these two effects make accurate redshifting more difficult over this redshift range and would create an apparent under-density. To test this we split 2CSz into pointings with high and low spectroscopic success rate, with half having a success rate greater than 90% and half with less. The result of this is shown for $z \geq 0.38$ in Fig. 2.3(a) by the dashed histogram. All fitting used the full dataset. The success rate of the 2dF field strongly affects the depth of the $z = 0.42$ void indicating that it is affected by systematic incompleteness.

Also, at $z > 0.4$, small differences in the homogeneous model will lead to large differences in the derived $\delta_m(z)$. To investigate whether the model $n(z)$ could be over-predicting the galaxy density at the higher redshifts creating spurious under-densities, we have explored a model $n(z)$ constructed from random catalogues built for the GAMA survey (Farrow et al., 2015) and find that indeed this different model $n(z)$ decreases the depth of the $z = 0.42$ void. When compared to the mean GAMA $n(z)$ however this model $n(z)$ appears to under-predict the galaxy density at higher redshift and therefore we do not replace our homogenous model with the GAMA random catalogue constructed $n(z)$. Whether the void seen by 2CSz in this $z$ range

is accentuated by such systematics or not does not matter for our main conclusion since even including this void's contribution the total ISW decrement from Table 2.3 is still only $\sim -32\mu$K compared to the $\sim -150\mu$K needed to explain the Cold Spot.

Additionally we note that the bias of galaxies will not be constant throughout the redshift range as assumed. Because the survey is magnitude limited the galaxies at the high redshift end of the survey will be brighter than the low redshift end. The brighter 2CSz galaxies at $z = 0.42$ may actually be as large as $b \sim 2$ (e.g. Zehavi et al., 2011) and increasing the bias would linearly decrease the depth of the void $\delta_0$ (by eqn 2.3.6) and hence its ISW imprint.

Together these arguments cast doubt on the existence of the $z = 0.42$ void and for this reason we neglect it in our conclusions. A sample of galaxies with a magnitude limit intermediate between that of 2CSz and Bremer et al. (2010) et al is needed to determine finally the status of the $z = 0.42$ void.

### 2.4.2 Photo-$z$ and spectroscopic $n(z)$

In order to assess why the spectroscopic 2CSz survey results apparently differ from the photometric redshift survey of Kovács & García-Bellido (2016) we convolved the 2CSz spectroscopic redshift distribution with an estimated error of $0.034(1 + z)$ photo-$z$ error, which is the quoted photo-$z$ error from Szapudi et al. (2015). Kovács & García-Bellido (2016) used 2MPZ with a very small photo-$z$ error of $0.015(1 + z)$, but the 2MPZ sample is limited by low number densities at higher redshifts so we do not compare to this directly.

The resulting model $\delta_m(z)$ is shown in Fig. 2.5 where we see that there is limited consistency with the model result of Kovács & García-Bellido (2016) with $r_0 = 500\mathrm{h}^{-1}\mathrm{Mpc}$ and $\delta_0 = -0.25$ when convolved with a photo-$z$ error. The main source of disagreement is the lack of an under-density at $z \sim 0.2$ in 2CSz which seems difficult to reconcile with the model void but we note that at $z > 0.15$ the 2MPZ data is consistent with no under-density due to a large uncertainty. While our data

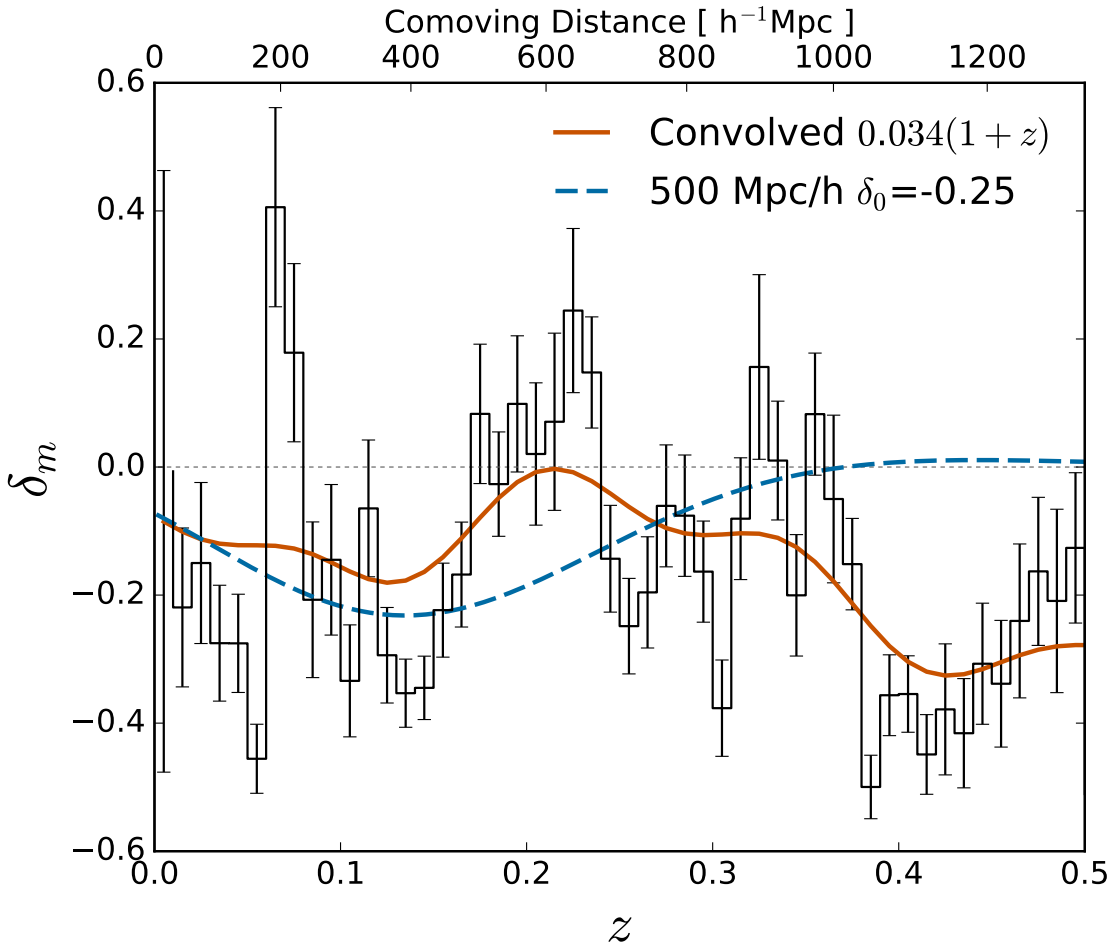


Figure 2.5: The 2CSz $\delta_m(z)$ (black), the 2CSz $\delta_m(z)$ convolved with the photo-$z$ error of the PanSTARRS data of Szapudi et al. (2015) (orange) and compared to the fitted $\delta_m(z)$ model of Kovács & García-Bellido (2016) (blue)

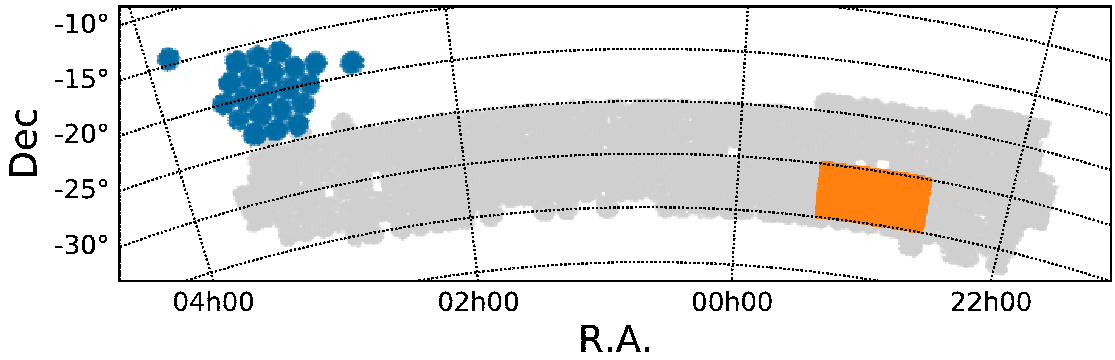


Figure 2.6: The position of the 2dFGRS SGC strip (grey) relative to the 2CSz 2dF fields (blue) and the GAMA G23 area (orange).

is not consistent with an $r_0 = 500\mathrm{h}^{-1}\mathrm{Mpc}$ void we believe it is consistent with the photo-$z$ data.

When we compare our predicted ISW central decrement to previous work we see some consistency. With the 3-void model of the Cold Spot line of sight the combined temperature decrement is $-9.1 \pm 6.1\mu K$ which is consistent with the $\sim -20\mu K$ of Szapudi et al. (2015) but not with the $\sim -40\mu K$ of Kovács & García-Bellido (2016). One could argue the 4-void model at $-31.7 \pm 16.0\mu K$ is consistent with Kovács & García-Bellido (2016) values, but $\sim \frac{2}{3}$ of that decrement is due to the $z = 0.42$ void which is likely to be contaminated by systematic effects as discussed previously. Additionally the void of Kovács & García-Bellido (2016) did not extend to $z > 0.4$ and it is beyond the range of the 2MPZ data.

### 2.4.3 A coherent SGC galaxy distribution?

We have already discussed the important question of the normalisation of the Cold Spot $n(z)$. Both G23 and the Cold Spot areas are contained in the Local Hole under-density known to extend at least to $z = 0.06$ across the SGC. Moreover we have noted that the galaxy count in the 5° radius Cold Spot area is $\sim 16\%$ under-dense relative to G23 and the rest of the SGC at our $i_{AB} < 19.2$ limit. When compared to a surrounding $\sim 1000\mathrm{deg}^2$ area the 5° core of the Cold Spot is 7.4% under-dense. The Cold Spot area therefore appears to exist in an environment exhibiting a significant

global gradient stretching across the SGC. Finally we have noted the similarity of the 2CSz and GAMA G23 redshift distributions which again may suggest evidence for coherent structure extending between them.

To investigate further this possibility, we now exploit the 2dF Galaxy Redshift Survey (2dFGRS, Colless et al., 2001) which spans the SGC between GAMA G23 and the Cold Spot at $-35° < \mathrm{Dec} < -25°$ (see Fig. 2.6). With a magnitude limit $b_J(\sim g) \leq 19.6$, 2dFGRS is shallower than the $i_{AB} \leq 19.2$ surveys so only probes the low $z$ structures but has a large area. Busswell et al. (2004) shows the redshift distribution of the 2dFGRS survey in the SGC in their Fig. 14 (also shown in Norberg et al., 2002 Fig. 13). The distribution shows peaks at $z = 0.06$, $z = 0.11$ and $z = 0.21$ which are very similar to those shown in 2CSz and roughly similar to those shown in G23. We have attempted to track these features across 2dFGRS to see if they do in fact span the sky between G23 and 2CSz. When we split 2dFGRS by R.A. as in Fig. 2.7 we generally see coherence in that at $z < 0.06$ we consistently see under-density in this range. This is the 'Local Hole' of Whitbourn & Shanks (2014) (see their Fig. 2b) which covers $\sim 3500$ deg$^2$ of the SGC (the 6dFGS-SGC area marked in orange in their Fig. 1 with coordinate ranges given in their Table 3). Based on the $0.06 < z < 0.11$ void seen in the 2dFGRS $n(z)$ shown in Fig. 14 of Busswell et al. (2004), these authors have speculated that the void runs to $z \sim 0.1$. In passing, we note that the $\sim 8\%$ gradient between the regions surrounding G23 and the Cold Spot may represent Local Hole sub-structure.

In Fig. 2.7 we see that the eastern half of 2dFGRS ($0 < \mathrm{R.A.} < 4$hrs) more clearly exhibits the peaks at $z = 0.06$ and $z = 0.11$ (with intervening under-density) than does the range at $21 < \mathrm{R.A.} < 0$hr. We have checked that restricting 2dFGRS to the G23 area produces very good agreement in $\delta_m(z)$ out to $z < 0.25$. More speculatively, even the $z = 0.21$ peak may be seen in at least some of the R.A. ranges If so, this possible coherence may also explain why 2CSz and G23 have such similar $n(z)$ distributions. However in the $23 < \mathrm{R.A.} < 1$hr and $0 < \mathrm{R.A.} < 2$hr ranges the feature at $z$ =0.21 is less obvious and perhaps argues against coherence extending to

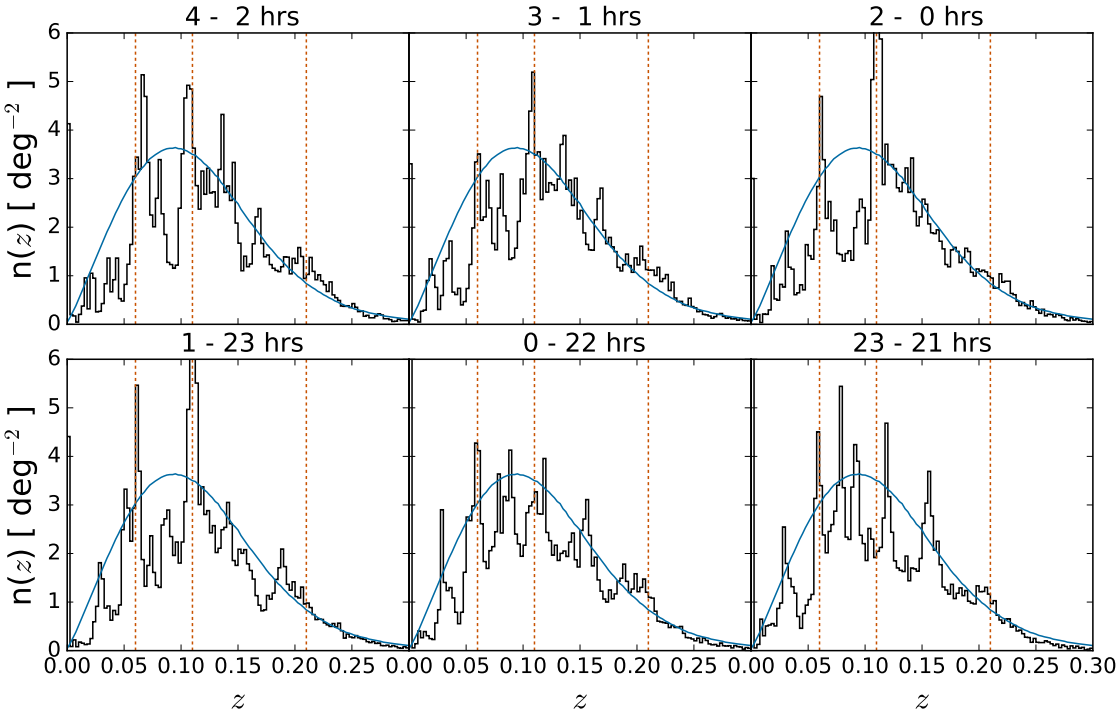


Figure 2.7: The 2dFGRS SGC galaxy redshift distributions, $n(z)$ in overlapping 2hr ranges of R.A. at Dec$\sim -30 \pm 5$deg (black). The homogeneous model prediction of Metcalfe et al. (2001) to the 2dF limit of $b_j = 19.6$ is plotted (blue). The redshifts corresponding to the peaks in the average 2dFGRS $n(z)$ at $z = 0.06$, 0.11 and 0.21 are marked (orange dashed lines).

$z \sim 0.2$. This would leave the similarity of the 2CSz and G23 $n(z)$'s at $0.1 < z < 0.2$ appearing accidental. We note that the absence of these structures from the NGC 2dFGRS survey (c.f. Figs. 13, 14 of Busswell et al., 2004) makes systematic effects unlikely as the cause.

How likely is it, in the standard cosmological model, that coherent structure extends out to $z < 0.2$ across the 2dFGRS SGC strip? We assume an $\sim$1000deg$^2$ area for 2dFGRS SGC and a power-law correlation function, $\xi(s) = (s/s_0)^{-\gamma}$, with $s_0 \sim 6.92\mathrm{h}^{-1}$Mpc and $\gamma \sim 1.51$ for $s < 50\mathrm{h}^{-1}$Mpc, as measured for 2dFGRS by Hawkins et al. (2003). The variance, $\sigma_N^2$, of galaxy numbers, $N$, around average $\bar{N}$ in a volume, $V$, where the galaxy space density, $n$ $(= N/V)$, is (e.g. Peebles, 1980)

$$\sigma_N^2 = \langle (N - \bar{N})^2 \rangle = \bar{N} + n^2 \int_V \xi(s_{12}) dV_1 dV_2, \qquad (2.4.1)$$

implying $\sigma_N \sim 20 \times \sqrt{\bar{N}}$. Given that $\bar{N} \sim 140000$ galaxies in the 2dFGRS SGC volume, a nominal 10% under-density (or over-density) across 2dFGRS SGC even out to $z \sim 0.2$ would amount to a $\sim 1.9\sigma$ fluctuation. On the same assumptions, a similar over- or under-density out to $z = 0.1$ would represent a significance of $\sim 1.3\sigma$. Now these may be taken as a rough measure of the significance of coherence in a survey modeled by some of its $z$ range being 10% overdense and the rest being 10% underdense. So at $\sim 1.3 - 1.9\sigma$, we conclude that galaxy clustering coherence across 2dFGRS SGC can plausibly explain the 2CSz-G23 coherence out to $z \sim 0.1$ and more speculatively to $z \sim 0.2$. However the observational evidence for coherence at $z \sim 0.2$ is mixed.

### 2.4.4 Origin of The CMB Cold Spot

As noted in Section 2.1, several authors have calculated the significance of the Cold Spot with respect to the coldest spots in CMB sky simulations (e.g. Nadathur et al., 2014, Planck Collaboration et al., 2016c). The significances are typically at the $\sim 1\%$ level. As shown by these authors, the significance of the Cold Spot in the standard cosmology comes not from the central temperature but from the temperature profile seen in Fig. 2.8 which closely matches the compensated SMHW that was originally used to detect it (Vielva et al., 2004). On this basis when assessing what impact the detected voids have on the significance of the Cold Spot we have to go beyond central temperature and look at the significance of the SMHW filtered temperature subtracted for the detected voids. This removes the ISW imprinted signal and assesses the significance of the residual primordial profile. Following Naidoo et al. (2016), subtracting our best 3-void (i.e. the voids with $z_0 < 0.4$ in Table 2.3) model ISW contribution would reduce the significance of the Cold Spot only slightly, typically to $\sim 1.9\%$ (Naidoo et al., 2016) i.e. only 1 in $\sim 50$ $\Lambda$CDM Universes would produce such a feature by chance. Fig. 2.8 shows the ISW imprints of the 3 and 4-void models and the measured CMB Cold Spot temperature profile. This significance would be reduced if our 4-void model was trusted but, as previously

argued, the void at $z = 0.42$ may be unduly affected by systematics.

Kovács & García-Bellido (2016) claimed the Cold Spot supervoid is an elongated supervoid at $z = 0.14$ with $r_0 = 500\mathrm{h}^{-1}\mathrm{Mpc}$ in the redshift direction and $r_0 = 195\mathrm{h}^{-1}\mathrm{Mpc}$ in the angular direction with $\delta_0 = -0.25$. The ISW effect on the central decrement is estimated to be a reduction of $\sim 40\mu$K. At the central redshift of $z = 0.14$ this supervoid would extend 27.5° on the sky. We note that the 2dFGRS SGC strip covers the area to the South of the Cold Spot. In the 2h<R.A.<4h range, all of this R.A. bin is within 27.5° of the Cold Spot. Fig. 2.7 shows that although there is a 2dFGRS void at $z = 0.08$ within the supervoid redshift range, the peak at $z = 0.11$ and plateau out to $z = 0.15$ is near the claimed $z = 0.14$ centre of the supervoid; there seems little evidence of a void at $0.1 < z < 0.25$ in this 2dFGRS 2h<R.A.<4h range. The $z = 0.2$ peak may still be present indicating there may be an under-density at $0.15 < z < 0.2$. So at least in the direction South of the Cold Spot, evidence for an extended simple void structure around its centre is again not present.

Various authors (e.g. Cai et al., 2014a,b; Kovács et al., 2017 and references therein) have also discussed the possibility of an enhanced ISW effect in voids being produced by modified gravity models. This has been done to explain observations where a larger than expected ($2 - 4\times$ under ΛCDM) ISW-like signal has been found around voids (Granett et al., 2008, Cai et al., 2017a), these results are however low significance. It may be speculated whether our 2CSz Cold Spot results may also be explained similarly. But again the similarity between the galaxy redshift distributions in 2CSz and the G23 control field tends to argue against this possibility. If some modified gravity model did give an enhanced ISW effect to explain the Cold Spot then why is there no similar Cold Spot seen in the G23 line-of-sight? This argument should be tempered with the facts that, first, the $n(z)$ agreement between the Cold Spot and G23 is inexact given that the $n(z)$ peak at $z = 0.21$ is more pronounced in G23. This difference is reflected in the predicted ISW decrements, $-9.1 \pm 6.1\mu K$ and $-3.6 \pm 7.5\mu K$ for the Cold Spot and G23 respectively. Second, the $n(z)$'s used

to construct the $\delta_m(z)$'s were normalised with respect to their surroundings and so don't contain all the information of the largest scale fluctuations. As discussed previously the region surrounding the Cold Spot is under-dense with respect to the region surrounding G23 by $\sim 8\%$ so the two fields are not exactly equivalent and the structures detected in this analysis are embedded in different large scale potentials. This could have an effect on the Cold Spot ISW imprint but likely at larger scales than the 5° radius feature we have mainly investigated here. One could argue that the alignment of the CMB Cold Spot and the large $z = 0.14$ void implies a causal link though the improbability of alignment but voids of this scale are not expected to be unique (Nadathur et al., 2014, Kovács et al., 2017) and our search was not blind nor the only attempt to detect for a void.

If not explained by a ΛCDM ISW effect the Cold Spot could have more exotic primordial origins. If it is a non-Gaussian feature, then explanations would then include either the presence in the early universe of topological defects such as textures (Cruz et al., 2007) or inhomogeneous re-heating associated with non-standard inflation (Bueno Sánchez, 2014). Another explanation could be that the Cold Spot is the remnant of a collision between our Universe and another 'bubble' universe during an early inflationary phase (Chang et al., 2009, Larjo & Levi, 2010). It must be borne in mind that even without a supervoid the Cold Spot may still be caused by an unlikely statistical fluctuation in the standard (Gaussian) ΛCDM cosmology.

To conclude, based on the arguments and caveats above we have ruled out the existence of a void at which could imprint the majority of the CMB Cold Spot via a ΛCDM ISW effect. The predicted decrement is consistent with some previous studies (Szapudi et al., 2015), although certainly at the low end of literature values. We have additionally placed powerful constraints on any non-standard ISW-like effect which must now show how voids, apparently unremarkable on 5° scales, can imprint the unique CMB Cold Spot. The presence of the detected voids only slightly relaxes the significance of the primordial residual of the CMB Cold Spot in standard cosmology to approximately 1 in 50, tilting the balance towards a primordial and also possibly

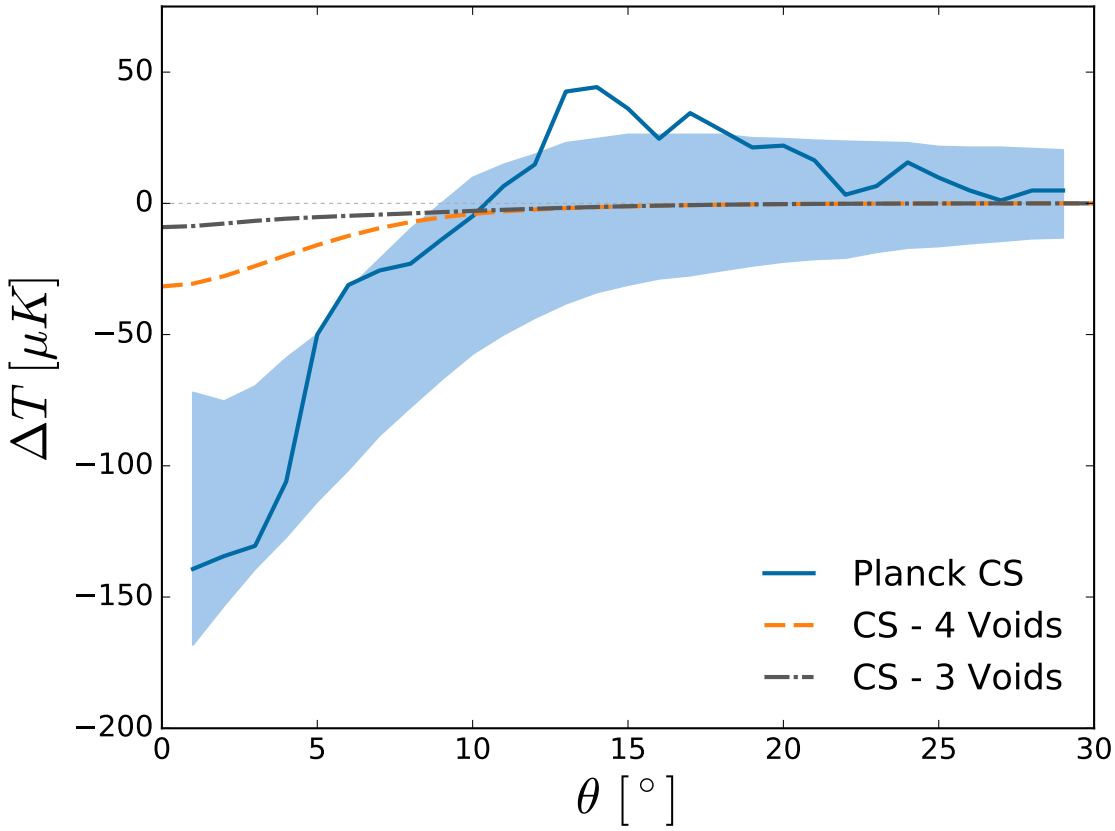


Figure 2.8: The Cold Spot temperature profile (Planck Collaboration et al., 2016c) (blue line) and the ISW imprints of the 3- and 4-void models (grey dot-dashed and yellow dashed respectively) fitted to the Cold Spot region. The void temperature profiles from Table 2.3 have been summed and the result fitted to eq. (3) of Naidoo et al. (2016). The shaded region (light blue) is the 68% confidence interval from the coldest spots identified in Gaussian simulations (see Nadathur et al., 2014, Fig 6).

non-Gaussian origin. But at this level of significance clearly any exotic explanation will have to look for further evidence beyond the Cold Spot temperature profile.

## 2.5 Conclusions

We have conducted a spectroscopic redshift survey of the CMB Cold Spot core in order to test claims from photo-$z$ analyses for the existence of a large low-$z$ void that could be of sufficient scale and rarity to explain the CMB Cold Spot.

• We have detected an 119 $h^{-1}$Mpc, $\delta_m = -0.34$ under-density at $z = 0.14$. This under-density is much less extended than found in photo-$z$ analyses in the literature

but is more under-dense. The estimated ΛCDM ISW effect from this void is estimated at $-6.25\mu$K, much too small to explain the CMB Cold Spot.

• Two further small under-densities were observed at $z = 0.26$ and 0.30. The effect of these voids is even smaller than the $z = 0.14$ void.

• A further candidate void was detected at $z = 0.42$ although we conclude this is most likely due to redshift incompleteness in the survey. Even if real this void would still not explain the CMB Cold Spot.

• Without detailed calculation we have shown that the rarity of this void is not sufficient to motivate it as the cause of the CMB Cold Spot because of the similarity with GAMA G23. The comparability of under-densities at $z \sim 0.4$ between G23 and the Cold Spot again means that even if the $z = 0.42$ void in the Cold Spot was not a systematic effect, it is not unique enough to suggest an effect beyond standard cosmology.

• Combining our data with previous work (Bremer et al., 2010) the presence of a very large void which can explain the CMB Cold Spot can be excluded up to $z \sim 1$, beyond which the ISW effect becomes significantly reduced as the effect of the Cosmological Constant is diluted.

• The similarity between the 2CSz and G23 $n(z)$ distributions may have some explanation in the similar $n(z)$ seen in the 2dFGRS SGC strip that spans the $\sim 60°$ angle between these sightlines. This includes the 'Local Hole' at $z < 0.06$ but may also include further structures out to $z \sim 0.2$.

Our 2CSz results therefore argue against a supervoid explaining a significant fraction of the Cold Spot via the ISW effect. This suggests a primordial origin for the Cold Spot, either from an unlikely fluctuation in the standard cosmology or as a feature produced by non-Gaussian conditions in the early Universe.

# Chapter 3

# Witnessing galaxy assembly in an extended $z \approx 3$ structure

We present new observations acquired with the Multi Unit Spectroscopic Explorer instrument on the Very Large Telescope in a quasar field that hosts a high column-density damped Ly$\alpha$ absorber (DLA) at $z \approx 3.25$. We detect Ly$\alpha$ emission from a nebula at the redshift of the DLA with line luminosity $(27\pm1)\times10^{41}$ erg s$^{-1}$, which extends over 37$\pm$1 kpc above a surface brightness limit of $6\times10^{-19}$ erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$ at a projected distance of $30.5 \pm 0.5$ kpc from the quasar sightline. Two clumps lie inside this nebula, both with Ly$\alpha$ rest-frame equivalent width $> 50$ Å and with relative line-of-sight velocities aligned with two main absorption components seen in the DLA spectrum. In addition, we identify a compact galaxy at a projected distance of $19.1 \pm 0.5$ kpc from the quasar sightline. The galaxy spectrum is noisy but consistent with that of a star-forming galaxy at the DLA redshift. We argue that the Ly$\alpha$ nebula is ionized by radiation from star formation inside the two clumps, or by radiation from the compact galaxy. In either case, these data imply the presence of a structure with size $\gg 50$ kpc inside which galaxies are assembling, a picture consistent with galaxy formation in groups and filaments as predicted by cosmological simulations such as the EAGLE simulations.

## 3.1 Introduction

In the current cosmological paradigm, galaxies form in overdense regions of the Universe when gas is funnelled from the intergalactic medium inside dark matter halos, where it settles into disks and cools to form stars (White & Rees, 1978; Blumenthal et al., 1984). A distinctive feature of this model is the clustering of galaxies on scales ranging from individual halos of tens of kiloparsecs, up to structures on scales of few megaparsecs. This picture is supported by hydrodynamic simulations, as well as by the statistics recovered from observations of the starlight emitted by young galaxies that populated the Universe more than ten billion years ago.

Direct observations of the link between hydrogen and star formation within and around galaxies provide key insight into the astrophysical processes that regulate galaxy evolution in this hierarchy of groups and filaments (e.g. Warren & Møller, 1996; Weatherley et al., 2005; Cucciati et al., 2014; Fumagalli et al., 2016; Bielby et al., 2017; Péroux et al., 2017; Cai et al., 2017b). However, due to the difficulty in mapping neutral hydrogen in emission at high redshift, until recently it has proven challenging to study in detail the connection between gas and galaxies inside overdensities beyond $z \gtrsim 2$.

Most of our knowledge of the neutral gas content of the high-redshift Universe originates from the study of damped Ly$\alpha$ absorbers (DLAs), which are detected in absorption against bright background sources (Wolfe et al., 2005). Defined as systems with H I column density $N_{\rm HI} \geq 2 \times 10^{20}$ cm$^{-2}$, DLAs trace the bulk of the neutral hydrogen in the distant Universe and are the major repository of the fuel for the formation of the stars seen at present days (e.g. Prochaska & Wolfe, 2009; Noterdaeme et al., 2009). Many pieces of evidence link DLAs to star formation in high-redshift galaxies. For instance, the high column density of DLAs provides the necessary conditions for hydrogen to shield from the ambient ionizing radiation, potentially allowing the condensation of a cold and molecular phase that is thought to be essential for the formation of new stars (Krumholz et al., 2009; Neeleman et al.,

2015; Noterdaeme et al., 2015; Rafelski et al., 2016). Furthermore, DLAs contain significant amounts of heavy elements with a mean metallicity of $\log Z/Z_{\odot} \approx -1.5$ at $z \gtrsim 2$, which is more than an order of magnitude above what is found in the intergalactic medium (e.g. Prochaska et al., 2003; Rafelski et al., 2012, 2014).

However, establishing a connection between DLAs and star-forming galaxies has been a non-trivial task. Despite decades of searches, only tens of DLAs have been associated directly to counterparts in emission, particularly thanks to recent campaigns (e.g. Møller et al., 2004; Fynbo et al., 2010; Noterdaeme et al., 2012; Péroux et al., 2012; Fynbo et al., 2013; Neeleman et al., 2017; Krogager et al., 2017). This low detection rate across the entire DLA population is generally attributed to the difficulties of imaging faint emission from stellar populations or hydrogen recombination lines at high redshifts. It also likely reflects the fact that only a fraction of the DLA population is directly connected to active *in situ* star formation, or that only a subset of DLAs reside near galaxies which are sufficiently bright for direct detection (such as high column-density and/or metal-rich DLAs; Fynbo et al., 2008; Møller et al., 2013; Christensen et al., 2014; Fumagalli et al., 2015; Joshi et al., 2017; Krogager et al., 2017).

Together with recent advancements at infrared and millimetre wavelengths (e.g. Péroux et al., 2012; Neeleman et al., 2017), the deployment of the Multi Unit Spectroscopic Explorer (MUSE; Bacon et al., 2010) at the Very Large Telescope (VLT) represents a breakthrough to address the long-standing issue of connecting DLAs to galaxy counterparts. Indeed, MUSE allows us to image Ly$\alpha$ emission within $\approx$250 kpc from the position of known DLAs to flux limits of $\approx 10^{-18}$ erg s$^{-1}$ cm$^{-2}$ in exposure times of only few hours, thus enabling deep searches of DLA host galaxies without the need of pre-selecting targets for spectroscopic follow-up observations.

In this work, we present new MUSE observations of a field centred on the quasar J025518.58+004847.6 at $z_{\rm qso} \approx 3.996$, where high-resolution absorption spectroscopy reveals the presence of a high-column density DLA at the redshift $z_{\rm dla} = 3.2552 \pm 0.0001$ (Prochaska et al., 2001; Fumagalli et al., 2014). This target has been selected

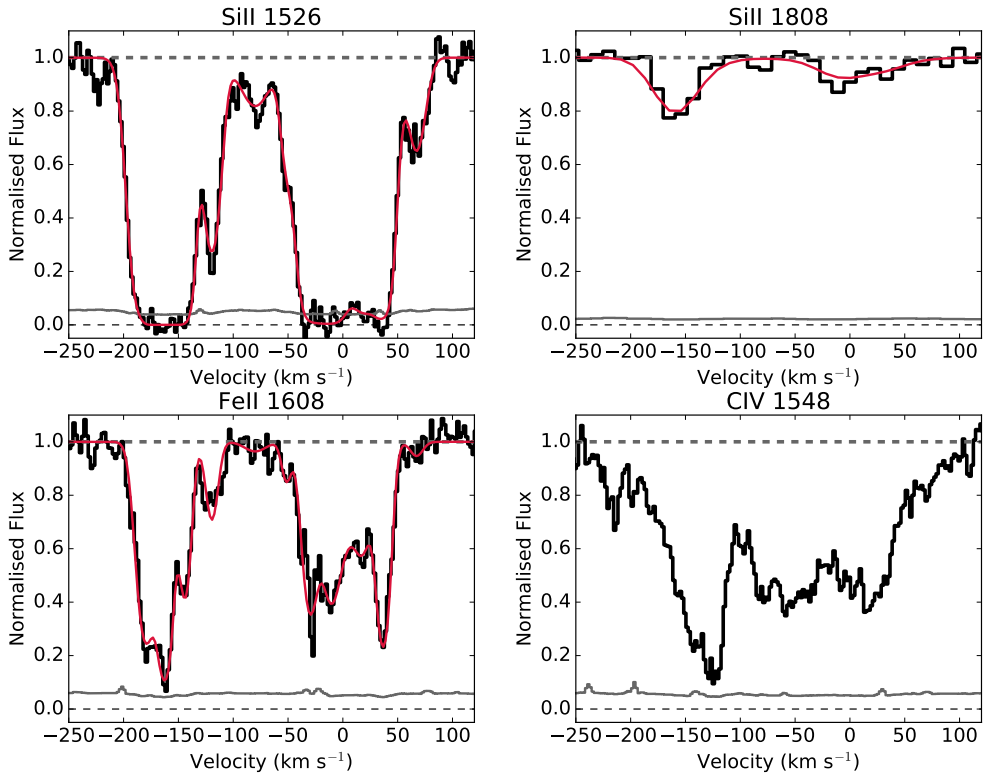


Figure 3.1: Gallery of selected metal transitions associated with the $z_{\rm dla} = 3.2552 \pm 0.0001$ DLA. Data from HIRES or X-SHOOTER (in the case of Si II $\lambda$1808) are shown in black, together with the associated $1\sigma$ error on the flux (grey solid line). The red solid line shows a multi-component Voigt profile fit to the low ionization lines. The grey dashed line at the top of each panel marks the continuum level. Two main components separated by $155 \pm 6$ km s$^{-1}$ are clearly visible, each with sub-components.

from a larger parent sample of quasar fields with available deep optical imaging (see Sect. 3.2). The layout of this paper is the following. Details on the data acquisition and reduction are presented in Sect. 3.2, while the analysis and modelling of these observations are in Sect. 3.3, Sect. 3.4, and Sect. 3.5. Summary and conclusions are presented in Sect. 3.6. Throughout this work, we use the Planck 2015 cosmological parameters ($\Omega_{\rm m} = 0.307$, $H_0 = 67.7$ km s$^{-1}$ Mpc$^{-1}$; Planck Collaboration et al., 2016b) and we quote magnitudes in the AB system.

## 3.2 Data acquisition and reduction

The quasar field J025518.58+004847.6 is one of the six fields selected from the parent sample of quasars included in our imaging survey designed to probe the *in situ* star formation rate (SFR) of high-redshift DLAs (O'Meara et al., 2006; Fumagalli et al., 2010, 2014, 2015). The six fields have been selected to host DLAs at $z > 3$, the redshift for which Ly$\alpha$ enters the wavelength range covered by MUSE, with no other constraints on the physical properties of the absorbing gas (e.g. on metallicity). MUSE observations for this sub-sample have, at the time of writing, been recently completed at VLT as part of the ESO programmes 095.A-0051 and 096.A-0022 (PI Fumagalli). The line of sight presented in this work is the first field we have analysed, and the entire sample will be presented in a forthcoming publication.

MUSE observations of the quasar field J025518.58 + 004847.6 have been conducted during the nights 17-20 September 2015 in a series of 1500 s exposures totalling 2.5 hours on source, under good seeing (requested to be $\leq 0.8''$) and clear sky, but with non-photometric conditions. Data have been reduced using the standard ESO MUSE pipeline (v1.6.2; Weilbacher et al., 2014) that includes basic data processing (bias subtraction and flat-fielding), plus wavelength and photometry calibrations. Throughout this analysis, barycentric corrections have been applied to the MUSE data and the wavelengths have been converted into vacuum for consistent comparisons with the spectroscopic data described below. We further post-process the resulting datacubes to enhance the quality of sky subtraction and flat-fielding using the CubExtractor code (CubEx, Cantalupo in prep.), following our earlier work (Borisova et al., 2016; Fumagalli et al., 2017a).

The final data product is a cube that samples the instrument field of view of $1 \times 1$ arcmin$^2$ in pixels of size 0.2 arcsec and, for each pixel, contains a spectrum covering the wavelength range $4750 - 9350$ Å in bins of 1.25 Å. Following comparisons with photometric data from the Sloan Digital Sky Survey (Eisenstein et al., 2011), we apply a small correction factor of 1.12 to the flux calibration to account for low

levels of atmospheric extinction during the observations. We further correct fluxes to remove Milky Way dust extinction in the direction of our observations following Schlafly & Finkbeiner (2011). At $\lambda \approx 5170$ Å, which corresponds to the Ly$\alpha$ wavelength at the redshift of the DLA, we achieve an effective image quality of $\approx 0.6$ arcsec, and a noise level of $\approx 6 \times 10^{-19}$ erg s$^{-1}$ cm$^{-2}$ Å$^{-1}$ arcsec$^{-2}$ (root mean square).

Throughout this analysis, we further make use of broad-band imaging data collected with the *Hubble Space Telescope* (HST) in the F390W filter using Wide Field Camera 3, and in the $u'$, $V$, and $R$ filters using the LRIS camera at the Keck Telescope. The observations and reduction techniques for these data are described in Fumagalli et al. (2014). Finally, we use archival spectroscopic data on the DLA from the HIRES instrument at the Keck Telescope and the X-SHOOTER instrument at VLT. These data are described in Prochaska et al. (2001) and López et al. (2016). The HIRES spectrum covers the wavelength range $5800 - 8155$ Å with a signal-to-noise ratio ($S/N$) of $\approx 15$ per pixel and a spectral resolution of $\approx 6$ km s$^{-1}$. The X-SHOOTER spectrum covers instead the wavelength range $3100 - 18000$ Å with a $S/N \approx 30$ per pixel at a resolution of $\approx 40 - 60$ km s$^{-1}$, depending on the instrument arm.

## 3.3 Analysis of the imaging and spectroscopic data

### 3.3.1 The physical properties of the $z \approx 3.25$ DLA

A gallery of selected absorption lines at the redshift of the $z \approx 3.25$ DLA, which are chosen to be representative of low ionization transitions, is shown in Figure 3.1. The profiles of low ionization metal lines reveal the presence of two distinct main components, each showing additional sub-components. Velocity structure is also seen in C IV, although this ion does not track the low ionization transitions in velocity space, suggesting the presence of multi-phase gas in this DLA (e.g. Fox et al.,

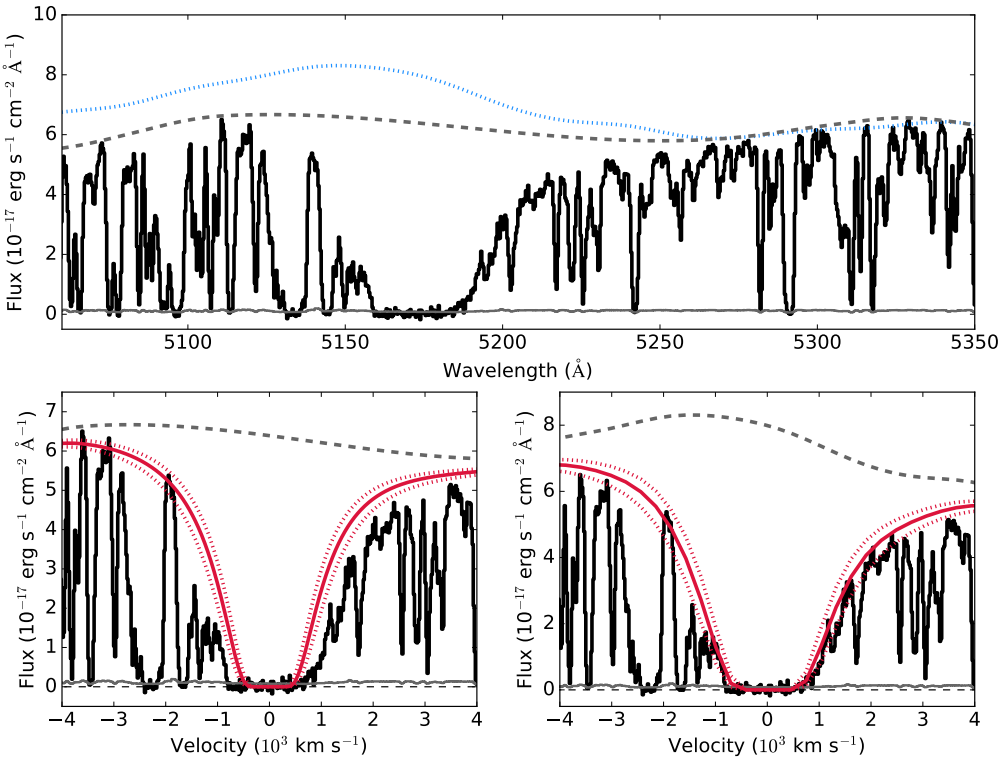


Figure 3.2: (*Top panel*) X-SHOOTER spectrum of the J025518.58 + 004847.6 quasar (black), together with the associated error on the flux (grey solid line). The H I profile of the $z_{\rm dla} = 3.2552 \pm 0.0001$ DLA is visible at $\lambda \approx 5173$ Å. The grey dashed line marks the local continuum, computed with a spline function that interpolates regions deemed free from absorption in the Ly$\alpha$ forest. The blue dotted line shows instead the assumed continuum from the quasar template by Telfer et al. (2002), which is normalised to reproduce regions of the spectrum with no absorption. This continuum level deviates from the local one around $\approx 5150$ Å because of the presence of Ly$\beta$ and O VI emission lines of the quasar. *(Bottom panels)* Closer view of the H I profile of the DLA, together with a Voigt profile model (red solid line) and associated errors (red dotted lines). The two panels differ in the choice of the adopted continuum level (grey dashed lines), which is either based on a local determination (left panel) or on a quasar template (right panel) as described above.

2007). By fitting Voigt profiles to the Si II $\lambda$1526 and the Fe II $\lambda$1608 lines using the VPFIT code (Carswell & Webb, 2014), we find a column-density weighted redshift of $z_1 = 3.2530 \pm 0.0001$ and $z_2 = 3.2552 \pm 0.0001$ for the two main components.

Further, a Voigt profile centred at $z_2$ provides a good description of the Ly$\alpha$ absorption line and, in the following, we assume $z_2$ as the fiducial redshift for the DLA. As shown in Figure 3.2, the measurement of the H I column density is complicated by the fact that the Ly$\alpha$ transition at the DLA redshift lies next to the O VI and Ly$\beta$ emission lines of the quasar, a coincidence that makes our estimate of the continuum level rather uncertain. For this reason, we derive at first a lower limit on the DLA column density by assuming a "local" continuum, which we obtain with a spline function constrained to follow regions deemed free from absorption inside the Ly$\alpha$ forest (grey dashed line in the top panel of Figure 3.2). The resulting model for the DLA (bottom-left panel of Figure 3.2) yields a column density of $\log(N_{\rm HI}/{\rm cm}^{-2}) > 20.55 \pm 0.10$, which is primarily constrained by the blue wing.

However, it is evident that additional absorption is present in the red side of the H I transition. Therefore, we construct a second model adopting the continuum level from the quasar template by Telfer et al. (2002), which we normalise to reproduce the observed spectrum in regions not absorbed by the Ly$\alpha$ forest (blue dotted line in the top panel of Figure 3.2). With this continuum, we derive a second model for the DLA with $\log(N_{\rm HI}/{\rm cm}^{-2}) = 20.85 \pm 0.10$ (bottom-right of Figure 3.2), which is primarily constrained by the red wing of the absorption profile. As the red wing is less affected by the quasar emission lines, our model is not very sensitive to the exact shape of the template. For instance, the template by Vanden Berk et al. (2001) yields a consistent model within errors. Qualitatively, this model appears to reproduce well the shape of the Ly$\alpha$ transition and hereafter, we assume $\log(N_{\rm HI}/{\rm cm}^{-2}) = 20.85 \pm 0.10$ as our best estimate for the H I column density. The quoted error does not reflect the systematic uncertainty on the continuum placement.

Before continuing with our analysis, we note that the low velocity separation of the two main metal components allows for a further decomposition of the H I profile, with

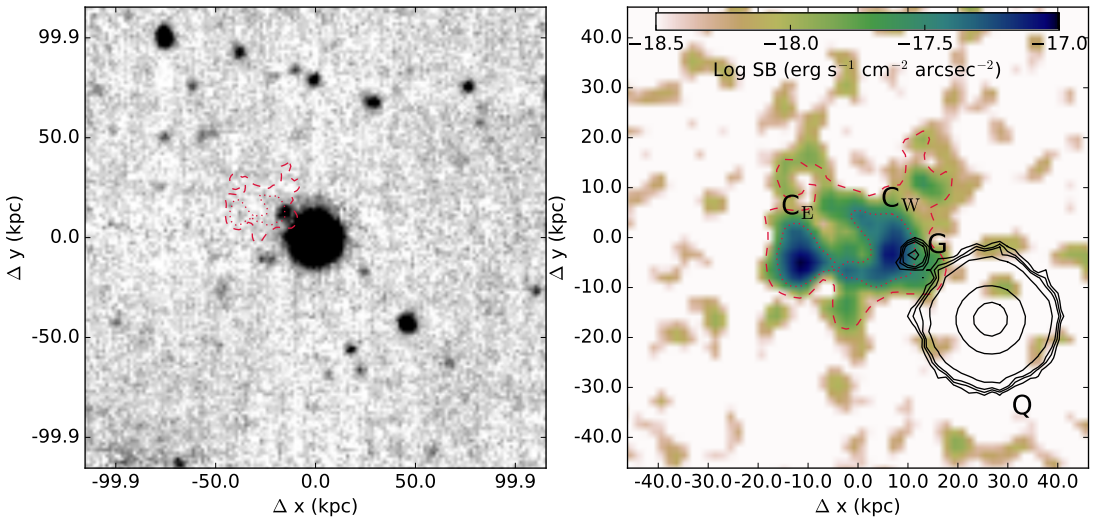


Figure 3.3: (*Left panel*) The white-light image reconstructed from the MUSE datacube showing the quasar at the centre, where absorption spectroscopy reveals the presence of a DLA with column density $\log N_{\rm HI} = 20.85 \pm 0.10$ cm$^{-2}$ at $z_{\rm dla} = 3.2552 \pm 0.0001$. The red contours mark the $1\sigma$ and $5\sigma$ surface brightness levels of the source detected in Ly$\alpha$ at the DLA redshift. North is upward and East to the left. (*Right panel*) The pseudo narrow-band image of the Ly$\alpha$ emission at the DLA redshift, smoothed by 2 pixels. The red contours identify the $1\sigma$ and $5\sigma$ surface brightness levels, respectively. MUSE observations reveal the presence of a $37 \pm 1$ kpc structure, with two distinct clumps, labelled "$C_{\rm E}$" and "$C_{\rm W}$". The black contours mark the position of the quasar (labelled "Q") and the continuum-detected galaxy assumed at the DLA redshift (labelled "G").

components at $z_1$ and $z_2$. Both in the cases of the local continuum determination and the continuum based on a quasar template, solutions with two components yield a total $N_{\rm H\,I}$ in agreement with the single-component analysis. However, the lack of high-order hydrogen lines, which fall beyond the Lyman limit of a higher redshift absorber, makes an unambiguous decomposition difficult. Further, sub-structures are evident within the metal lines beyond the two main components, suggesting that the physical interpretation of a two-cloud model is still limited. For this reason, we proceed with a single-component fit for H I, which is nevertheless effective in describing the total amount of neutral hydrogen within this system.

Strong metal lines in the DLA spectrum reveal that the absorbing gas is significantly

enriched, with the majority of the low-ionization transitions being saturated, with the exception of the weak Si II $\lambda$1808 transition. After correcting the spectrum for partial blending arising from the telluric band between $\approx 7600 - 7700$ Å, we apply the kinematic model derived from the Si II $\lambda$1526 line to the Si II $\lambda$1808 transition, finding a column density of $\log(N_{\rm SiII}/{\rm cm}^{-2}) = 15.29 \pm 0.05$ which is consistent with earlier measurements based on the apparent optical depth method (Prochaska et al., 2001; Fumagalli et al., 2014). Combined with the measurement of the H I column density and assuming the relative abundance pattern of the Sun from Asplund et al. (2009), we derive a metallicity of $\log Z/Z_\odot < -0.8 \pm 0.1$ if $\log(N_{\rm HI}/{\rm cm}^{-2}) > 20.55 \pm 0.10$, or $\log Z/Z_\odot = -1.1 \pm 0.1$ assuming the fiducial value of column density ($\log(N_{\rm HI}/{\rm cm}^{-2}) = 20.85 \pm 0.10$). No ionization corrections are included, as they are known to be negligible for DLAs. Similarly, no corrections for dust depletion have been applied, although we note they may be non-negligible given the metallicity of this system (up to $\approx 0.5$ dex, but generally $\lesssim 0.2$ dex for Si; Rafelski et al., 2012; De Cia et al., 2016).

From our analysis, we conclude that this system is enriched above the median value of $\log Z/Z_\odot \approx -1.5$ for the DLA population at $z \approx 2 - 3$ (Rafelski et al., 2012). Moreover, the line profiles exhibit a complex kinematic structure with two main velocity components stretching over $> 200$ km s$^{-1}$, placing this DLA in the upper end of the metallicity and kinematics distribution of the DLA population at these redshifts (Ledoux et al., 2006; Neeleman et al., 2013).

### 3.3.2 Search for galaxy counterparts

The search for Ly$\alpha$ emitting galaxies inside the MUSE cube is performed using the three-dimensional (3D) source extraction code CubEx (Cantalupo in prep.), following the techniques we developed for the analysis of MUSE observations in quasar fields at $z \approx 3 - 4$ (e.g. Fumagalli et al., 2016).

Briefly, after convolving the cube with a boxcar filter of 2 pixels on a side in the

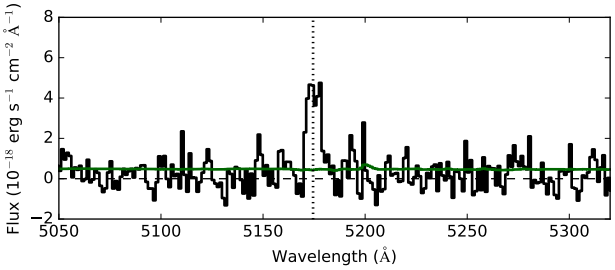


Figure 3.4: Spectrum of the Ly$\alpha$ nebula, extracted from the MUSE cube in a mask that encompasses the full extent of the emission. The corresponding error array is shown in green. The vertical dotted line marks the mean redshift of the nebula.

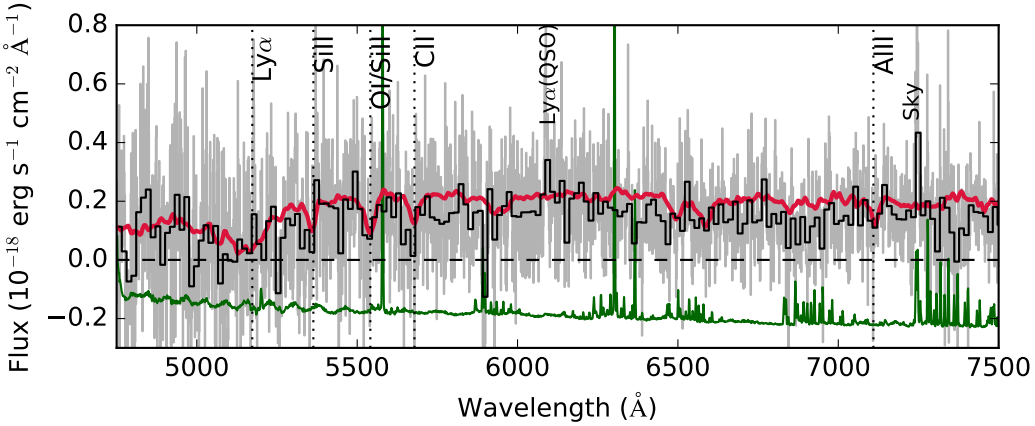


Figure 3.5: Spectrum of galaxy G from MUSE data, before (grey) and after (black) rebinning by 12 spectral pixels (corresponding to 15Å). The error array (green) has been offset for visualisation purposes. The red line shows the composite LBG spectrum from Bielby et al. (2013). Ly$\alpha$ and metal lines (as labelled) are tentatively detected in absorption. These spectral features, together with colour information and the source position, place this galaxy at a likely redshift $z_{\rm G} = z_{\rm dla}$.

spatial direction, we search for candidate sources which are detected with $S/N > 3$ within a minimum volume of 30 voxels. After producing 3D segmentation maps of all the detected objects, we create pseudo narrow-band images by projecting the cube along the wavelength direction as described in Borisova et al. (2016). Next, we visually inspect these images and the associated 1D spectra to exclude both spurious sources (cosmic rays and noise artefacts at the edge of the field of view) and lower-redshift line emitters, which we identify by virtue of resolved [OII] doublets or multiple emission lines.

In the end, our search for Ly$\alpha$ emitting sources within $\pm$500 km s$^{-1}$ around the DLA redshift reveals the presence of an extended structure ($1\sigma$ level of $6{\times}10^{-19}$ erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$) near the quasar, with two distinct clumps (hereafter the East "$C_{\rm E}$" and West "$C_{\rm W}$" clumps), as shown in Figure 3.3. A 1D spectrum extracted from a mask that encompasses the extent of the nebula (see Sect. 3.3.3) is in Figure 3.4. The Ly$\alpha$ emission line is detected at high significance.

No other Ly$\alpha$ sources are detected to a flux limit of $> 2 \times 10^{-18}$ erg s$^{-1}$ cm$^{-2}$ ($3\sigma$ confidence level). Based on the completeness tests we performed in Fumagalli et al. (2016), our search is assumed to be $\gtrsim$ 80% complete to a flux of $\gtrsim 4.5 \times 10^{-18}$ erg s$^{-1}$ cm$^{-2}$, after accounting for the shorter exposure time of these observations.

The detected source is clearly associated with the DLA, being separated by only $90 \pm 20$ km s$^{-1}$ in velocity space and having a centroid offset of $30.5 \pm 0.5$ kpc in projection with respect to the quasar sightline. The velocity offset increases to $\approx$ 250 km s$^{-1}$ if we assume $z_1$ instead of $z_2$ as the redshift of the DLA. In any case, we note that redshifts of $\approx$ 300 km s$^{-1}$ relative to systemic velocity are quite common for the resonant Ly$\alpha$ transition in Lyman break galaxies (LBGs; e.g. Steidel et al., 2010; Bielby et al., 2011; Rakic et al., 2011), and are also seen in DLAs with detected Ly$\alpha$ emission (e.g. Srianand et al., 2016; Joshi et al., 2017, and reference therein).

Finally, we measure spectroscopic redshifts for the continuum-detected sources, fol-

lowing the procedures detailed in Fumagalli et al. (2016), which are here only briefly summarised. After generating a deep white-light image by collapsing the cube along the wavelength axis, we run the SExtractor code (Bertin & Arnouts, 1996) with a detection area of 8 pixels and a threshold of $1.5\sigma_{\rm rms}$, where $\sigma_{\rm rms}$ is the background root mean square. For each detected source, we then extract the 1D spectrum, which is visually inspected by two of the authors (MF and RM) to assign a redshift. Of all the galaxies to which we could assign a spectroscopic redshift following this procedure (typically with $m_{\rm R} \lesssim 25$ mag), none is found to lie within $\pm 500$ km s$^{-1}$ of the DLA.

Due to its close separation to the quasar in projection, source G (Figure 3.3) requires an additional local background subtraction to correct for the contamination from the quasar light in the wings of the point spread function. To account for this, we extract an annulus centred on the quasar that encompasses the full width of galaxy G and contains $\approx 200$ pixels, without including the emission from the galaxy itself. We next generate a composite spectrum of the pixels contained in this annulus, which we subtract from the source spectrum.

As shown in Figure 3.5, the residual spectrum is very noisy. However, this source has a continuum emission consistent with a galaxy at $z \approx 3.2$ when compared to an LBG template (Bielby et al., 2013). The high-redshift nature of this galaxy is also confirmed by the lack of detection in the LRIS $u'$ filter ($m_{\rm u'} < 27.8$ mag) and the marginal detection in the HST F390W filter ($m_{\rm f390w} = 27.98 \pm 0.23$ mag). Combined with the $R$ and $V$ magnitudes measured in the Keck imaging ($m_{\rm V} = 26.19 \pm 0.24$ mag; $m_{\rm R} = 25.33 \pm 0.18$ mag), we derive colours of $U - V = 1.8 \pm 0.3$ mag and $V - R = 0.9 \pm 0.3$ mag, which are consistent with those of $z \approx 3$ LBGs (Steidel et al., 2003; Bielby et al., 2011). Moreover, after rebinning the spectrum, H I in absorption and perhaps metal lines at the DLA redshift are tentatively detected. We observe no Ly$\alpha$ emission from this galaxy, but this is not unusual even for star-forming LBGs (Steidel et al., 2011). The best-fit redshift obtained comparing the spectro-photometry of galaxy G with the spectral energy distribution of an LBG is $z_{\rm G} = 3.22 \pm 0.03$.

Given the presence of neutral hydrogen at the DLA position and that neutral hydrogen is likely present also at the location of the Ly$\alpha$ emission, it is possible that the features seen in the spectrum of galaxy G originate from foreground absorption against the UV continuum of a source at $z > z_{\rm dla}$ (see Cooke & O'Meara, 2015). For this reason, and given the marginal quality of the data, we cannot determine a secure redshift for galaxy G. However, the colour information combined with the source position near the nebula makes galaxy G a compelling candidate for a system within the same structure that gives rise to the DLA and the extended Ly$\alpha$ emission. Thus, in the following, we consider this galaxy at the redshift $z_{\rm G} = z_{\rm dla}$. Future follow-up observations targeting rest-frame optical lines in the IR will test our working hypothesis.

### 3.3.3 Properties of the detected galaxies

With a flux of $f_r = (16 \pm 1) \times 10^{-20}$ erg s$^{-1}$ cm$^{-2}$ Å$^{-1}$ measured in an image generated from the MUSE cube convolved with the SDSS $r$-band filter, galaxy G is forming stars at a rate of $3.7 \pm 0.2$ M$_\odot$ yr$^{-1}$, when adopting the calibration described in Fumagalli et al. (2010) and neglecting internal dust extinction. The impact parameter between the quasar sightline and galaxy G is $19.1 \pm 0.5$ kpc, sufficiently large to make it quite unlikely that we are probing directly the galaxy interstellar medium in absorption.

As our data cover only the rest-frame UV ($\lesssim$ 2000 Å), we do not have a direct estimate of the galaxy stellar mass. Using empirical correlations between the SFRs and stellar masses of LBGs between $z \approx 2-4$, galaxy G is expected to have a stellar mass of $\approx 10^9$ M$_\odot$ (Whitaker et al., 2014; Salmon et al., 2015). Conversely, the stellar mass expected from scaling relations calibrated on known DLAs is a factor $\approx 5$ higher, as it is predicted to be $\approx 5 \times 10^9$ M$_\odot$ for a DLA with metallicity of $\log Z/Z_\odot \approx -1.1$ at the observed impact parameter (Møller et al., 2013; Christensen et al., 2014). However, the large scatter in both relations combined with the unknown

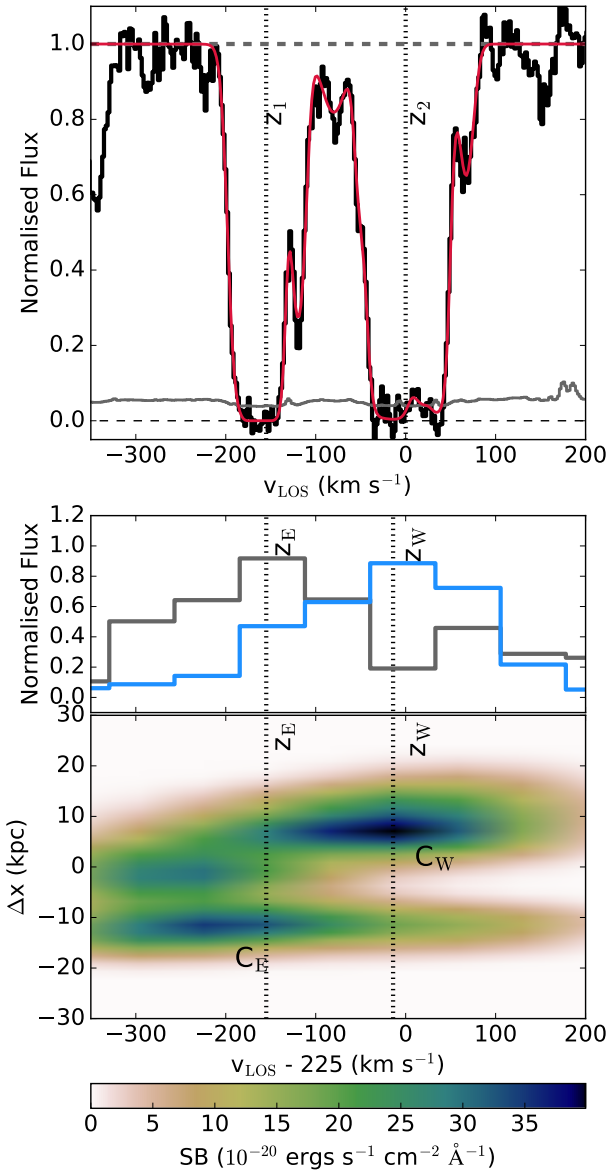


Figure 3.6: *(Top panel)* The Si II line profile from HIRES data as in Figure 3.1, with the redshifts of the two main components marked by vertical dotted lines. *(Bottom panel)* Ly$\alpha$ narrow-band velocity/position map reconstructed from the MUSE datacube, with the inset showing the projected 1D spectrum. The grey and blue lines in the inset refer to the $C_{\rm E}$ and $C_{\rm W}$ clumps, respectively. The redshifts of the two clumps are also marked with vertical dotted lines. Once an offset of $\Delta v = -225$ km s$^{-1}$ is applied to align the East clump with the first absorption component, the absorption and emission spectra appear to track each other in velocity space.

dust extinction for galaxy G makes reliable comparisons between these methods difficult without additional observations that target the rest-frame optical and IR wavelengths. It is also interesting to note that this detection occurs in a metal-rich and high column-density DLA, in agreement with the trend that has emerged from searches of DLA counterparts completed in the past decades (Krogager et al., 2017).

Focusing on the extended Ly$\alpha$ emission next, we use the surface brightness map extracted from the MUSE cube to define a region that encloses the full source (with a contour at the $1\sigma$ level of $6 \times 10^{-19}$ erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$), and the two clumps (with a contour at the $5\sigma$ level). These apertures are shown in Figure 3.3 and are used throughout this analysis to compute Ly$\alpha$ line fluxes and for aperture-matched broadband photometry. Under these assumptions, the two knots $C_{\rm E}$ and $C_{\rm W}$ are separated by $16.5 \pm 0.5$ kpc, and they contain respectively $(28 \pm 2)\%$ and $(35 \pm 3)\%$ of the total line flux of $(2.8 \pm 0.1) \times 10^{-17}$ erg s$^{-1}$ cm$^{-2}$. As shown in Figure 3.3, these clumps are embedded in an extended component that stretches for $37 \pm 1$ kpc on a side. To date, this is one of the largest Ly$\alpha$ emitting structures associated with a DLA (see e.g. Kashikawa et al., 2014).

When examining the line-of-sight velocity derived from the MUSE cube, differential velocity is seen across the structure, with the two clumps being separated by $\Delta v_{\rm E-W} = 140 \pm 20$ km s$^{-1}$ in velocity space (Figure 3.6). This velocity separation is consistent, within errors, with the velocity separation inferred from the two main absorption components ($\Delta v_{1-2} = 155 \pm 6$ km s$^{-1}$), suggesting that the Ly$\alpha$ emission from the clumps follows the absorption profile of the DLA in velocity space. We note, however, that radiative transfer effects modulate the Ly$\alpha$ line profile, potentially altering the relative velocity separation that is intrinsic for the two clumps.

The kinematic alignment is further highlighted in Figure 3.6, where we compare the emission and absorption spectra. After shifting the Ly$\alpha$ emission by $\Delta v = -225$ km s$^{-1}$ to align the first main absorption component with the redshift of $C_{\rm E}$, the DLA profile and the Ly$\alpha$ line emission show remarkable similarity, reinforcing the hypothesis of a physical association between the two clumps and the DLA. Physically,

$\Delta v$ reflects a combination of the offset between emission and absorption arising from radiative transfer effects and intrinsic velocity difference due to peculiar motions along the line of sight. With current observations, we cannot separate these two effects, but future IR observations that target rest-frame optical lines will be able to further constrain the kinematics of this system.

Focusing on the extended Ly$\alpha$ emission, we note that it is largely non-overlapping with galaxy G. When masking this compact source, our analysis of deep imaging data from Keck, VLT, and HST reveals that no continuum emission is detected from the rest-frame wavelength interval $821 - 2198$ Å, to a limit of $< 1.7 \times 10^{-19}$ erg s$^{-1}$ cm$^{-2}$ Å$^{-1}$ ($2\sigma$) that is calculated on the deepest imaging obtained by collapsing the MUSE cube along the wavelength axis. Hence, without considering galaxy G, the lack of extended UV continuum across this structure sets a limit for the average SFR of $< 5.1$ M$_\odot$ yr$^{-1}$ for the entire system, and $< 0.9$ M$_\odot$ yr$^{-1}$ and $< 1.3$ M$_\odot$ yr$^{-1}$ for the E and W clumps. These estimates do not account for possible dust-obscured components. The corresponding Ly$\alpha$ rest-frame equivalent width for the two clumps (E and W) are $> 60.7$ Å and $> 52.3$ Å. Here, non-detections are computed at $2\sigma$ confidence level accounting for correlated noise by measuring the standard deviation of fluxes recovered in empty regions of the images within apertures of equal size to those defined above. With its UV flux, however, galaxy G contributes to the total equivalent width of the structure, which becomes $41 \pm 3$ Å in the rest frame.

## 3.4 Nature of the extended emission

Our MUSE observations reveal an emitting structure of $\approx 40$ kpc on a side, the centre of which lies at an impact parameter of $\approx 30$ kpc from the quasar sightline hosting a high-column density DLA. Further, a compact UV source is found at $\approx 20$ kpc in projection from the quasar sightline, near the nebula. Combined, these pieces of evidence suggest the presence of a structure that contains neutral hydrogen

and stretches for over $\approx 50$ kpc in projection, giving rise to the DLA at the quasar position and the nebula. This structure further hosts one or multiple galaxies (see below).

A structure with these properties is reminiscent of the filamentary structures predicted by cosmological simulations, at the intersection of which galaxies form. Indeed, recent hydrodynamic simulations (Fumagalli et al., 2011; van de Voort et al., 2012; Rahmati & Schaye, 2014) consistently predict the presence of extended ($\gg 100$ kpc) filaments which give rise to absorption systems with $\log(N_{\rm HI}/{\rm cm}^{-2}) > 17$ in the surroundings of dark matter halos, and which host multiple satellite galaxies clustered to the main parent halo. Thus, interpreted in this context, our observations are probing the denser gas distribution either inside the circumgalactic medium (CGM) of galaxy G (with the caveat on the redshift discussed above) or inside a group of galaxies, with the observed Ly$\alpha$ emission being powered by star formation. These two scenarios are discussed in the reminder of this section. A more direct comparison between the observed system and the results of a recent cosmological hydrodynamic simulation follows in Sect.3.5.

### 3.4.1 Photoionization from galaxy G

With an average surface brightness $\Sigma_{\rm Lya} = (2.20 \pm 0.08) \times 10^{-18}$ erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$ within $\approx 30$ kpc from the star-forming galaxy G, the observed Ly$\alpha$ emission may arise from photoionization and/or collisional ionization in the CGM of this galaxy. Indeed, extended Ly$\alpha$ emission in excess of $\approx 10^{-18}$ erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$ is typically seen around LBGs or low-mass star-forming galaxies and it is ubiquitously predicted by numerical simulations (e.g. Steidel et al., 2011; Wisotzki et al., 2016; Furlanetto et al., 2005; Faucher-Giguère et al., 2010; Rosdahl & Blaizot, 2012).

Considering the morphology and location of the Ly$\alpha$ emission, we note that a model in which galaxy G is powering this nebula needs to account for both the off-centre position of the UV source compared to the Ly$\alpha$ emitting structure and for the

presence of substructures inside the nebula. Both requirements can be satisfied, at least qualitatively. First, offsets between the sites of star formation and regions of Ly$\alpha$ emission have been reported in the literature (e.g. Sobral et al., 2015). Moreover, density fluctuations (or the presence of sub-halos) in the CGM could account for the observed clumps.

The next condition that must be satisfied is whether, from an energetic point of view, galaxy G can power the observed emission. Given the observed SFR, we estimate that this source produces ionizing radiation at a rate of $\log(Q_{\rm H0}/s^{-1}) \approx 53.9$, according to a STARBURST99 model with constant star formation rate, Geneva stellar tracks without rotation (Leitherer et al., 1999; Lagarde et al., 2012), and a metallicity $Z = 0.002$ comparable to the one of the DLA ($\approx 0.1$ Z$_\odot$). We note that stellar population models predict an anti-correlation between metallicity and flux of ionizing radiation, thus galaxy G will produce a lower $Q_{\rm H0}$ if it has a metallicity higher than what is observed in the DLA. Assuming that a fraction $f_{\rm Ly\alpha} = 0.68$ of all recombinations gives rise to Ly$\alpha$ (in case B), galaxy G can account for a total luminosity of $L_{\rm Ly\alpha,G} \approx 8.5 \times 10^{42}$ erg s$^{-1}$, which is a factor of $\approx 3$ above the total observed luminosity of the nebula, $L_{\rm Ly\alpha,neb} = (2.7 \pm 0.1) \times 10^{42}$ erg s$^{-1}$. However, if galaxy G emits isotropically and the nebular emission arises from $\approx 1/4$ of the solid angle around the UV source, an escape fraction of $\approx 100\%$ in the direction of the nebula is required to power Ly$\alpha$ with photoionization alone.

Our calculation is complicated by an unknown dust extinction, which is likely present inside this structure. Indeed, the resonant Ly$\alpha$ emission can be easily absorbed by dust, making the luminosity inferred for the nebula only a lower limit. At the same time, galaxy G is likely to have an higher intrinsic star formation than what is inferred from UV observations, making also our current estimate for $Q_{\rm H0}$ a lower limit. As dust can easily absorb Ly$\alpha$ photons, as well as ionizing radiation, galaxy G appears to be marginally sufficient for powering the observed Ly$\alpha$ emission. Although potentially viable from an energy point of view, this interpretation requires a very high and possibly anisotropic escape fraction, together with some degree of fine-

tuning to reproduce the observed clumpiness inside the nebula. It also does not naturally explain the apparent kinematic alignment between the absorption and emission, which would require correlation of substructures within the CGM, but on opposite side of the galaxy star-forming region. Thus, additional ionization mechanisms are likely responsible for, or contribute to, the observed Ly$\alpha$ emission.

### 3.4.2 Contribution from *in situ* star formation

A perhaps more natural explanation for the extended emission is the presence of *in situ* star formation within the two clumps. It is in fact possible that clump $C_{\rm E}$ and $C_{\rm W}$ are actively forming stars inside an extended gas rich structure, of which galaxy G is a member under our working assumption that this galaxy lies at the DLA redshift. Indeed, after masking galaxy G, the lack of extended UV continuum within the two clumps sets the rest-frame equivalent width to $> 60.7$ Å and $> 52.3$ Å, respectively, implying that $C_{\rm E}$ and $C_{\rm W}$ are being powered by a recent burst of star formation if the source of radiation is local (see below). If Ly$\alpha$ is powered *in situ*, star formation could be occurring inside a $\approx 40$ kpc proto-disk that is fragmenting in two clumps. However, the large separation of these clumps would imply the existence of a very extended star-forming disk at $z \approx 3$, and therefore it appears more likely that star formation occurs instead inside two low-mass galaxies that are on the verge of merging (e.g. Stierwalt et al., 2015; Ribeiro et al., 2017).

This scenario is quantitatively consistent with the results of STARBURST99 calculations, in which we follow the evolution of a single burst of $10^6$ $M_\odot$ for 1 Gyr using again the Geneva stellar tracks with no rotation and with metallicity $Z = 0.002$ ($\approx 10\%$ solar). As before, the Ly$\alpha$ luminosity is computed from the photon flux of ionizing photons assuming case B recombination for gas at $T = 10^4$ K. The Ly$\alpha$ equivalent width is derived mimicking the observational techniques by computing the UV continuum luminosity by averaging the stellar continuum over the wavelength interval covered by the MUSE observations. With this STARBURST99 model, we

recover the well-known result of a decreasing equivalent width as a function of time from the burst (e.g. Charlot & Fall, 1993), a trend which we use to constrain the age of the burst.

By comparing our stellar population synthesis model to the inferred lower limits for the equivalent width (which are independent on the assumed mass of the burst), we infer an upper limit on the age of the burst of $\lesssim 7$ Myr. At the same time, an estimate of the mass formed during this star formation event is obtained by scaling the predicted Ly$\alpha$ luminosity (which is directly proportional to the mass of the burst) to match the observed value. As the luminosity is a function of time, we conservatively assume the maximum age of the burst allowed by the observed equivalent widths. With this comparison, we find that the two bursts in $C_{\rm E}$ and $C_{\rm W}$ account for a total stellar mass of $\approx 2 \times 10^7$ M$_\odot$ and $\approx 3 \times 10^7$ M$_\odot$, respectively for the E and W clump. It should be noted that this simple mass estimate does not provide a direct measurement for the total mass associated with the DLA in this structure, as it accounts only for stars formed recently inside the two clumps under the assumptions described above.

A more extended star formation history would indeed contribute to additional mass in stars, a possibility that can be tested with observations in the rest-frame optical. Further, we emphasise that radiative transfer effects combined with an unknown dust extinction make these estimates uncertain. However, our limits can be considered conservative, as plausible corrections would increase the equivalent width, resulting in an even younger and more massive burst. Future deep observations in the far IR rest-frame will be able to assess the importance of dust-obscured star formation in this system.

Owing to more degrees of freedom, a model in which *in situ* star formation powers Ly$\alpha$ appears to provide a more natural explanation for the observed clumps and for the kinematic alignment between emission and absorption spectra. Indeed, the DLA absorption can be interpreted as extended gas in the CGM of the two interacting galaxies, each of which retains the bulk of the halo systemic velocity. The interaction

between the two systems may also induce a tidal force that pushes neutral hydrogen to large distances inside the CGM, as observed for instance in nearby interacting dwarfs (e.g. Pearson et al., 2016).

Gravitational interactions that remove enriched material from the interstellar medium of these galaxies can also account for the observed metallicity of the DLA, which is significantly above the mean metallicity of the IGM and CGM at these redshifts (e.g. Schaye et al., 2003; Fumagalli et al., 2016). Indeed, a $\approx 10\%$-solar metallicity resembles the metal content of the interstellar medium of $\approx 10^{8.5}-10^{9}$ $M_{\odot}$ galaxies according to the $z \approx 3$ mass-metallicity relation (Mannucci et al., 2009). In addition to tidal interactions, galactic outflows are also a plausible mechanism to enrich the DLA. However, a wind originating from the two bursts in clump $C_{\rm E}$ and $C_{\rm W}$ would have to travel at a velocity $v > 4000$ km s$^{-1}$ to reach the DLA in $< 7$ Myr. Thus, it is more plausible that older episodes of star formation, or winds from galaxy G and other galaxies below our detection limit, have contributed to the observed enrichment.

Given the above discussions, it is clear that we cannot draw a firm conclusion on the nature of this system based on current data. Our analysis, however, favours a model in which the quasar sightline is probing an extended gas rich structure, within which two low-mass galaxies undergo a recent burst of star formation triggered by their encounter. Galaxy G likely lies within the same structure, or at comparable redshift within a group. Future observations targeting rest-frame optical lines and dust from these galaxies will be able to further refine the relative position/velocities of the members of this structure, as well as to refine our estimates on what powers the extended Ly$\alpha$ emission, thus providing more definite information on this system.

## 3.5 Comparison with the EAGLE simulations

With only a handful of DLA sightlines being targeted by MUSE at present, we have limited empirical information on the frequency of extended nebulae near DLAs. To

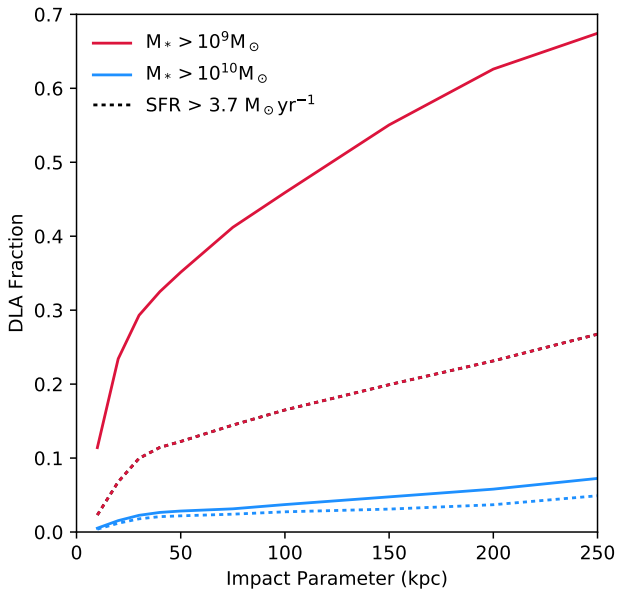


Figure 3.7: Fraction of DLAs in the EAGLE L0025N0752 simulation with column density $\log(N_{\rm HI}/\ {\rm cm}^{-2}) \geq 20.55$ that intersect a galaxy at an impact parameter smaller than the value plotted along the horizontal axis, in proper kpc. Red and blue curves are computed requiring the galaxy to have a stellar mass greater than $10^9$ and $10^{10}$ $M_\odot$, respectively. The dashed lines are computed requiring that, in addition to the mass constraints, the SFR of the galaxy exceeds 3.7 $M_\odot$ yr$^{-1}$.

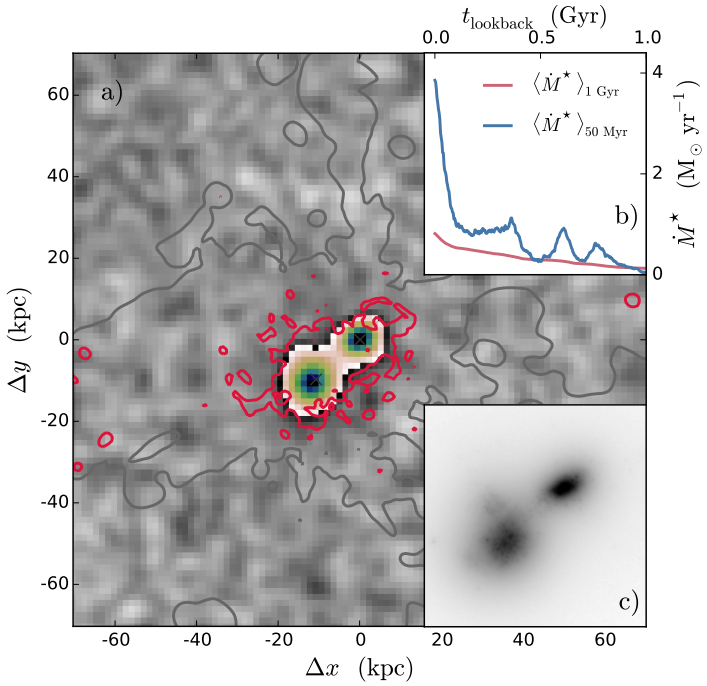


Figure 3.8: Mock observations of a system of two star-bursting galaxies associated with a DLA and exhibiting Ly$\alpha$ emission, extracted from the EAGLE L0025N0752 simulation, with properties similar to the system observed in MUSE. *Panel a:* Mock Ly$\alpha$ image (with no radiative transfer), with the grey scale representing the noise layer matched to the MUSE data, revealing two bright Ly$\alpha$ clumps at a separation of $\approx$ 20 kpc. The colour map represents the intensity of the emission for an unresolved source. The red contour marks the H I column density of $\log(N_{\rm HI}/\ {\rm cm}^{-2}) = 20.6$, showing that this system is embedded in an extended DLA with size comparable to that inferred by our observations. The grey contour at $\log(N_{\rm HI}/\ {\rm cm}^{-2}) = 16$ highlights how this simulated system is embedded in an extended gas filament. *Panel b:* The reconstructed star formation history of the two clumps combined, with $\dot{M}_\star$ averaged over the past 50 Myr and 1 Gyr shown in blue and red, respectively, illustrating that this system has a long-term star formation history consistent with the lack of UV detection in our observations, while it has experienced a recent burst that formed a stellar mass of $\approx 4 \times 10^7$ $\rm M_\odot$ (in agreement with observations) that could power the Ly$\alpha$ emission. *Panel c:* The $r$-band image reconstructed from the star formation history shows that this system is undergoing a merger event, which induced the two synchronous bursts of star formation.

offer additional insight into the probability of uncovering similar systems in the future, we turn to the results of the EAGLE cosmological simulations (Schaye et al., 2015; Crain et al., 2015). A general prediction of these and other simulations is that DLAs are clustered with galaxies of a wide range of masses in ΛCDM (e.g. Pontzen et al., 2008; Rahmati & Schaye, 2014). Here, we focus on the more specific question of how common systems like the observed one are in simulated volumes at $z \sim 3$. To address this question, we start by asking what fraction of DLAs lies in proximity to galaxies similar to galaxy G. Next, we estimate how common DLAs are in group environments, concluding with a search of a close-analogue to the observed system.

### 3.5.1 The EAGLE simulations and radiative transfer post-processing

EAGLE is a suite of cosmological hydrodynamic simulations performed in cubic volumes with co-moving lengths ranging from $L = 12.5$ to 100 co-moving Mpc, with cosmological parameters taken from the Planck Collaboration et al. (2014a) analysis. Simulations are performed using the GADGET-3 incarnation of the tree-SPH code described by Springel (2005), with modifications to the hydrodynamics and time-stepping schemes (see Schaller et al. (2015) for details), and the inclusion of a range of subgrid models to account for unresolved physics below the resolution scale, calibrated as described by Crain et al. (2015). Briefly, the simulations incorporate the major ingredients required to model galaxy evolution, including element-by-element radiative cooling and photoheating, a pressure-dependent star formation law above a metallicity-dependent density threshold, mass loss from stars and supernovae, and thermal feedback associated with star formation and active galactic nuclei[1].

Here we use two different EAGLE runs from table 2 in Schaye et al. (2015). To examine the correlation between galaxies and DLAs we use the higher resolution run L0025N0752 ($L = 25$ Mpc, SPH particle mass $m_g = 2.26 \times 10^5$ $M_\odot$, co-moving

[1]Details on these implementations can be found in Schaye et al. (2015), together with relevant discussion on differences with previous simulations, e.g. OWLS (Schaye et al., 2010).

Plummer-equivalent softening $\epsilon_{\rm com} = 1.33$ kpc). To investigate the presence of multiple galaxies in the same structure, we use instead run L0100N1504 ($L = 100$ Mpc, $m_g = 1.81 \times 10^6$ M$_\odot$, $\epsilon_{\rm com} = 2.66$ kpc), which has better galaxy statistics. Relevant to the current analysis is that EAGLE reproduces the $z \approx 3$ galaxy stellar mass function to within current observational uncertainties, although the specific SFRs are low by $0.2 - 0.5$ dex if the data are taken at face value (Furlong et al., 2015).

We use the URCHIN reverse-ray tracing code described by Altay & Theuns (2013) to compute the neutral fraction of all gas particles, assuming the IGM to be photoionized at the rate computed by Haardt & Madau (2001). Projecting the neutral hydrogen density of all gas particles onto a side of the simulation box allows us to evaluate the H I column density for $8192^2$ sightlines parallel to a coordinate axis of the cube. Applied to EAGLE, this procedure reproduces the observed number of DLAs (as well as Lyman limit systems) per unit absorption distance as a function of column density very well (see Rahmati et al., 2015), at least as well as the OWLS simulation analysed in the same way by Altay et al. (2011). Galaxies in EAGLE are identified as self-bound substructure inside dark matter halos, using the SUBFIND algorithm of Springel et al. (2001b) and Dolag et al. (2009). Next, we compute the stellar mass and SFR for each galaxy, the position and peculiar velocity of their centres of mass, as well as the mass of their parent dark matter halos (see McAlpine et al., 2016). Finally, we correlate DLAs to galaxies, by combining the H I column through the EAGLE volume with the location of galaxies.

We note how, in the previous section, we have explored the relevance of ionizing radiation produced by local sources in powering the observed Ly$\alpha$ emission in the nebula. However, when comparing with the results of the EAGLE simulations, we do not include photoionization from local sources for the following reason. While local radiation is clearly relevant to model in detail the particle-by-particle ionization state of the gas and its emissivity near galaxies, in practice, the inclusion of radiative transfer effects from local sources is subject to uncertainties related to the unresolved

structure of the interstellar medium within large cosmological simulations. These uncertainties are particularly significant at high column densities ($N_{\rm HI} \gtrsim 10^{21}$ cm$^{-2}$; Rahmati et al., 2013), which are relevant for this work.

Therefore, we refrain from deriving predictions of the Ly$\alpha$ luminosity from DLAs in EAGLE, focusing instead on the more basic task of comparing the general DLA population with galaxies in the EAGLE volume. As discussed in the literature (e.g. Fumagalli et al., 2011; Rahmati et al., 2013), while local sources of radiation may alter the cross section of individual absorbers, extended DLAs next to galaxies persist even after including local sources. This implies that associations between DLAs and galaxies in simulations are generally robust even without inclusion of local sources during the radiative transfer post-processing. Nevertheless, in the following, we provide a qualitative description of how sensitive our conclusions are to the assumptions made during radiative transfer post-processing.

### 3.5.2 Statistical analysis of DLAs associated to galaxies in EAGLE

We estimate the probability with which systems like the one uncovered by our study can be found in future observations by first quantifying the fraction of DLAs that resides near galaxies with properties similar to galaxy G (with the caveat on the redshift discussed above). For this, we begin by identifying all DLA sightlines with $\log(N_{\rm HI}/\ {\rm cm}^{-2}) \geq 20.55$, which is the minimum column density of the observed DLA as discussed in Sect. 3.3.1. The likelihood of observing a galaxy with properties similar to that of galaxy G (see Sect. 3.3.3) within a given impact parameter from a DLA with this column density is shown in Figure 3.7.

The observed projected distance between the DLA and galaxy G is $\approx 20$ kpc, but the galaxy stellar mass is relatively uncertain, and therefore we show curves for two choices, $M_\star = 10^9$ and $10^{10}$ M$_\odot$ (red and blue solid curves, respectively). From this analysis, it is clear that finding a DLA-galaxy pair with such a close impact

parameter has a probability of $\approx 23\%$ and $\approx 2\%$, respectively for the two choices of mass. We note that this probability would further increase if we had to consider a higher column density for the DLA. When we require that the galaxy undergoes star formation at a rate exceeding 3.7 $\mathrm{M_\odot\ yr^{-1}}$, we find that these probabilities drop to 7% and 1%, respectively (dashed lines in Figure 3.7).

As discussed above, since we do not include local sources in our calculation, the total number of DLAs recovered in this calculation is formally an upper limit. However, DLAs sightlines are clustered in structures with a range of column densities, implying that neutral gas with column densities above the $\log(N_{\rm HI}/\ \mathrm{cm^{-2}}) \geq 20.55$ threshold is likely to be found in the same regions identified by our study even when radiative transfer effects from local sources are included. For this reason, we regard our conclusions as largely insensitive to the details of the radiative transfer post-processing.

Next, we investigate how frequently a DLA sightline with $\log(N_{\rm HI}/\ \mathrm{cm^{-2}}) \geq 20.55$ passes near a galaxy with properties similar to those of clumps $C_{\rm W}$ and $C_{\rm E}$ combined. In this case, we require that the SFR averaged over the past 10 Myr is $\dot{M}_\star(10\ \mathrm{Myr}) > 3\ \mathrm{M_\odot\ yr^{-1}}$, whereas when averaged over the past 1 Gyr is $\dot{M}_\star(1\ \mathrm{Gyr}) < 5\ \mathrm{M_\odot\ yr^{-1}}$. This is in line with our estimates from Sect. 3.3.3. For the observed impact parameter between the clumps and the DLA of 30 kpc, we find that these criteria apply to $\approx 3\%$ of the selected DLA sightlines in EAGLE. Again, we regard these statistics as largely independent of radiative transfer effects for the reasons discussed above.

Finally, we use EAGLE to quantify the joint probability of finding a galaxy with properties consistent with those of galaxy G near two star-forming galaxies consistent with $C_{\rm W}$ and $C_{\rm E}$. For this calculation, we turn to the $L = 100$ Mpc box to increase the statistical significance of this analysis. To mimic our observations, we require that one galaxy has $M_\star \geq 10^9\ \mathrm{M_\odot}$ (as galaxy G), which is within a projected separation $\leq 50$ kpc of two star-bursting galaxies with $\dot{M}_\star(10\ \mathrm{Myr}) > 2\ \mathrm{M_\odot\ yr^{-1}}$, in line with the inferred properties of the two clumps which form $\approx 2 \times 10^7\ \mathrm{M_\odot}$ of stars in $\approx 10$ Myr. To match the observed equivalent widths, we further require

that $\dot{M}_\star(1\ \mathrm{Gyr}) < 5\ \mathrm{M}_\odot\ \mathrm{yr}^{-1}$. Finally, we demand that the two galaxies are less than 30 kpc apart (projected distance), to mimic the observed separation between $C_\mathrm{E}$ and $C_\mathrm{W}$.

After imposing these constraints, we find 21 such triples in the EAGLE run with $L = 100$ Mpc. In most of the cases, the triple corresponds to a situation where all three galaxies are members of the same galaxy group, with halo mass in the range $10^{12} - 10^{13}\ \mathrm{M}_\odot$. We note that for each pair ($C_\mathrm{E}$ and $C_\mathrm{W}$), there are several nearby galaxies that are consistent with galaxy G, implying that other star-forming galaxies are predicted in the surroundings of DLAs associated with these groups. The velocity offset between these galaxies ranges between $100 - 500\ \mathrm{km\ s}^{-1}$, consistent with the virial velocity[2] of

$$V_c = (10\mathrm{G}H(z)M_{200})^{1/3} = 240\ \mathrm{km\ s}^{-1}\,(M_{200}/10^{12}\mathrm{M}_\odot)^{1/3}, \qquad (3.5.1)$$

in groups of this halo mass at this redshift. Such offsets are consistent with those found between galaxy G, the clumps, and the DLA. We conclude from this that the observed system is consistent with the DLA being associated with a galaxy group, of which G, $C_\mathrm{E}$ and $C_\mathrm{W}$ are members under our working hypothesis that $z_\mathrm{G} = z_\mathrm{dla}$. Such groups are relatively rare at $z = 3.25$, with a number density of $\approx 10^{-4}$ per comoving Mpc$^3$, with typical radius of $R_{200} = 100$ physical kpc, a size which is consistent with the observed projected separations. Only a small fraction of such groups contain two star bursting and merging galaxies of the types of the observed clumps.

Altogether, we conclude from this comparison with the results of the EAGLE simulation that systems such as the one observed - consisting of a DLA, with a nearby relatively massive galaxy and a further two lower mass star-bursting galaxies - do occur in simulations, although not with high number density. In the majority of cases, such systems are associated with a group of galaxies, with the DLA also

[2]We employ the usual definition where $R_{200}$ is the radius of a halo, defined such that the mass within $R_{200}$ equals $M_{200} = 200 \times (4\pi/3)\rho_c\, R_{200}^3$, where $\rho_c$ is the critical density.

associated with the group. The velocity offsets amongst the galaxies themselves, and between the galaxies and the DLA, are predicted to arise from virial motions in a dark matter halo of mass $10^{12} - 10^{13}$ $M_\odot$. This result implies that, for the case of groups, individual galaxy-absorber associations cannot be trivially established, and that these do not provide a complete description of the systems probed in absorption.

### 3.5.3 Close analogue of the observed system

We conclude this section by briefly discussing the properties of a close analogue of the observed association of two star bursting galaxies near a DLA, which we have identified during this analysis within the EAGLE L0025N0752 run. This simulated object offers an example that configurations like the observed one - with two galaxies in synchronous star-burst phases inside an extended DLA - are plausible within the current cosmological model.

The simulated system, which is shown in Figure 3.8, is composed of two galaxies that are undergoing a merger that triggers two synchronous bursts of star formation, and which form a total of $\approx 4 \times 10^7$ $M_\odot$ of stars within $\approx 10$ Myr, similar to the observed clumps $C_\mathrm{W}$ and $C_\mathrm{E}$. Prior to this encounter, the two galaxies were forming stars at a rate $\lesssim 1$ $M_\odot$ yr$^{-1}$, and hence, again similar to clumps $C_\mathrm{W}$ and $C_\mathrm{E}$, would not have been detected in our current UV continuum observations. Inset *b* shows the recent star formation history of both clumps combined, measured back in time with look back time $t = 0$ chosen to correspond to the time at which the system is observed in the simulation. This close analogue is further embedded within a $\log(N_\mathrm{HI}/\ \mathrm{cm}^{-2}) \geq 20.6$ DLA with an extent of $\gtrsim 40$ kpc (red contours), which is part of a large-scale filament, as traced by lower column density neutral hydrogen (grey contours). Being related to the properties of a single object, we caution that this part of the analysis is more prone to uncertainties introduces by our assumptions on radiative transfer. The simulation further predicts a mass-weighted metallicity for this system of about $\approx 10\%$ solar, again close to the observed value.

Using the star formation history known from the simulation, we generate a mock Ly$\alpha$ image (background grey scale in the main panel) with depth and resolution comparable to the MUSE data, by scaling the Ly$\alpha$ luminosity from the stellar population synthesis model described above according to the stellar mass generated within the simulation during the burst. As before, we neglect radiative transfer effects and dust absorption. The resulting map is computed for the Ly$\alpha$ luminosity close to the peak of the burst (at $< 1$ Myr). We also generate a rest-frame $r$-band image of the stellar continuum at full resolution following the method discussed in Trayford et al. (2017), shown in panel *c*.

## 3.6 Summary and Conclusions

We have presented new MUSE observations of the quasar field J025518.58+004847.6 ($z_{\rm qso} \approx 3.996$), which is known to host an intervening DLA at $z_{\rm dla} = 3.2552 \pm 0.0001$ with $\log N_{\rm HI} > 20.55 \pm 0.10$ cm$^{-2}$ (with a best estimate of $\log N_{\rm HI} = 20.85 \pm 0.10$ cm$^{-2}$) and metallicity $\log Z/Z_{\odot} < -0.8 \pm 0.1$ (with a best estimate of $\log Z/Z_{\odot} = -1.1 \pm 0.1$). By reaching a flux limit of $> 2 \times 10^{-18}$ erg s$^{-1}$ cm$^{-2}$ ($3\sigma$), our observations uncover a $37 \pm 1$ kpc extended structure that emits in Ly$\alpha$ with a total luminosity of $(27 \pm 1) \times 10^{41}$ erg s$^{-1}$. The centre of this nebula is located at a projected distance of $30.5 \pm 0.5$ kpc from the quasar sightline.

This nebula also contains two distinct clumps, for which no continuum is detected to rest-frame equivalent limits of $> 60.7$ Å and $> 52.3$ Å. The two clumps are separated by $16.5 \pm 0.5$ kpc in projection and have a line-of-sight velocity difference of $\Delta v_{\rm E-W} = 140 \pm 20$ km s$^{-1}$, which is consistent with the velocity difference of $\Delta v_{1-2} = 155 \pm 6$ km s$^{-1}$ measured in between the two main absorption components seen in the metal transitions associated with the DLA.

Furthermore, a compact galaxy is detected in the continuum at a projected distance of $19.1 \pm 0.5$ kpc from the quasar sightline. The MUSE spectrum of this source is noisy, but with properties consistent with a $z \approx 3.2$ star-forming galaxy. Given its

close proximity to the nebula and its photometric properties, we consider this galaxy to be physically associated with the structure hosting the DLA.

Based on current observations, two mechanisms appear plausible to power the extended Ly$\alpha$ emission. First, with a star formation rate of $\approx 4$ M$_\odot$ yr$^{-1}$, photoionization from the continuum-detected galaxy may ionize the galaxy CGM, powering the nebula. However, this scenario requires an elevated escape fraction ($\approx 100\%$) of ionizing radiation from the sites of star formation, although an unknown dust extinction hampers robust conclusions. A second viable model for powering the nebula is that Ly$\alpha$ emission arises from *in-situ* star formation inside the two embedded clumps, in the form of two young ($\lesssim 7$ Myr) starbursts likely triggered by a merger of two galaxies. This scenario also provides a satisfactory explanation, at least qualitatively, for the observed kinematic alignment between the two clumps and the main absorption components of the DLA. Finally, a combination of *in-situ* star formation and ionization from galaxy G may also account for the total energy budget. Additional follow-up observations at rest-frame optical and IR wavelengths are now required to better characterise the nature of this system, especially by confirming the redshift of the continuum detected galaxy and by constraining the presence of dust inside this structure.

Regardless of what is powering the observed emission, the position of the DLA and the extent of the Ly$\alpha$ nebula jointly suggest the presence of a gas-rich structure that stretches for $\gg 50$ kpc on a side, inside which one or multiple galaxies are forming stars. This structure is reminiscent of the filaments predicted by hydrodynamic cosmological simulations, at the intersection of which galaxies form. Within this model, we would expect a fraction of DLAs to be embedded within structures hosting galaxies that are surrounded by extended halos or filaments which radiate in Ly$\alpha$ emission (Cantalupo et al., 2012; Rauch et al., 2013; Wisotzki et al., 2016).

The association between extended Ly$\alpha$ nebulae discovered in deep long-slit spectroscopy or narrow-band images and the population of optically-thick absorbers, and particularly DLAs, has already been explored from a statistical point of view

by Rauch et al. (2008). This study suggests that low-surface brightness nebulae have sufficiently large number density and covering factor to account for the observed incidence of DLAs. Compared to the properties of the nebulae presented in this work, however, the emitters uncovered in the deep long-slit observations by Rauch et al. (2008) are generally more compact, with sizes $\lesssim 20$ kpc measured at $10^{-19}$ erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$. Only a few cases (4/27 objects) are found with sizes $\gtrsim 30$ kpc, which is similar to the extent of the nebula discovered in our MUSE observations. Moreover, the objects from Rauch et al. (2008) have a typical luminosity $\lesssim 10^{42}$ erg s$^{-1}$, again placing the nebula near the $z \approx 3.25$ DLA in the extreme tail of the population of these emitters.

The Ly$\alpha$ sources discussed in Rauch et al. (2008) represent "field" objects which are not explicitly associated to a DLA. However, examples of diffuse Ly$\alpha$ nebulae in proximity of DLAs, which exhibit characteristics similar to the system described in this work, can be found in the literature (e.g. Fynbo et al., 1999; Zafar et al., 2011; Kashikawa et al., 2014). In particular, while the system reported in Fynbo et al. (1999) and Zafar et al. (2011) lies in proximity to a quasar, the object presented by Kashikawa et al. (2014) is a close analogue to the system discussed here, being associated to an intervening DLA with a luminosity of $\approx 10^{42}$ erg s$^{-1}$ and a size of $\approx 15$ kpc as measured in a 2D spectrum at a depth of $2.5 \times 10^{-19}$ erg s$^{-1}$ cm$^{-2}$ Å$^{-1}$.

A question that arises from our study and previous work is how common extended nebulae near DLAs are. Our discovery of a $\approx 40$ kpc nebula in a field that had been studied previously with optical imaging campaigns yielding no detections (Fumagalli et al., 2014, 2015) hints that previous imaging and, to some extent, spectroscopic observations have been unable to identify these types of associations with high completeness. Hence, our study offers a clear example of the need for deep multiwavelength campaigns to map simultaneously the galaxy population and the more diffuse gas environment near DLAs.

Based on the number density estimates in Rauch et al. (2008) and our comparisons with the EAGLE simulations, it is however unlikely that nebulae with sizes $\gtrsim 30$ kpc

and luminosity $\geq 10^{42}$ erg s$^{-1}$ are common occurrence next to DLAs. Nevertheless, this system offers a tantalising example of how continuum-detected sources (especially near metal-rich DLAs; Krogager et al., 2017) may not represent isolated associations but trace instead clustered galaxy formation inside a web of filaments. In fact, the presence of DLAs in interacting groups has already been confirmed with observations in the literature. For instance, Warren & Møller (1996) and Weatherley et al. (2005) reported the detection of multiple emission-line galaxies in the fields of two DLAs at $z \approx 2$ and $z \approx 3$. The plausibility of DLAs being associated with clustered galaxy formation is also confirmed by our analysis of the EAGLE simulations.

At present, we cannot derive firm conclusions on the number density of these types of associations, as a systematic exploration of the incidence of extended nebulae and groups in proximity to DLAs has not been possible so far due to the difficulties of reaching flux limits of $10^{-18}$ erg s$^{-1}$ cm$^{-2}$ (Christensen et al., 2007). However, the deployment of sensitive large-format integral field spectrographs at 8-meter telescopes and of the full ALMA array will soon enable deep and complete multiwavelength searches for both star-forming galaxies (including dust-obscured systems) and diffuse emission near DLAs. These observations, in tandem with refined model predictions of the emission properties from dense filaments that are traced by strong absorption line systems, will offer a novel way to test if galaxy formation proceeds inside gas-rich filaments as predicted by current cosmological simulations.

# Chapter 4

# A MUSE Integral Field Survey of DLAs at $z \simeq 3.5$

We present results from a survey of galaxies in the field of six $z \geq 3$ damped Lyman $\alpha$ systems (DLAs) using the Multi Unit Spectroscopic Explorer (MUSE) at the Very Large Telescope (VLT). We report an unprecedented detection rate of up to $\approx 80\%$ for a representative sample that is not preselected in metallicity, finding five high-confidence Ly$\alpha$ emitters associated with three DLAs, plus up to 9 additional detections across five of the six fields. Among our detections, we report four counterparts associated with the most metal-poor DLA in our sample ($Z/Z_{\odot} = -2.33 \pm 0.22$), which trace an overdense structure resembling a filament. By comparing our detections with predictions from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) cosmological simulations and a semi-analytic model designed to reproduce the observed bias of DLAs at $z > 2$, we conclude that our observations support a scenario in which DLAs trace the neutral regions within and around galaxies clustered inside dark matter halos. Moreover, current detections suggest halo masses of $M_{\rm h} \approx 10^{11} - 10^{12}$ M$_{\odot}$ as typical for at least some $z \approx 3$ DLAs, in agreement with the inference made from the large-scale clustering of DLAs. We finally show how surveys conducted with integral field spectroscopy targeting $\approx 25$ absorbers have the potential of constraining the characteristic masses

of halos hosting high-redshift DLAs with sufficient accuracy to discriminate between different models.

## 4.1 Introduction

Since the discovery of damped Ly$\alpha$ absorbers (DLAs) in the spectra of quasars in the 1970s (Beaver et al., 1972; Carswell et al., 1975), significant efforts have been made to identify the properties of the galaxy population that gives rise to this class of absorption line systems. The interest in connecting DLAs to galaxies stems from the fact that these absorbers, defined to have neutral hydrogen column density in excess of $\log(N_{\rm HI}/{\rm cm}^{-2}) \geq 20.3$ (Wolfe et al., 2005), act as signposts of significant reservoir of neutral hydrogen within or around high-redshift galaxies (Wolfe et al., 1986). For this reason, direct associations between DLAs detected in absorption and galaxies detected in emission provide a powerful way to probe links between gas supply in the form of neutral hydrogen and ongoing star formation. The combination of absorption and emission techniques thus provides at present the only means to study the star formation law in atomic gas beyond $z > 2$ (Wolfe & Chen, 2006; Prochaska & Wolfe, 2009; Rafelski et al., 2011; Fumagalli et al., 2015; Rafelski et al., 2016).

Starting from earlier searches on the ground and with the *Hubble Space Telescope* (Warren et al., 2001; Møller et al., 2004), several surveys have attempted to identify the galaxy population that is giving rise to DLAs. Despite decades of searches, progress has been scarce until recently, when the advent of integral field spectrographs have enabled more complete spectroscopic surveys (e.g. SINFONI at the Very Large Telescope, VLT; **?**). Moreover, building on the evidence that galaxies obey a defined mass-metallicity relation (Tremonti et al., 2004; Maiolino et al., 2008), searches have focused on the high end of the metallicity distribution of DLAs, yielding higher detection rates with the discovery of tens of DLA hosts (Fynbo et al., 2010; Krogager et al., 2017). The much increased sensitivity of the Atacama Large Millimeter Array

(ALMA) is now also enabling searches of DLA galaxies via molecular and atomic lines, a technique that is being successfully pioneered, still, at the high end of the metallicity distribution (Neeleman et al., 2017; Kanekar et al., 2018).

While these recent identifications offer a way to finally study the link between column density and metallicity in absorption, and stellar masses and star formation rates (SFRs) in emission (Møller et al., 2004; Christensen et al., 2014; Krogager et al., 2017), these studies are likely to only probe the bright end of the DLA population, and may even not be fully representative of the diverse population of DLA host galaxies. Most simulations and models (Haehnelt et al., 1998; Pontzen et al., 2008; Barnes & Haehnelt, 2009; Rahmati & Schaye, 2014; Bird et al., 2014; Fynbo et al., 2008) consistently indicate that DLA hosts are generally to be found at the very faint end of the luminosity function, typically below the sensitivity limit of current searches. Indeed, our own survey of galaxies designed to image the DLA hosts at all impact parameters (O'Meara et al., 2006; Fumagalli et al., 2010, 2014, 2015) has yielded a series of non-detections within a few kiloparsecs of the location of the DLAs, despite removing the primary source of observational bias, that is the bright quasar emission which hampers the detection of faint galaxies.

While there seems to be general consensus that DLAs are primarily associated with faint galaxies, the question of what the typical range of halo masses giving rise to DLAs is still remains open. At face value, following a similar argument than the one adopted in abundance matching studies (e.g. Conroy et al., 2006), DLAs are expected to arise primarily in faint galaxies (e.g. Fynbo et al., 2008) and hence in low mass halos ($M_{\rm halo} \lesssim 10^{11} M_{\odot}$). However, this appears to be in tension with other pieces of evidence. Firstly, the distribution of velocity widths measured from metals in low ionisation states shows a prominent tail at high velocity, which has been used to argue for the existence of a population of large disks hosting DLAs (Prochaska & Wolfe 1998; but see Bird et al. 2015 for more recent work on this topic). Furthermore, Font-Ribera et al. (2012), using the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS), found a unexpected large linear bias of DLAs ($b_{DLA} = 2.17 \pm 0.20$, with

an assumption on the bias of the Ly$\alpha$ forest) by cross-correlating these absorbers with the Ly$\alpha$ forest. This measurement was more recently updated by Pérez-Ràfols et al. (2018b), who found a linear bias of $b_{DLA} = 2.00 \pm 0.19$, which is only slightly lower than the clustering amplitude of Lyman Break Galaxies (LBGs; see also Cooke et al., 2006). This value of the bias, which implies masses of $\gtrsim 10^{11}$ M$_\odot$, appears uncomfortably high for some galaxy formation simulations and models (see e.g Pontzen et al. 2008; Barnes & Haehnelt 2014) if DLAs sample a very wide range in galaxy sizes (e.g. Krogager et al. 2017).

Building on our previous searches for galaxies at small impact parameters from the quasars (Fumagalli et al., 2015), this study exploits the power of wide field integral field spectroscopy provided by the Multi Unit Spectroscopic Explorer (MUSE) at the VLT (Bacon et al., 2010) to carry out the first highly-complete spectroscopic survey in an unbiased sample of DLAs with the goal of searching for faint Ly$\alpha$ emission from galaxies associated with DLAs. Our survey improves upon previous searches of this type, which were characterised by a lower sensitivity (e.g. Christensen et al., 2007) or smaller field of view (e.g. **?**). By achieving a flux-limited search up to $\approx 250$ kpc from the DLAs, we are finally able to search at the same time for the DLA hosts and for other associations, which provide a means of characterising the small-scale clustering of DLAs with tracers that are independent of what adopted by previous studies.

Throughout this work, we define a host as a galaxy that is physically connected with the absorbing gas. For detected galaxies at small impact parameters, i.e. within a projected distance $b_{\rm impact} \leq 50$ kpc, we will consider if they could plausibly be linked to the absorbing material. These values, albeit arbitrary, define a reasonable range of distances and velocities that encompass the inner circumgalactic medium (CGM) of a galaxy. We define more generally an association a galaxy that is physically connected to the DLA (e.g. in the same halo or clustered on small scales of up to a few hundred kiloparsecs), but not necessarily the one from which the absorption arises. When comparing our observations with simulations we use a definition of

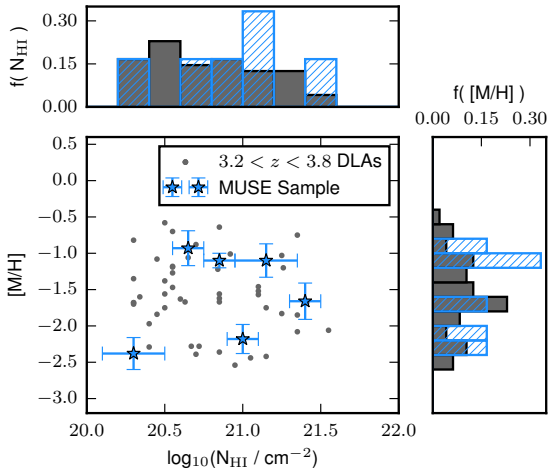


Figure 4.1: The metallicity and column density distributions of the DLAs targeted with MUSE, compared to the parent DLA population over the same redshift interval ($3.2 < z < 3.8$) from Rafelski et al. (2012). The normalized distribution of metallicity is shown on the right for both samples, and similarly at the top for column density. Our sample spans a wide range of DLA properties, thus extending the parameter space covered by most recent surveys.

$\Delta v_{\rm DLA,gal} \leq 650$ km s$^{-1}$ as a definition of associated.

This paper is structured in the following way. The observations, data reduction and analysis are presented in Sect. 4.2 and 4.3, followed by a discussion of the properties of each field is continued from Sect. 4.4 in appendix A.2 for the remaining fields. Sect. 4.5 presents technical details on the simulations and semi-analytic models used in the interpretation. We briefly describe some highlighted detections in Sect. 4.4. The discussion of results is presented in Sect. 4.6, followed by a summary and conclusion in Sect. 4.7.

Magnitudes reported in this paper follow the AB system and fluxes and magnitudes have been corrected for Galactic extinction following Schlafly & Finkbeiner (2011). All spectral data products are reported in vacuum wavelengths, and we adopt the Planck Collaboration et al. (2016b) ΛCDM cosmological parameters ($\Omega_{\rm m} = 0.308$, $h = 0.677$).

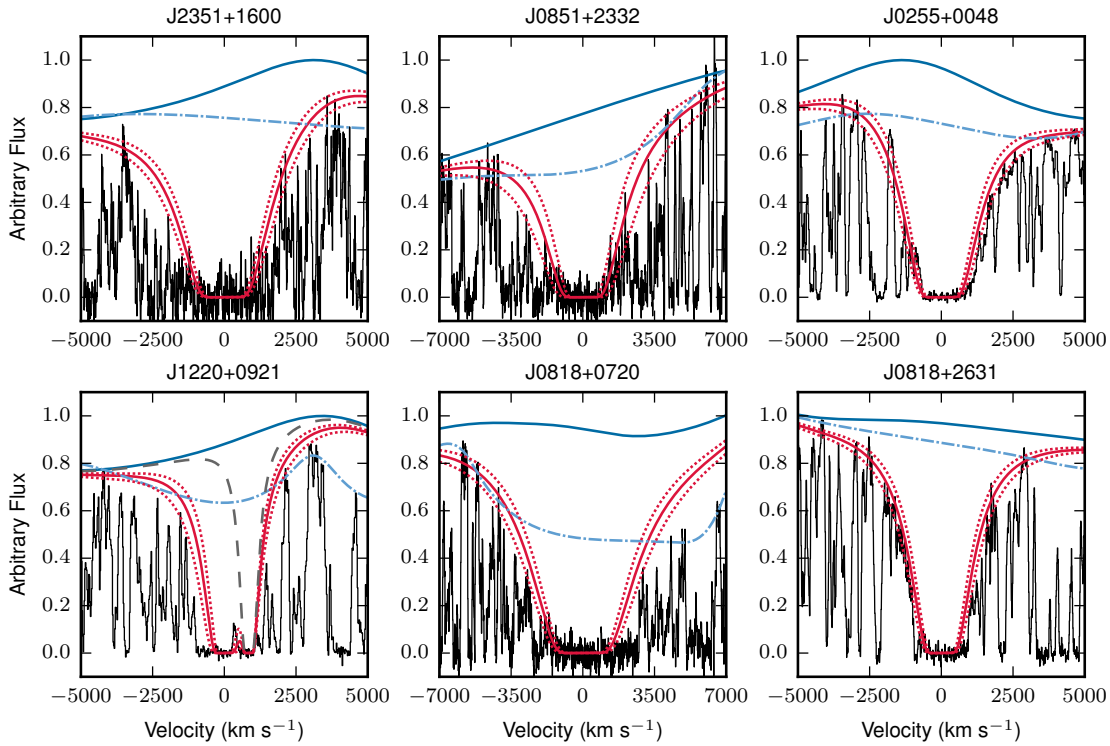


Figure 4.2: The Ly$\alpha$ absorption profiles for the six DLAs in the MUSE sample. The quasar spectra are shown in black, the fitted quasar composite template are in blue (solid line), and the template multiplied by the Voigt profile is in red (solid line). Errors ($1\sigma$) on the column density are show by red dotted lines. The original continuum normalisation from Fumagalli et al. (2014) is also shown as a blue dash-dotted line. For J0851+2332 a manual spline continuum normalisation was adopted instead of the quasar composite. The fitted contribution of the Ly$\beta$ line from a pDLA contaminating the profile for J1220+0921 is also plotted with grey dashed line.

## 4.2 Observations and data reduction

### 4.2.1 Sample selection

The six quasar sightlines observed with MUSE are chosen from the parent sample of Fumagalli et al. (2014) which is selected among the general DLA population purely based on the presence of two optically-thick absorbers along the quasar sightline, with the goal of searching for the rest-frame UV emission at close impact parameters from the DLAs (see O'Meara et al., 2006, for details). For this study, we select DLAs that are observable from VLT and at $z > 3.2$, which is the redshifts at which Ly$\alpha$ falls at wavelengths where the throughput of MUSE is $> 25\%$. Excellent high-resolution spectroscopy are available for this sample, which spans a wide range of metallicity and column density.

The properties of the selected DLAs in our sample are summarised in Table 4.1, while Figure 4.1 shows the MUSE sample in context with the wider DLA population in terms of metallicity and column density, highlighting how our targets span a representative range of both parameters. This is in contrast to many recent searches for DLA hosts, such as those conducted with X-Shooter (Krogager et al., 2017) and ALMA (Neeleman et al., 2017), which typically select high metallicity systems ([Si/H] $\gtrsim -1$) to exploit the metallicity-luminosity relation between the metallicity of the DLA in absorption and the luminosity of the host galaxy to ensure higher detection ratio compared to samples selected, e.g., only on $N_{\rm H\,I}$ (e.g. Péroux et al., 2012). This approach, however, leaves the hosts of the majority of "typical" DLAs at $z > 2$ unexplored. Our DLA sample contains instead three systems above average metallicity at this redshift (Rafelski et al., 2012) and three below, thus extending the parameter space targeted by previous searches. Notably, DLA J1220+0921 in our sample is among the most metal-poor DLAs, sitting at $Z/Z_{\odot} = -2.33$ just above the metallicity floor at $Z/Z_{\odot} \simeq -3$ (Prochaska et al., 2003; Rafelski et al., 2012).

| QSO Name | Field[a] | R.A. | Dec | $z_{\rm DLA}$ | $\log(N_{\rm HI}/\ {\rm cm}^{-2})$ | $\log\ (Z/Z_{\odot})$ | Element[b] | Exp. Time (s)[c] |
|---|---|---|---|---|---|---|---|---|
| J2351+1600 | 06G6 | 23:51:52.80 | +16:00:48.9 | 3.7861 | 21.00 ± 0.10 | -2.18 ± 0.20 | Fe | 6x1480 |
| J0851+2332 | 10G11 | 08:51:43.72 | +23:32:08.9 | 3.5297 | 21.15 ± 0.20 | -1.10 ± 0.23 | Zn | 4x1480 |
| J0255+0048 | 15H3 | 02:55:18.58 | +00:48:47.6 | 3.2530 | 20.85 ± 0.10 | -1.10 ± 0.10 | Si | 6x1480 |
| J1220+0921 | 19H7 | 12:20:21.39 | +09:21:35.7 | 3.3090 | 20.30 ± 0.20 | -2.33 ± 0.22 | Si | 6x1480 |
| J0818+0720 | 23H11 | 08:18:13.14 | +07:20:54.9 | 3.2332 | 21.15 ± 0.10 | -1.66 ± 0.25 | Si;Zn | 6x1480 |
| J0818+2631 | 24H12 | 08:18:13.05 | +26:31:36.9 | 3.5629 | 20.65 ± 0.10 | -0.93 ± 0.24 | Si;Zn | 2x1480 |

Table 4.1: The properties of the MUSE DLA sample.[a] The naming system from Fumagalli et al., 2014. [b] The element(s) used to estimate the DLA metallicity. [c] The exposure time of the MUSE integral field spectroscopy.

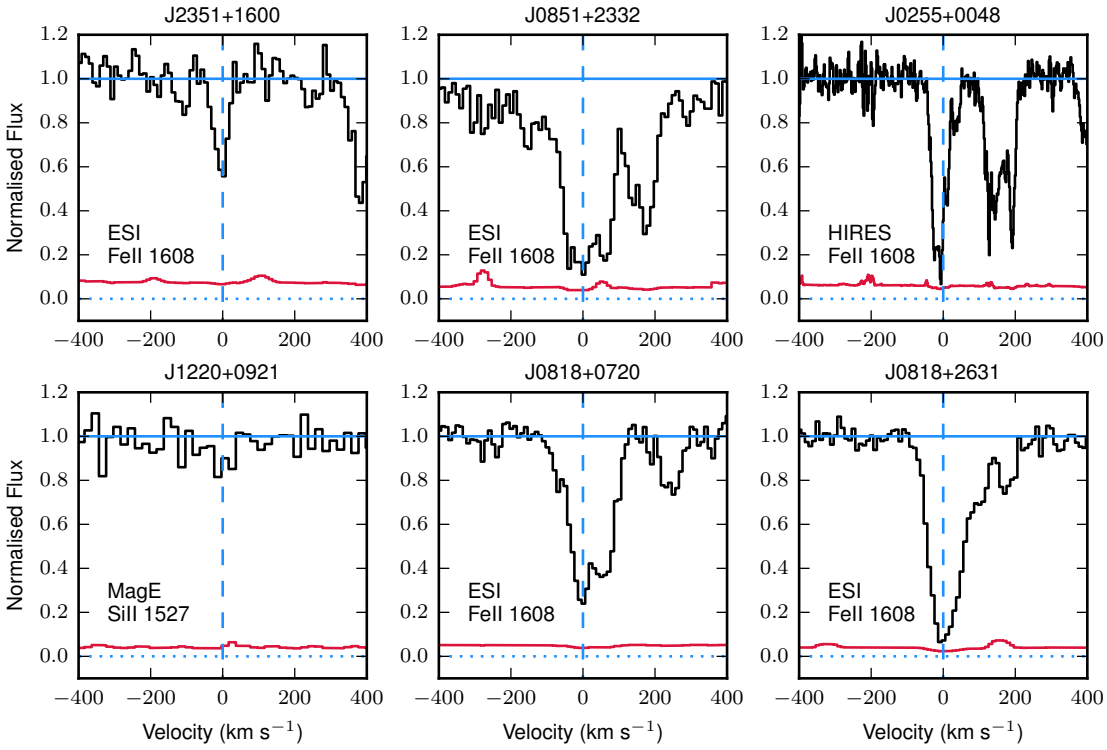


Figure 4.3: Selected low-ionisation lines from each DLA in our sample. The spectra (black) are continuum normalised, with the $1\sigma$ flux error shown in red. In each panel the instrument and transition are listed. Lines indicating zero flux (blue dotted line), the normalised continuum level (1.0, blue solid line) and the zero velocity (blue dashed line) are included.

### 4.2.2 MUSE observations and data reduction

MUSE integral field spectroscopy of the six DLA fields was acquired in service mode between June 2015 and April 2016 under ESO programmes 095.A-0051 and 096.A-0022 (PI Fumagalli). Observations were split into sets of 1480s exposures. While four fields were completed with 6×1480s exposures, J0851+2332 and J0818+2631 were only partially completed with 4×1480s and 2×1480s respectively. The observations were taken in dark time with seeing ranging from 0.7 arcsec to 0.9 arcsec using the Nominal Wide Field Mode, with clear conditions. In each sub-exposure the quasar was centred in the field of view, and small dithers combined instrument rotations in increments of 90° were made to improve the quality of the final data product.

The initial reduction of the data is carried out using the ESO MUSE pipeline (Weilbacher et al., 2014) (v1.6.2). The pipeline carries out bias, dark and flat field corrections, calibrates the data in wavelength and astrometry, and applies a basic illumination correction. This initial reduction is however limited by the accuracy of the illumination correction and therefore we additionally post-process the data with CubExtractor (Cantalupo, in prep.) to further improve the quality of the illumination correction and sky subtraction (see, e.g. Borisova et al. 2016; Fumagalli et al. 2016, 2017a for details).

Following this post-processing, the final datacubes are created from the mean of all sub-exposures, and we additionally produce median coadds and datacubes from even and odd numbered sub-exposures to produce independent sets of data for verification processes (see Sect 4.3.2). The final data product for each field covers a field of view of approximately $1 \times 1$ arcmin$^2$, covering $4750 - 9354$ Å with 1.25 Å binning. Regions of sub-exposures affected by stray-light near the edges of the cube are masked before stacking. We also mask 2-5 pixels around the edge of the combined cubes to remove low quality data and very noisy pixels. The field of DLA 15H3 further requires substantial masking due to a bright (V= 11.2) star lying at 46 arcsec from the quasar, resulting in a slightly smaller final field of view.

As a last step, we calibrate the absolute astrometry and verify the quality of the data against imaging and spectra from the Sloan Digital Sky Survey (SDSS; York et al. 2000). The absolute astrometry of the datacubes is calibrated relative to the quasar position, keeping the relative astrometry within each field derived with the MUSE pipeline. The spectrophotometric calibration of the datacubes is then checked against SDSS by extracting broadband $r$ and $i$ images from the datacubes and carrying out aperture photometry on brighter stars. Only J0255+0048 is found to require an offset in the form of a constant multiplicative factor of 1.12 applied to the MUSE data, which brings the flux scale in line with SDSS. All spectra reported in this work have been converted to vacuum wavelength and have had a barycentric correction applied. Comparison to SDSS and high-dispersion spectra of the quasars show excellent agreement.

### 4.2.3 Absorption line spectroscopy

In order to measure the absorption properties of the DLAs, spectra with higher resolution than MUSE (R$\simeq$2000 at 5500Å) are required, particularly for narrow absorption features and to resolve saturation in strong absorption lines. In this work, we use the same data originally presented in Fumagalli et al. (2014), but we refine the measurement of the H I column density by improving the determination of the quasar continuum.

The spectroscopic data available are described in Table 4.2. For all DLAs in our sample, we have moderate dispersion spectra ($R \approx 5000$) over the optical range, encompassing the DLA Ly$\alpha$ and common low ions (e.g. Si II and Fe II), which are used to estimate the DLA metallicity. For the majority of the sample, this is ESI data, while in the case of J1220+0921 the spectrum is from MagE (Jorgenson et al., 2013). Finally, J0255+0048 has additional higher dispersion data from HIRES (Prochaska et al. 2001) covering some low ions at $R \approx 50,000$, and an X-Shooter spectrum ($R \approx 6000$ López et al., 2016) with higher signal-to-noise ratio (SNR) than the ESI

| Field | Instrument | Exp. Time | S/N | Resolution | Wavelength Coverage | Reference |
|---|---|---|---|---|---|---|
| | | (s) | (per pixel) | (km s$^{-1}$) | (Å) | |
| J2351+1600 | ESI | 3600 | 12 | 37 | 3995-10140 | Fumagalli et al. (2014) |
| J0851+2332 | ESI | 3600 | 19 | 37 | 3995-10140 | Fumagalli et al. (2014) |
| J0255+0048 | ESI | 2400 | 16 | 37 | 3995-10140 | Fumagalli et al. (2014) |
| | X-Shooter | 3320-3600[a] | 30 | 40-60[a] | 3100-18000 | López et al. (2016) |
| | HIRES | 20200 | 15 | 6.3 | 5800-8155[b] | ? |
| J1220+0921 | MagE | 3000 | 22 | 71 | 3100-10360 | Jorgenson et al. (2013) |
| J0818+0720 | ESI | 3600 | 18 | 56 | 3995-10140 | Fumagalli et al. (2014) |
| J0818+2631 | ESI | 3600 | 28 | 56 | 3995-10140 | Fumagalli et al. (2014) |

[a]Differs between spectograph arms. [b]With gaps between echelle orders.

Table 4.2: Summary of the spectroscopic data used to establish the absorption properties of the DLAs.

data and coverage of the near infrared.

### H I column densities

Following the re-analysis in Fumagalli et al. (2017b) of the H I column density for DLA J0255+0048, we revise the original values presented in Fumagalli et al. (2014) for the entire sample. In this original work, we made use of a local continuum determination, which was later found to underestimate the column density by $0.1 - 0.2$ dex, due to the fact that the Ly$\alpha$ absorption lines of the DLAs often coincide with the H$\beta$ and O VI emission lines of the quasars. This overlap is a consequence of the selection in Fumagalli et al. (2014), where DLAs have similar redshift separations with respect to the quasar in order to exploit the "double-DLA" technique. In this work, we refit the $N_{\rm H\,I}$ column densities replacing the original local continuum determination by the Telfer et al. (2002) composite quasar spectrum. For each sightline, the template continuum power-law slope and normalisation are adjusted to fit the quasar continuum over the Ly$\alpha$ forest, and the H I column densities of the DLAs are then estimated by overlaying a Voigt profile on the top of the Ly$\alpha$ transition. A characteristic uncertainty of 0.1 dex is assigned to our determinations. Final values of column density are listed in Table 4.1, while a gallery of Voigt profile fits is in Fig. 4.2.

During this analysis, we noted that the composite quasar spectrum provides a poor fit to the continuum of quasar J0851+2332, which appears to be significantly lower (approximately a factor of 2) blueward of the DLA Ly$\alpha$. This feature cannot be modelled as a high redshift partial Lyman limit system, as it would require a redshift above that of the quasar. Furthermore, this break is observed in the MUSE, SDSS and ESI spectra, ruling out an artefact with the data. For this reason, in this sightline, we use a manually estimated continuum, found by fitting a spline though points in the forest believed to represent the unabsorbed continuum. Using this model we reach a column density of $\log(N_{\rm HI}/{\rm cm}^{-2}) = 21.15$. We note that, despite some degree of subjectivity in this estimate, the column density is mostly insensitive

to changes in the continuum, as its upper bound is set by the width of the core of the Ly$\alpha$ line. Indeed, if we fit the composite quasar to the spectrum between 5800 and 9000 Å, we obtain a column density of $\log(N_{\rm HI}/{\rm cm}^{-2}) = 21.25$, only marginally different from our previous estimate. To capture this discrepancy, for this DLA we adopt an uncertainty of $\pm 0.2$ dex. Finally, when fitting DLA J1220+0921, we note that the Ly$\alpha$ of the DLA at $z_{\rm DLA} = 3.090$ lies close to the Ly$\beta$ of a proximate DLA (pDLA) at $z_{\rm pDLA} = 4.1215$. For this line of sight, we therefore also include the Ly$\beta$ transition of this pDLA.

#### Metallicities

In order to calculate the metallicities of the DLAs, we adopt the metal ion column densities compiled in Fumagalli et al. (2014). These values are based on the apparent optical depth method using unsaturated transitions, or are bracketed by upper and lower limits. Table 4.1 indicates the ions used to estimate the metallicity of the DLAs in the sample. Fumagalli et al. (2017b) showed that in the case of J0255+0048 the Si II column density estimated with Voigt profile decomposition of HIRES and X-Shooter data was consistent with the value obtained with ESI and the apparent optical depth method. We therefore conclude the ionic column densities to be robust and do not re-estimate them. As the transitions used do not lie in the Ly$\alpha$ forest the continuum level can be estimated accurately from the data without the need of a quasar template as done for the H I transition. Fig. 4.3 shows strong low-ionisation absorption lines for each DLA. From the ionic column densities and the $N_{\rm H\,I}$ values measured above combined with the solar elemental abundance pattern (Asplund et al., 2009), we calculate the metallicity of the DLAs, which are listed in Table 4.1.

## 4.3 Search for Galaxy Associations

In order to identify galaxies in association to the DLAs, we have conducted a search for both Ly$\alpha$ emitters (LAEs) and Lyman break galaxies (LBGs) by searching for

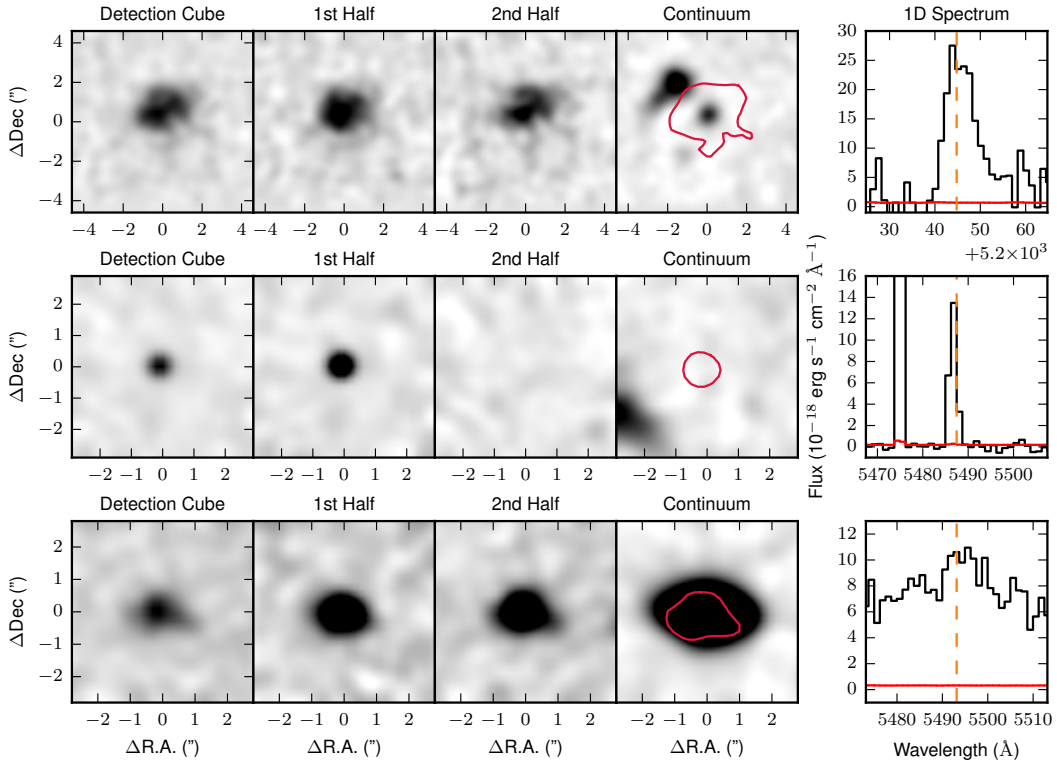


Figure 4.4: Three examples of LAE candidates that illustrate the three most common occurrences in our identification procedure. The first object (top row) is a high confidence LAE, while the later two are examples of false candidates: a cosmic ray and residual from continuum subtraction, respectively. The first three columns in each row show optimally extracted images from the detection cube after continuum subtraction, and the two independent coadds (labelled 1$^{st}$ and 2$^{nd}$ half). All three images are shown on a linear scale with the same stretch for a single candidate. The fourth column shows the reconstructed $r$-band image of the same fields with Ly$\alpha$ surface brightness contours taken from the detection map. All images are smoothed with a Gaussian kernel with full width at half maximum of 3 pixels. The right column shows 1D spectra of the detected emission line for each object (black) with the flux error (red) and the line centre (orange dashed vertical line).

emission line objects in the MUSE datacubes and fitting redshifts for continuum objects detected in deep white-light images reconstructed from the datacubes. No candidates were found in the continuum object search to a magnitude limit of $m_r < 25$ mag, thus the emission line search has been adopted as our primary method of identifying galaxy associations. The search of continuum object is nevertheless described for completeness.

### 4.3.1 Search for Ly$\alpha$ emitters

We have conducted a search for LAEs with $z_{LAE} \simeq z_{DLA}$ with a velocity window of $\pm 1000$ km s$^{-1}$ over the full MUSE field of view. Initially, the mean combined cubes are trimmed in the wavelength direction to restrict the wavelength range to that of Ly$\alpha$ over the velocity interval of interest around the DLA of each field, plus a margin of 300 MUSE channels on either side.

This first slice of the cube retains a sufficient number of channels to perform continuum subtraction of the quasar and other continuum-detected objects in the field using the utilities distributed in CubeExtractor. The $\pm 1000$ km s$^{-1}$ velocity range around the DLAs is masked during this process, to ensure that no emission-line objects are subtracted close to the DLA redshift. Because of this masking, the continuum subtraction in this region is performed by extrapolating the continuum from the unmasked region.

After this step, the resulting cubes are more finely trimmed to the wavelength range of interest, plus a margin of one channel to prevent extended objects from being truncated. This continuum-subtracted datacube then becomes the detection cube for our search. The mean, median and two independent coadded datacubes are also trimmed to the same wavelength range as the detection cube for quality control purposes, as described below. Finally, we run CubeExtractor over the detection datacubes. The first step in this process is to rescale the data variance in order to match the root-mean-square (RMS) of the flux. This rescaling is needed to correct

for small albeit significant degree of correlation in the data following the drizzling process (see Appendix A.1).

We then convolve the detection datacubes with a two pixel boxcar spatially, grouping connected pixels above a SNR threshold of 2. These groups become a detection if the following criteria are met: i) the object has at least 40 voxels above the threshold; ii) it spans at least 3 wavelength channels at a single spatial position; 3) it has an integrated SNR$\geq$ 7. Segmentation maps are further examined to ensure that connected regions resemble point-like or extended sources (i.e. they are not extremely elongated in a direction). The main source of spurious candidates are cosmic rays and sky line residuals near the edges of the IFUs, which can generate narrow spatial and spectral fluctuations. In order to filter some of these artefacts, we extract spectra using the 3D segmentation map produced by CubExtractor from the even and odd datacubes, and we retain only candidates with SNR $\geq 5$ in both coadds. This cut effectively rejects cosmic rays, which appear only in one of the two datacubes. Furthermore, we reject objects that differ significantly in SNR between the two independent coadds ($\Delta$SNR $> 3$), as these are found to be associated with noise fluctuations.

The remaining candidate LAEs are then inspected, using optimally-extracted images (see Borisova et al. 2016) from the mean, median, independent coadds and detection datacubes. During this step, we find that comparing the two independent coadds is the strongest discriminant to reject objects which appeared to differ significantly in morphology or position between the images. Finally, the 3D segmentation map is projected onto a 2D grid to extract the object spectrum over the full MUSE wavelength range. Inspection of these spectra enables us to cull other emission lines (commonly [OII]) or the residuals of continuum objects, retaining only *bona-fide* LAEs. This procedure yields 14 candidate LAEs over the six DLA fields, as summarised in Table 4.3.

Fig. 4.4 provides examples for three illustrative cases in our identification procedure. The first example is a high confidence LAE (ID85 from field J1220+0921), where

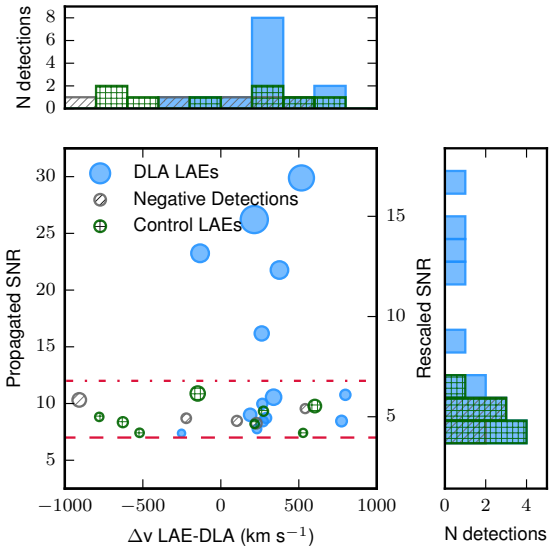


Figure 4.5: The distribution of velocity separations ($\Delta$v) between DLAs and candidate LAEs against integrated emission line signal-to-noise ratio. Real detections (blue, filled), LAEs selected in control windows (green, gridded) and negative false detections (grey, hatched) are shown. The number of voxels in the segmentation map for each candidate are indicated by the sizes of the symbols. The histogram on each axis shows the distribution of candidates over that single parameter.

the bright core of the object does not shift in the independent coadds indicating the detection is robust. The second example is a highly-significant detection which however only appears in one of the independent coadds, indicating that this candidate is a cosmic ray strike. The last example is a source that appears strong and marginally extended in the detection cube, however it is much brighter in both independent coadds once the continuum is not subtracted. This feature, combined with the properties of the $r$-band image and 1D spectrum, makes it clear that this candidate is a bright low-redshift galaxy and that the detected feature is in the residual associated with continuum subtraction.

### 4.3.2 Testing the robustness of LAE identifications

The resulting candidates from the procedure described above vary in integrated SNR from 7.4 to 29.9. While all detections are robust from a statistical point of view, it is worth examining where an additional cut in SNR is warranted to avoid

additional spurious detections that are not rejected in our procedure, especially given that the noise field is non Gaussian. To this end, we have repeated the detection procedure described above, but with the flux values in the datacubes flipped in sign (hereafter the negative datacubes) to explore whether noise fluctuations appear in our selection, and with what SNR. When searching the negative datacubes, we adopt identical parameters as in the search of real source. However, since real sources with absorption features can generate negative residuals during the continuum subtraction procedure, we remove detections overlapping with bright continuum sources. This task generates 5 detections in the negative cubes meeting our requirements. The properties of these detections are compared to the real candidates in Fig. 4.5 in terms of the velocity offset from the DLA redshift and the detection SNR. The SNR values are calculated by propitiating the error on the flux from the variance cube created by the data pipeline. Additionally figure 4.5 also shows rescaled SNR values, which is claimed to better capture the effect of correlated noise between pixels (Wisotzki, L. private communication).

Additionally, as a further test, we perform the extraction method described in Sect. 4.3 over the six datacubes, but this time shuffling the central redshifts (i.e. the DLA redshift) across the fields. This experiment yields control source catalogues containing both real LAEs, at a redshift far from the DLAs, and additional spurious detections, if any. The 8 detections produced by this search are displayed in Fig. 4.5. This method, unlike the use of the negative detection cubes, does not require symmetry in the noise properties and provides a baseline which we can use to assess if an excess of LAEs clustered to the DLAs is detected in our data.

Two key points are apparent from Fig. 4.5. First, the detections from the negative datacubes are all SNR $< 11$, while several of the the true detections are at high SNR, up to SNR $= 29.9$. Secondly, sources identified in the detection datacubes are significantly more clustered around $\Delta v \simeq +200$ km s$^{-1}$ than both the detections from the negative cubes and the control windows. As, due to radiative transfer effects, the Ly$\alpha$ line is expected to be redshifted compared to the DLA redshift by

$\approx 100 - 300$ km s$^{-1}$ (e.g Steidel et al. 2010; Rakic et al. 2011), the clustering of sources around $\Delta v \simeq +200$ km s$^{-1}$ compared to both the control sample and the detection in the negative datacubes indicates that our identification procedure yields primarily a sample of true associations.

Our MUSE programme has therefore identified associations with a very high-detection rate, possibly approaching 83%. However, the fact that a non-negligible number of detections in the negative datacubes pass our selection criteria, we take a conservative approach and establish SNR = 12 as a threshold for identifying LAEs with high purity, although at the expenses of sample completeness. In the following, we refer to these objects (SNR$>$ 12) as high-confidence confirmed LAEs, while the remaining sources form a sample of candidate associations that for most part are believed to be real, but for which we cannot exclude the presence of some spurious sources. Deep follow-up observations will be required to determine the nature of each of these sources. In rescaled SNR our 7 and 12 $\sigma$ thresholds become 4.0 and 6.7 $\sigma$ respectively. We further note that observations for DLA J0818+2631 suffer quite badly from only having 2 out of 6 exposures completed, meaning that independent coadds contain only a single exposure that presents significant gaps in the reduced datacubes due to the masking around the gaps between stacks. Thus, the search for sources in this field is likely to be incomplete.

As a last check, we investigate the robustness of our detections, in particular considering whether correlated noise affects the estimates of SNR values and the degree of incompleteness in the detected LAEs. The results of these tests are detailed in Appendix A.1, where we conclude that correlated noise is not a substantial effect in MUSE data (see also Bacon et al. 2017) and that our sample of high-confidence LAEs is highly complete.

### 4.3.3 Identification of continuum objects

For the identification of continuum detected objects, we first extract objects using the $r$-band images reconstructed from the cubes running SExtractor (Bertin & Arnouts, 1996) with minimum area 6 pixels each above an SNR of 2.0. The SNR threshold was raised for DLA J0818+2631, which had many resuduals due to being only partially completed. The resulting segmentation maps are used to define apertures on the datacubes over which we extract 1D and 2D spectra for the selected objects. In this work, we attempt to determine redshifts only for objects with $m_{\rm r} <= 25.0$ mag (corrected for Galactic extinction). This choice is motivated by previous MUSE analyses (Fumagalli et al., 2015) which have shown how high confidence redshifts for objects fainter than this approximate limit are only obtained in presence of bright emission lines (usually Ly$\alpha$). As we already search for Ly$\alpha$ emission close to the DLAs, LBGs associated to DLAs will not be missed.

These spectra are then inspected for emission and absorption lines, as well as characteristic continuum features. Their redshifts are measured by two authors (RM and DJH), either by fitting Gaussian functions to the detected emission lines, or by comparing the 1D spectra with a range of stellar and galaxy templates, including low redshift galaxies and high redshift LBGs. The quality of redshifts obtained were ranked from 0 to 4, with $3-4$ indicating a reliable redshift. Stars were graded 6. As no $m_{\rm r} <= 25.0$ objects were found to lie close to a DLA, we do not include further details on the redshifts.

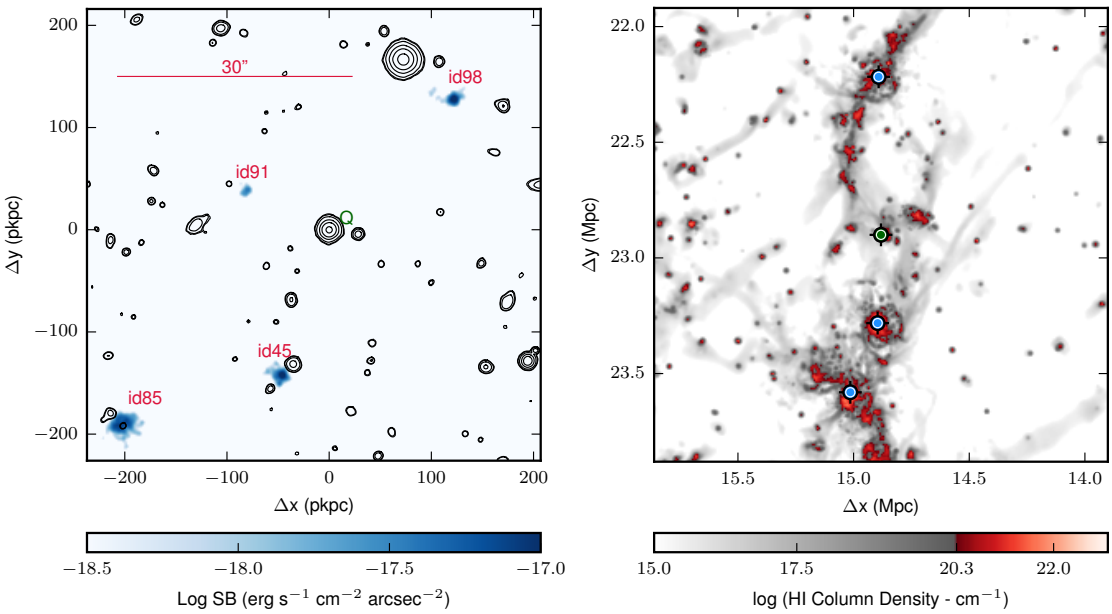


Figure 4.6: **Left:** A map of the MUSE field around the DLA J1220+0921 sightline, showing an optimally extracted Ly$\alpha$ surface brightness map of all four objects detected around the DLA. The pixels inside the segmentation map for each detection are projected onto an image. The angular scale is shown with the quasar position and the LAE ids (red). $r$-band contours are shown in black. **Right:** An analogue of the J1220+0921 system extracted from the EAGLE simulation for illustrative purposes. The H I column density projected though the 25 Mpc simulation is shown in greyscale, with the transition to the red colour-map indicating DLA column density. The FoV matches that of the MUSE DLA observations in co-moving distance. Also marked are LAEs (blue) and the central DLA (green).

## 4.4 Properties of the high-confidence associations

### 4.4.1 Notes on individual fields

In Sect. 4.2, we have described the reduction and data analysis of MUSE observations collected in six quasar fields hosting DLAs at $z > 3$, drawn from the parent sample described in Fumagalli et al. (2014). Following the search of LAEs and redshifting of continuum-detected sources, we identify five high-confidence LAE associations (with sufficient high SNR to be at very high purity) in three out of six DLA field (J0851+2332, J1220+0921 and J0255+0048, which was previously published in Fumagalli et al. 2017b). The derived properties of the detected objects are summarised in Table 4.3, with both Ly$\alpha$ and $r$-band images shown in Fig. 4.10. Spectra of the Ly$\alpha$ lines are also shown in Fig. 4.11. In addition to the high-confidence associations, we further identify 9 LAE candidates across five fields, which are shown in Fig. A.4. These are detections at lower SNR and likely include a high fraction of true associations, but for which we cannot guarantee the sample purity. In the following, we provide a brief discussion of the key features of the high-confidence associations.

#### J0851+2332

A Ly$\alpha$-bright LBG was detected at +260±20 km s$^{-1}$ from the DLA redshift with an impact parameter of 25±2 kpc, as summarised in Table 3. Fig. 4.10 shows the continuum and Ly$\alpha$ detection of this object. With an $r$-band magnitude of $25.34 \pm 0.14$ mag this galaxy is forming stars at a rate[1] of $5.6 \pm 0.7$ M$_\odot$yr$^{-1}$. DLA J0851+2332 is a somewhat high-metallicity DLA, for which LBGs associations have been previously detected (e.g. Møller et al. 2002). As the velocity offset is estimated

[1]The SFR we quote here is higher than reported in Fumagalli et al. (2017b), as we measure $r$-band magnitudes from the MUSE data and not from the Keck LRIS imaging.

| Id | RA [a] (hrs) | Dec [a] (deg) | $b$ [a] (arcsec) | Ly$\alpha$ Flux [b] ($10^{-18}$ erg s$^{-1}$ cm$^{-2}$) | $m_r$ [c] (AB mag) | $L_{Ly\alpha}$ [b] ($10^{41}$ erg s$^{-1}$) | $\Delta v$ [a] (kms$^{-1}$) | $D_{LOS}$ [a] (kpc) | $SFR_{UV}$ [d] ($M_\odot$ yr$^{-1}$) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | J2351+1600 | | | | | |
| 54 | 23:51:51.6 | 16:00:56 | 18.2±0.3 | 1.27±0.17 | >28.40 | 1.80±0.24 | -252±16 | 132.4±2.4 | <0.37 |
| 86 | 23:51:53.7 | 16:00:58 | 17.2±0.8 | 2.60±0.29 | >28.09 | 3.68±0.41 | +188±16 | 124.9±5.5 | <0.49 |
| | | | | J0851+2332 | | | | | |
| 83* | 08:51:43.6 | 23:32:12 | 3.3±0.3 | 9.98±0.62 | 25.34±0.14 | 11.94±0.74 | +262±17 | 24.9±2.1 | 5.55±0.72 |
| 95 | 08:51:42.0 | 23:32:26 | 29.6±0.5 | 2.77±0.33 | >27.68 | 3.32±0.39 | +270±17 | 221.0±4.0 | <0.64 |
| | | | | J0255+0048 | | | | | |
| 56* | 02:55:18.8 | 00:48:49 | 4.0±0.5 | 30.99±1.18 | 24.58±0.22 | 30.53±1.16 | +215±18 | 30.9±3.6 | 9.81±1.96 |
| 78 | 02:55:16.7 | 00:49:06 | 33.7±1.6 | 4.72±0.46 | >27.46 | 4.65±0.45 | +343±18 | 259.0±12.1 | <0.69 |
| 108 | 02:55:18.7 | 00:48:57 | 10.9±0.6 | 3.18±0.38 | >27.36 | 3.13±0.37 | +774±18 | 83.8±4.5 | <0.76 |
| | | | | J1220+0921 | | | | | |
| 45* | 12:20:21.8 | 09:21:17 | 19.6±0.2 | 14.62±0.63 | >26.77 | 14.97±0.64 | -279±18 | 150.1±1.9 | <1.34 |
| 85* | 12:20:23.1 | 09:21:10 | 36.2±0.2 | 27.45±0.92 | 25.87±0.20 | 28.12±0.94 | +370±18 | 277.1±1.9 | 3.08±0.58 |
| 91 | 12:20:22.1 | 09:21:40 | 11.7±0.3 | 3.03±0.30 | >27.57 | 3.10±0.31 | +120±18 | 89.3±2.0 | <0.64 |
| 98* | 12:20:20.3 | 09:21:52 | 22.9±0.3 | 13.39±0.62 | >26.81 | 13.72±0.63 | +229±18 | 175.0±2.1 | <1.29 |
| | | | | J0818+2631 | | | | | |
| 87 | 08:18:14.5 | 26:31:47 | 22.9±1.0 | 3.24±0.42 | >26.98 | 3.96±0.51 | +232±17 | 170.7±7.4 | <1.24 |
| 91 | 08:18:13.5 | 26:31:33 | 7.1±2.0 | 4.87±0.56 | >27.11 | 5.96±0.68 | +291±17 | 52.9±14.5 | <1.10 |
| 119 | 08:18:11.3 | 26:31:18 | 29.1±0.7 | 4.79±0.44 | >27.23 | 5.86±0.54 | +799±17 | 216.9±5.2 | <0.99 |

Table 4.3: Full list of and DLA associations. * High confidence detections. [a] Position and velocities estimated using the Ly$\alpha$ centroids. [b] Ly$\alpha$ integrated flux measured from integrating inside the 3D segmentation map. [c] Upper limits use the apertures defined from the Ly$\alpha$ segmentation map, and specifically pixels where the map includes at least two wavelength channels. [d] Values derived using the SFR calibration from Fumagalli et al. (2010).

using the Ly$\alpha$ emission line which is commonly shifted from the systematic redshift with respect to the true velocity, it is quite plausible that the LBG and DLA are close in velocity space. Based on the clustering of LAEs at $z \simeq 3$ (Gawiser et al., 2007), the impact parameter of 25±2 kpc is on the order of the virial radius for the mean LAE halo masses. While there will be a large scatter in the halo mass of a single LAE, this does suggest that the DLA may be directly related to a fainter, undetected galaxy, rather than this LBG itself. Hydrodynamical simulations predict that DLAs typically have impact parameters of $\leq 0.1\mathrm{R}_{200}$ (Rahmati & Schaye, 2014).

**J0255+0048**

The detected galaxy near DLA J0255+0048 was discussed at length in Fumagalli et al. (2017b), its detection is summarised in Fig. 4.10. The extended Ly$\alpha$ structure spans $37 \pm 1$ kpc along its major axis and is dominated by two clumps. While most of the source has no broadband counterpart, a compact continuum source is embedded towards the edge of this structure. Although the MUSE spectrum of this continuum source is noisy, this source has spectrophotometry consistent with an LBG at $z_{LBG} \simeq z_{DLA}$ (see Appendix A.3). Thus, given its location between the Ly$\alpha$ structure and the DLA at the same redshift, it is quite likely that the LBG forms part of the same structure.

As shown in Fumagalli et al. (2017b) the double peak of the Ly$\alpha$ emission line for this object stems from a velocity offset between the two clumps, with a separation of $\Delta v = 140 \pm 20$ km s$^{-1}$. This velocity difference is consistent with the velocity offset of the two components seen in the DLA absorption lines, such as Si II shown in Fig. 4.3 ($\Delta v = 55 \pm 6$ km s$^{-1}$). It was argued that the morphology of this source and the correspondence between the two components in absorption and Ly$\alpha$ emission may hint that this system is a merger, which has triggered starbursts in two galaxies embedded in the clumps. This symmetry is strong evidence that the detected structure is physically linked to the DLA, i.e. a host. Alternatively the clumps could be part of some extended collapsing proto-disk, although the scale of

this system is difficult to reconcile with this picture. Scheduled cycle 25 WFC3/IR observations (PID: 15283, PI Mackenzie) will soon offer more details on the nature of this system.

**J1220+0921**

Three high-confidence LAEs are detected in the field of DLA J1220+0921, and all three lie within ±400 km s$^{-1}$ of the DLA redshift. With impact parameters between 150.4 to 278.2 kpc, these associations are unlikely to be the galaxies that give rise to the DLA system, but they trace the large scale structure in which DLA J1220+0921 is embedded in. Additionally, a lower significance LAE is detected in this field, id91. This candidate is much closer to the line of sight than the high-confidence detections at 89±2 kpc, with a velocity offset from the absorber of +120±20 km s$^{-1}$, and may be more closely associated with the DLA. With a metallicity of $Z/Z_{\odot} = -2.33$, DLA J1220+0921 is the most metal-poor DLA in our sample, and our MUSE observations reveal for the first time associations with a truly metal-poor DLA at high redshift.

Galaxy id85 is the brightest detection both in Ly$\alpha$ and $r$ band. Fig. 4.10 (b) shows the Ly$\alpha$ halo of id85 extends far beyond the UV continuum. This object has an impact parameter of 278.2 kpc (36.4 arcsec), it is offset from the absorption system by +370±20 km s$^{-1}$, and is forming stars at a rate of $3.1 \pm 0.6$ M$_{\odot}$yr$^{-1}$ based on the UV luminosity. The rest-frame UV spectrum of this object is consistent with an LBG with $z_{LBG} \simeq z_{DLA}$=3.309 (see Fig. A.3 lower panel), the Lyman break is convincingly detected but the spectrum is too noisy to detect the interstellar lines. The broadband spectrum combined with the Ly$\alpha$ halo of this object make its redshift unmistakable. The other two detections, instead, do not show a continuum counterpart at the depth of these observations.

Fig. 4.6 shows the distribution of the detected galaxies near the DLA redshift across the MUSE field of view, spanning approximately 50 arcsec diagonally across the image. Three LAEs and the DLA, at the position of the quasar, are roughly

joined by a line, suggestive of the fact that these objects trace a filament within which the DLA is embedded A similar configuration was detected with MUSE for a very metal-poor (pristine) Lyman Limit System, with multiple LAEs detected in filament-like arrangement around the absorber (Fumagalli et al., 2016). With J1220+0921 the spread in velocity space is smaller by a factor of $\sim 2.5$ and the Ly$\alpha$ luminosities are higher. To gain more insight into the nature of this system, and in particular what type of structure we might be observing, we search for analogues in the EAGLE simulations, and in particular the 25 Mpc box with DLAs identified using URCHIN. The details of the modelling of LAEs and DLAs are described in Sect. 4.5. 4.6 (right) shows a projected H I column density map over the same field of view as the one probed by MUSE at $z = 3.25$. The figure is centred on a DLA pixel, selected because of its similarity to J1220+0921. Specifically, we searched for DLAs with 3 LAEs within the area defined by the MUSE FoV, but none within the inner $15 \times 15$ arcsec$^2$. In this search, we require that LAEs have luminosity $L_{Ly\alpha} \geq 10^{42}$ erg s$^{-1}$ sub-sampled to match the luminosity function as described in Sect. 4.5.1. Out of all the matches, the example shown in Fig. 4.6 is selected due to its morphological similarity with the field J1220+0921. In this case, the DLA arises from a small galaxy at close impact parameter, which is in turn embedded in a filamentary structure hosting additional Ly$\alpha$ bright galaxies. A wider view of the selected region further reveals that this whole structure is just part of a filament extending beyond the scale probed by of MUSE-like observations.

While analogues to this system in the EAGLE simulation support the idea of a filamentary structure, we note however that this picture is complicated by the large velocity offset between id85 and id45 of $650 \pm 25$ km s$^{-1}$, which may be too large to be explained with the associations embedded in a single filament. Therefore, we cannot exclude that galaxies are instead embedded in a proto-group or cluster type environment, but not yet bound to the same halo.

### 4.4.2 Continuum counterparts of the LAEs

As Fig. 4.10 shows, three of the five LAEs detected have an $r$-band counterpart visible at the depth of our MUSE observations. We have extracted spectra for these objects, which we present in Fig. A.3. With integral field spectroscopy, we can also examine more closely the nature of the candidate DLA hosts identified in Fumagalli et al. (2015) based on impact parameters in deep $u-$band images. None of the sources are confirmed to be close to the DLA redshift according to MUSE data, with most of them being in fact low-redshift interlopers (see appendix A.2). This highlights the power of searching for DLA host galaxies with Ly$\alpha$ rather than relying on broadband detections, and emphasises once more the perils of relying on proximity to the quasar sightline as way to identify candidates host of the DLAs.

## 4.5 Description of models and simulations

In the following, we compare our observational results to simulations and semi-analytic models in order to understand better the constraints they put on the association between DLAs and galaxies. In this section, we provide a detailed description on how different models are produced and analysed.

The hydrodynamic simulations adopted on this work are taken from the EAGLE suite (Schaye et al., 2015), with snapshots post-processed using the URCHIN reverse ray-tracing code (Altay & Theuns, 2013) to identify DLAs. In our analysis, we use the processed EAGLE snapshots combined with simple prescriptions to populate halos with LAEs to produce mock observations of the correlation between DLAs and galaxies. We additionally use a mock catalogue based on GALICS semi-analytic model designed to produce realistic LAEs (Garel et al., 2015). For this model, DLAs are "painted" onto dark matter halos. This simple prescription allows us to quickly investigate the effects of varying the DLA cross section as a function of halo mass. Comparisons with the data then allow us to judge to what extent current or future

data can distinguish between various models. This simple model of "painting" DLAs onto halos is also applied to the EAGLE simulations, to further gain insight into the properties of DLAs identified with URCHIN.

### 4.5.1 The EAGLE simulations

Evolution and Assembly of GaLaxies and their Environments (EAGLE, Schaye et al. 2015) is a suite of cosmological hydrodynamical simulations performed used the GADGET-3 Tree/SPH code (Springel, 2005) with modifications to the hydrodynamics solver described by Schaller et al. (2015). The simulations incorporate the dominant cooling and heating processes of gas in the presence of the uniform but time-varying UV/X-ray background of Haardt & Madau (2001) as described by Wiersma et al. (2009). Physics below the resolution scale, such as star and black hole formation and their feedback effects, are incorporated as 'subgrid modules' with parameters calibrated against observations at redshift $z = 0$ of the galaxy stellar mass function, the relation between galaxy mass and size, and the relation between galaxy mass and black hole mass (Crain et al., 2015). The simulations reproduce a number of observations that were not part of the calibration, including the colour-magnitude diagram (Trayford et al., 2017), the small-scale clustering of galaxies at $z = 0$ (Artale et al., 2017), the evolution of the galaxy stellar mass function (Furlong et al., 2015) and of galaxy sizes (Furlong et al., 2017), and the evolution of the H I and $H_2$ contents of galaxies (Crain et al., 2017; Lagos et al., 2015).

The EAGLE simulations are performed in cubic periodic volumes, and the linear extent $L$ of the simulation volume and number of simulation particles is varied to allow for numerical convergence tests. In this work, we use simulations L0100N1504 and L0025N0752 from table 1 of Schaye et al. (2015). Briefly these have $L = 100$ co-moving megaparsecs (cMpc) and $L = 25$ cMpc, and a SPH particle masses of $1.8 \times 10^6 \mathrm{M}_\odot$ and $0.2 \times 10^6 \mathrm{M}_\odot$, respectively; the Plummer equivalent co-moving gravitational softening is 2.66 and 1.33 kpc, limited to a maximum physical softening of

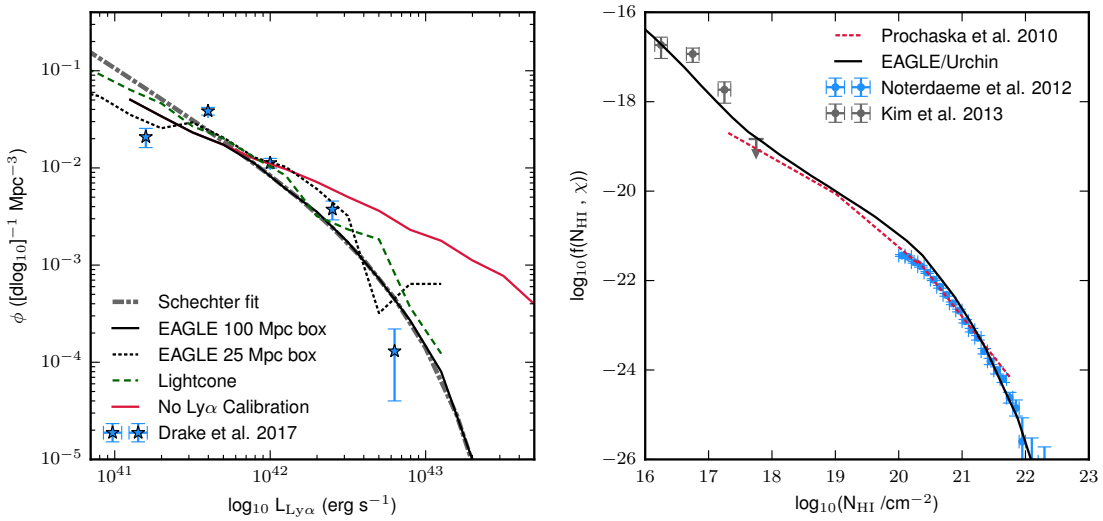


Figure 4.7: *Left:* The Ly$\alpha$ luminosity function from the EAGLE LAE sample compared against recent observations (Drake et al., 2017, blue stars) and a Schechter function fit to these observations (gray dot-dashed). The calibrated (black solid line) and original (red solid line) luminosity functions for the 100 Mpc box simulation are shown, together with the rescaled luminosity function from the 25 Mpc EAGLE box (dotted line). The luminosity function from GALICS is plotted (green, dashed), it shows excellent agreement with the data without need for calibration. *Right:* The column density distribution function derived from the EAGLE simulation after post-processing with the URCHIN radiative transfer code, compared against observations at comparable redshift.

0.7 and 0.35 kpc, respectively. The simulations assume the cosmological parameters of Planck Collaboration et al. (2014c), but the minor differences with the Planck Collaboration et al. (2016b) parameters adopted in the current paper are unlikely to be important.

To compare to the data presented earlier, we need to identify both DLAs and LAEs in the EAGLE. Since neither neutral hydrogen nor Ly$\alpha$ radiative transfer are directly incorporated in EAGLE, we compute these quantities in post-processing as explained next.

### Identifying DLAs in the EAGLE simulations

The ionising background in EAGLE is implemented in the optically-thin limit. Self-shielding of gas, allowing for the appearance of DLAs, is computed in post-processing

using the URCHIN radiative transfer code described by Altay & Theuns (2013). Briefly, this algorithm allows each gas particle to estimate the local ionising intensity it is subject to by sampling the radiation field in 12 directions, with neutral gas in neighbouring gas particles potentially decreasing the local photo-ionisation rate below the Haardt & Madau (2001) optically-thin value. Assuming the neutral fraction of a particle is set by the balance between photo- and collisional ionisation versus recombinations, a reduced ionisation rate increases the particle's neutral fraction, which in turn affects the ionisation rate determined by the particle's neighbours. The impact of a change in the neutral fraction on the photo-ionisation rate and vice versa is iterated until the neutral fraction of each particle converges from one iteration to the next. This post-process step thus yields the neutral hydrogen fraction $x_{\rm H\,I} \equiv n_{\rm H\,I}/n_{\rm H}$ of each SPH particle.

The resulting H I volume density is then projected onto a $8192^2$ grid along the coordinate $z$-axis of the simulation box, using the Gaussian smoothing described by Altay & Theuns (2013). Applying a column density threshold of $\log(N_{\rm H\,I}/{\rm cm}^{-2}) = 20.3$ allows us to identify DLAs[2]. We also calculate the $N_{\rm H\,I}$-weighted $z$-coordinate of particles along each DLA line of sight to obtain the 3D position of each DLA (two spatial coordinates $x$ and $y$, and a redshift coordinate $z$). The redshift of each DLA allows us to compare DLAs to galaxies in redshift space, although we note that observed DLA redshifts are derived from low-ionisation metal lines rather than H I directly, yet thee elements are believed to trace the same phase of the gas.

To compare to the MUSE observations, we analyse the L0025N0752 EAGLE snapshot at $z = 3.027$ which is closest in redshift to the data. Similarly to the results of Altay et al. (2011) based on the OWLS simulations (Schaye et al., 2010), the column density distribution function (CDDF, the number density of absorbers per unit column density, per unit absorption distance, $f(N_{\rm H\,I})$) of the post-processed snapshots using URCHIN is in good agreement with observations (Fig. 4.7,). The

[2]The redshift path through the simulation box is so small that the contribution of chance alignments of two high-column density systems to the DLA cross section is negligible.

data in this figure combines the low column density data ($\log(N_{\rm HI}/{\rm cm}^{-2}) < 18$) compiled by Kim et al., 2013 ($z = 2.4 - 3.2$), the multiple power-law fit to the Lyman-limit and DLA column density range derived by Prochaska et al. (2010) at $z = 3.7$, and the sub-DLA and DLA data from Noterdaeme et al. (2012) ($\langle z \rangle = 2.5$). At these redshifts, we find that the simulated and observed incidence of DLAs is very similar.

### Identifying LAEs in EAGLE

Gravitationally bound substructures in EAGLE are identified combining the friends-of-friends (FOF), (Davis et al., 1985) and SUBFIND (Springel, 2005; Dolag et al., 2009) algorithms. Physical properties of these 'galaxies' such as their centre of mass position and velocity, stellar mass and star formation rate, are computed and stored in a database (McAlpine et al., 2016). We populate the simulation with LAE by selecting simulated galaxies by star-formation rate, SFR, and associating a Ly$\alpha$ luminosity, $L_{Ly\alpha}$ using the conversion from Eq. 4.5.1 (Furlanetto et al., 2005).

$$\mathrm{L_{Ly\alpha}} = \frac{\mathrm{SFR}}{\mathrm{M_\odot yr^{-1}}} \times 10^{42}\ \mathrm{erg\ s^{-1}} \tag{4.5.1}$$

This prescription neglects diffuse emission from the low density intergalactic and circumgalactic medium, but captures the bulk of the emission associated with star formation inferred from stellar population synthesis models under the assumption that $\approx 2/3$ of the recombinations occurring in H II regions produces a Ly$\alpha$ photon. This is a reasonable approximation for the rate at which star formation in the galaxy produces Ly$\alpha$ photons, yet may nevertheless yield a large overestimate of the Ly$\alpha$ luminosity of a galaxy because a significant fraction of such photons are absorbed by dust.

The Ly$\alpha$ luminosity function of LAEs is shown in the left panel of Fig. 4.7, comparing results from L0025N0752 and L0100N1504 with the Schechter fit to the observed luminosity function from Drake et al. (2017), using the redshift bin $2.92 \leq z < 4.00$.

As anticipated, by neglecting dust absorption EAGLE significantly over predicts the observed number density of emitters at $L_{Ly\alpha} > 5 \times 10^{41}$erg s$^{-1}$, especially at high SFRs, $\dot{M}_\star \geq 10 M_\odot$ yr$^{-1}$. This trend is well documented observationally: star formation rates inferred from the UV compared to those based on Eq. (4.5.1) are discrepant if no correction is made for dust (e.g. Dijkstra & Westra, 2010). Moreover, Steidel et al. (2011) show that roughly half of the LBG population is not detected in Ly$\alpha$ at all.

We therefore need to make a dust correction to the simulated Ly$\alpha$-fluxes. We do so by sub-sampling the simulated star-forming galaxies until they match the Schechter fit to the luminosity function from Drake et al. (2017). This is done by dividing the Drake et al. Schechter fit by the measured EAGLE luminosity function, and applying this ratio as probability that a galaxy of a given luminosity will be added to the LAE catalogue. The result of this re-scaling is shown in Fig. 4.7, revealing that the sub-sampling performs well for $L_{Ly\alpha} > 5 \times 10^{41}$erg s$^{-1}$. Given that the high-confidence MUSE sample extend only down to $10^{42}$erg s$^{-1}$, the EAGLE model appears excellent for comparison to the observations.

As consistency check, we compute the clustering of sub-sampled LAEs identified in the L0100N1504 EAGLE simulation (Fig. 4.8). Ideally our identified LAEs should match observational measurements of the clustering amplitude, as typically quantified by fitting a simple power-law model to the correlation function,

$$\xi(r) = (r/r_0)^\gamma \,. \tag{4.5.2}$$

Observational estimates of $r_0$ for LAEs at this redshift range between 2-4 Mpc (Diener et al., 2017), with $\gamma = -1.8$ typically assumed (e.g. Bielby et al. 2017; Gawiser et al. 2007). Ly$\alpha$ luminosity typically varies from study to study, and so the samples are not necessarily the same. As discussed in section 4.3.2, our high-confidence sources with SNR>12 have luminosity $L_{Ly\alpha} > 10^{42}$ erg s$^{-1}$. Based on this selection, we draw comparison to the study of Gawiser et al. (2007), where a comparable luminosity

limit was adopted yielding a value of $r_0 = 3.6^{+0.8}_{-1.0}$ Mpc (shown in figure). Fig. 4.8 shows that the clustering of our selected LAEs in EAGLE is higher than $r_0 = 3.6$ Mpc, favouring instead a value of $r_0 = 5.1$ Mpc (between $1 - 15$ Mpc). Although there is large scatter in the estimates of $r_0$, our simulated LAEs have a clustering amplitude which is too large. Cross-correlations with DLAs will be less sensitive to this offset as the autocorrelation scales with the bias squared, the cross-correlation however is related to the product of the two biases. In comparison to observational uncertainties in the following work we consider this discrepancy to be minor, future datasets will require LAE populations which better match observations.

Altogether, these comparisons show how our simple model for the LAEs in EAGLE is able to capture all the basic properties (luminosity function) necessary to model the small-scale clustering of LAEs in the observations. In the following section, we further show how an independent semi-analytic model constructed to reproduce the clustering and luminosity function of LAEs agrees well with the prediction derived from the EAGLE simulations, validating further our model.

### 4.5.2 The GALICS semi-analytic model

We use mock catalogues of LAEs based on the model of Garel et al. (2015) which combines the GALICS semi-analytic model with numerical simulations of Ly$\alpha$ radiation transfer in galactic outflows (Verhamme et al., 2006; Schaerer et al., 2011) to predict the observed Ly$\alpha$ luminosities of galaxies. As shown in Garel et al. (2015, 2016), the model can reproduce various statistical constraints on galaxies at high redshift, such as the abundances of LAEs at $3 < z < 6$. The mock lightcones used in this study were extracted from the GALICS cosmological simulation volume ($L_{box} = 100h^{-1}$ comoving Mpc) and were specifically designed to match the redshift range, geometry and depth of typical Ly$\alpha$ surveys with MUSE (see Garel et al. 2016 for more details about GALICS and the mocks). We refer to these mock catalogues hereafter as GALICS, including the additional radiative transfer.

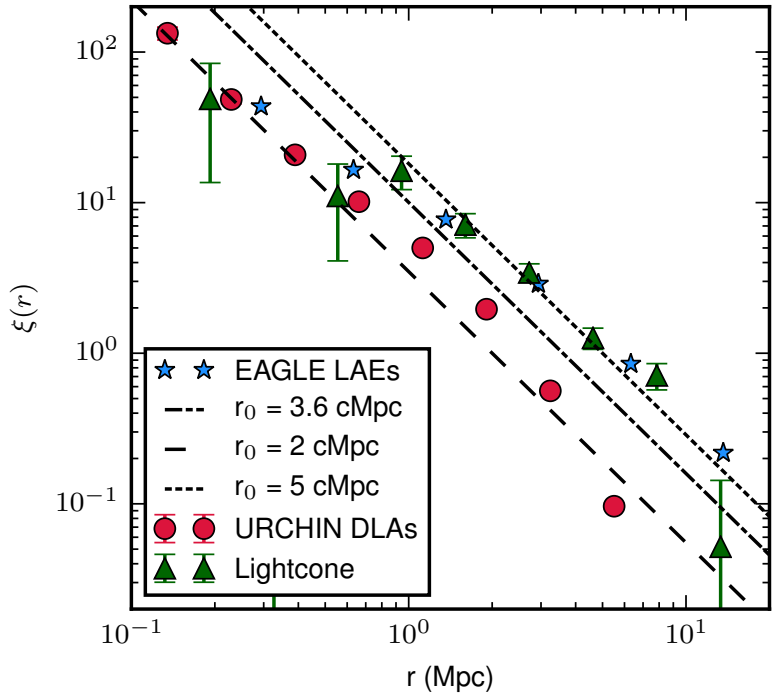


Figure 4.8: The two-point correlation function of LAEs ($L_{Ly\alpha} \geq 10^{42}$ erg s$^{-1}$, *blue stars:* sub-sampled LAEs identified in the L0100N1504 EAGLE simulation, *green triangles*: LAEs from the GALICS SAM) and of DLAs identified in the L0025N0768 EAGLE model (*red circles*). Three example power-law functions, $\xi(r) = (r/r_0)^\gamma$, with $\gamma = -1.8$ are shown for different values of $r_0$ as per the legend.

Fig. 4.7 (left) shows the predicted Ly$\alpha$ luminosity function from GALICS, compared to MUSE deep field observations (Drake et al., 2017). Fig. 4.8 also demonstrates that the clustering of the LAEs in the lightcone is in close agreement with that of EAGLE. Thus, the GALICS mock catalogue represents an excellent way to cross check predictions from EAGLE, and further validate our selection of LAE-like galaxies from the simulation. The LAE mock, however, does not simulate neutral hydrogen, and we describe next a simple model which can be applied to both GALICS and EAGLE to populate dark-matter halos with DLAs.

### 4.5.3 A halo prescription for DLAs

As well as exploiting the hydrodynamics of the EAGLE simulations to predict the position and properties of DLAs, we employ a second model by assigning DLAs

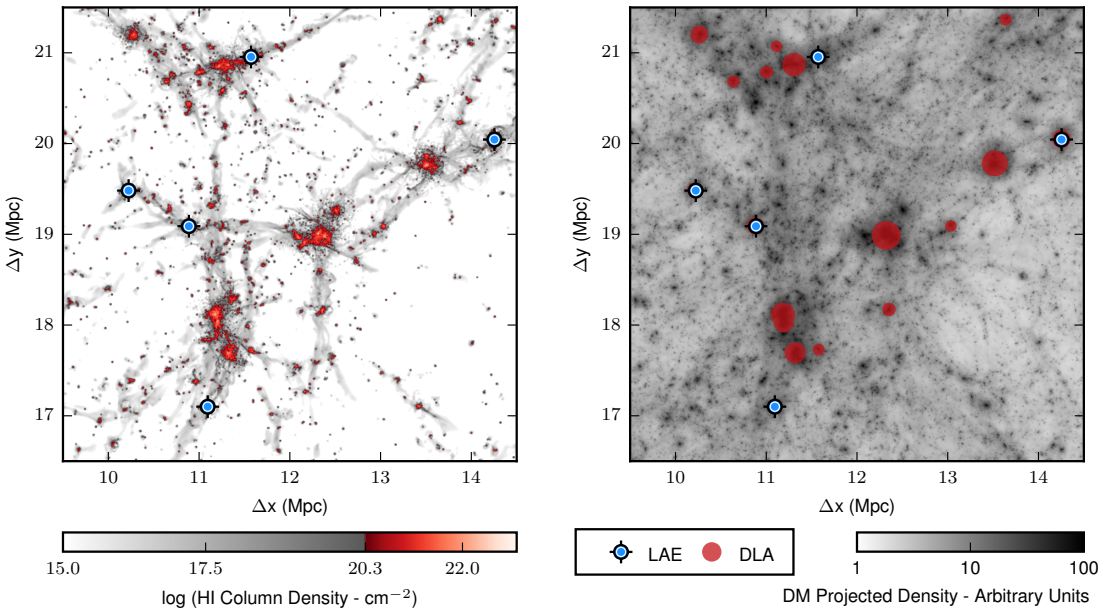


Figure 4.9: A comparison of the two DLA modelling techniques utilised in this paper. Each panel shows the distribution of DLAs and LAEs drawn from the same EAGLE 25 Mpc simulation, shown on the same physical scale. Both plots are projected over the full 25 Mpc box length. **Left:** The DLAs have been computed from the hydrodynamics of the EAGLE simulation using the URCHIN post-processing. The colour-map corresponds to $N_{\rm H\,I}$ with sightlines above the DLA column density threshold ($\log(N_{\rm H\,I}/\ {\rm cm}^{-2}) \geq 20.3$) highlighted in red. **Right:** A demonstration of the halo painting scheme used to populate the EAGLE simulation with DLAs based on the halo catalogue. Each halo above the mass threshold is painted with a circular cross-section that is proportional to its halo mass. The projected dark matter density is plotted for reference in greyscale, in arbitrary units.

to the dark matter halos from the simulations via a simple halo “painting" model. This is similar to the model put forward in Font-Ribera et al. (2012), and updated in Pérez-Ràfols et al., 2018b, in which DLA cross sections are assigned to halos as a function of halo mass. With this model, it is possible to quickly adjust the parameters to match observations, such as the large scale clustering of DLAs as measured in BOSS.

In this model, the relation between DLA cross section and halo mass is described by Eq. 4.5.3, where $\Sigma(\mathrm{M_h})$ is the DLA cross-section for a halo of mass $\mathrm{M_h}$ above some minimum halo mass $\mathrm{M_{min}}$ with a power-law slope of $\alpha$ and a zero-point $\Sigma_0$.

$$\Sigma(\mathrm{M_h}) = \Sigma_0(\mathrm{M_h}/10^{10}\mathrm{M_\odot})^\alpha \qquad (\mathrm{M_h} > \mathrm{M_{min}}) \tag{4.5.3}$$

The halo catalogues used in this work are obtained from the EAGLE database (McAlpine et al., 2016) Friends-of-Friends table for the 25 Mpc and 100 Mpc EAGLE boxes. While the 100 Mpc box suffers from resolution effects at low halo masses ($M_{200} \simeq 10^9\mathrm{M_\odot}$), the higher-resolution 25 Mpc box suffers from a limited volume that contains few massive halos (above $M_{200} = 10^{13}\mathrm{M_\odot}$). For these reasons, we combine both the 25 and 100 Mpc simulations, whereby the higher resolution simulation allows us to study halo model parameters which extend the cross-sections to low masses, while the larger box provides better convergence at high halo masses and on larger scales (see also Pontzen et al. 2008).

To populate halos with the cross sections specified in Eq. 4.5.3, we generate a $8192^2$ grid along the $z$-axis of the simulation with circular kernels representing DLAs centred on halos, the size of which varies with halo mass. While values of $\mathrm{M_{min}}$ and $\alpha$ are fixed to those from Font-Ribera et al. (2012) to reproduce the observed large scale clustering of DLAs, $\Sigma_0$ is fit for each value of $\mathrm{M_{min}}$ and $\alpha$ such that the umber of DLAs per unit path length ($\ell_{DLA}(\mathrm{X})$) in the 100 Mpc simulation matches $\ell_{DLA}(\mathrm{X})$ of the URCHIN DLAs in the 25 Mpc box. Hence, we calibrate the cross section to $\ell_{DLA}(\mathrm{X}) = 0.0948$, which is in good agreement with observational estimates at $z \simeq 3.5$ (e.g. Sánchez-Ramírez et al., 2016). We then transfer the same

calibration onto the other simulations. In order to obtain 3D coordinates for the DLAs, we use the location of the DLA parent halo, or take an average in the case where DLAs overlap. The periodic boundary of the box is taken into account during the projection.

A powerful feature of this model is the possibility to quickly explore how different parameters impact the small-scale clustering of galaxies around DLAs, which is the quantity probed by our observations. For this reason, we implement different values for the model parameters, as summarised in Table 4.4. While for this model we cannot generate predictions for the CDDF, we can still predict the number density of DLAs, as listed in Table 4.4. As this simple scheme is independent of the hydrodynamics of a simulation, we have also applied it to the GALICS mock catalogue described in Sect. 4.5.2.

Fig. 4.9 shows the halo painting model applied to a section of the EAGLE 25 Mpc simulation for the $\alpha = 0.75$ model (right) alongside the $N_{\rm H\,I}$ column density map from the EAGLE/URCHIN post-processing. Qualitatively, this choice of parameters produces a covering factors of DLAs which closely resembles the result of the simulation, with the difference of a sharp cut-off at a fixed minimum mass, which does not apply to the EAGLE simulations.

Moreover, it is also apparent in Fig. 4.9 (left) that the EAGLE simulations contain small clumps of H I, some of which reach DLA column density. Being close to the resolution limit of the simulations, we cannot assess whether the H I properties of these halos are fully converged and physically meaningful. Although small in cross-section, the clumps are numerous, and often far from massive galaxies. It may be these clumps which suppress the clustering of the URCHIN DLAs with respect to the results from the halo-painting model.

| $\alpha$ | $M_{min}$ ($M_\odot$ ) | $M_{max}$ ($M_\odot$ ) | $\Sigma_0$ (Mpc$^2$) |
|---|---|---|---|
| 1.1 | $3.0\times10^{9}$ | $1.5\times10^{13}$ | 0.000678 |
| 0.75 | $5.9\times10^{10}$ | $1.5\times10^{13}$ | 0.00327 |
| 0.0 | $2.2\times10^{11}$ | $1.5\times10^{13}$ | 0.123 |

Table 4.4: Different parameters of the DLA halo model by Font-Ribera et al. (2012) which we adopt in this work.

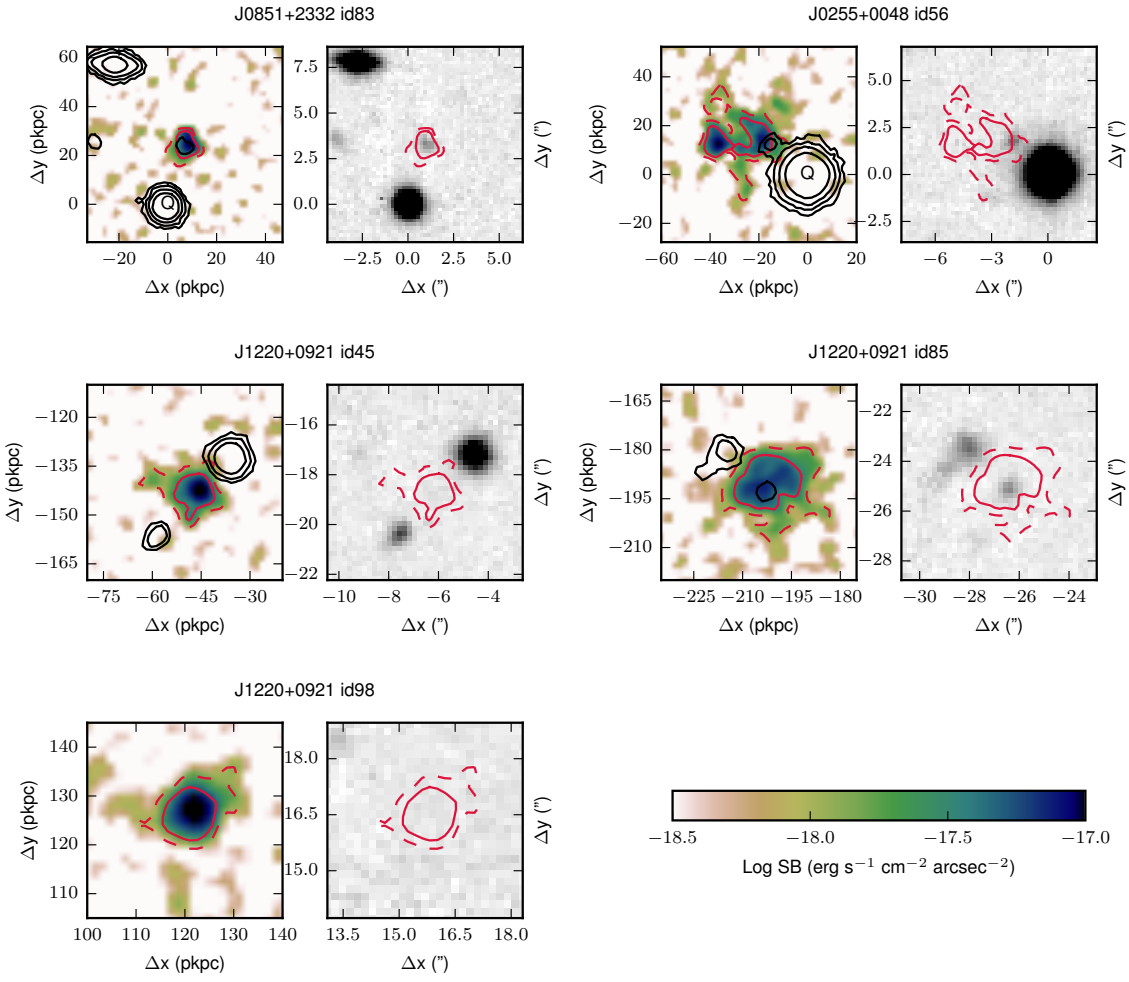


Figure 4.10: Postage stamps of the confirmed DLA associations. Each object has an optimally extracted Ly$\alpha$ surface brightness map (left) and an $r$ band image (right). Both images have Ly$\alpha$ SB contours (red), drawn at $10^{-18}$ erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$ (dashed lines) and $10^{-17.5}$ erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$ (solid lines) . The Ly$\alpha$ SB map is smoothed with a Gaussian kernel ($\sigma = 1$ pixel) and has $r$ band continuum contours overlaid in black. The scale of the SB map is in physical kpc from the DLA, while the $r$ band image is shown in arcsec from the quasar.

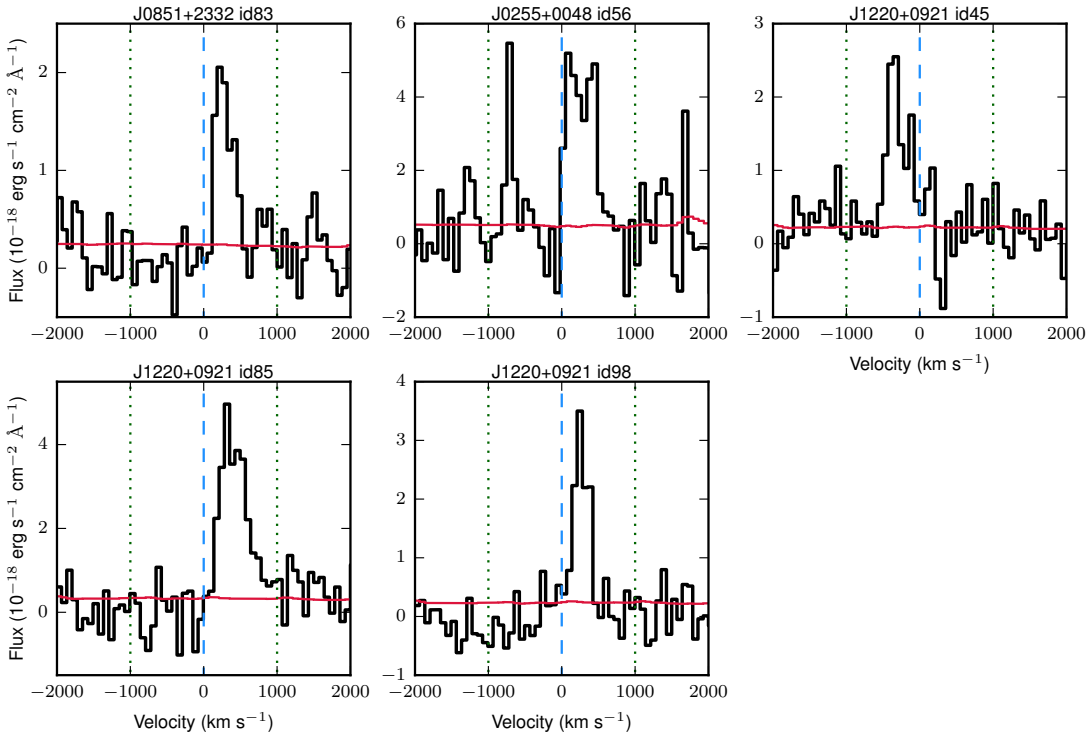


Figure 4.11: The Ly$\alpha$ emission lines for the high-confidence DLA associations. The 1D extracted spectrum is shown in black, with the 1$\sigma$ error in red, plotted as a function of velocity offset from the DLA absorption redshift. Vertical blue dashed lines indicate the DLA redshift, while green dotted lines mark the velocity window over which LAEs were extracted ($\pm$1000 km s$^{-1}$). Fluxes have been corrected for Galactic extinction.

## 4.6 Discussion

Following the analysis of the observations in Sect. 4.2 and Sect. 4.3, we have detailed the models we use in this work for the interpretation of our data in Sect. 4.5. Starting with the EAGLE simulations, we derive DLAs either by post-processing the simulation box with the URCHIN radiative-transfer code (hereafter URCHIN model), or by "painting" DLAs to halos from the simulations using a simple prescription that can be calibrated to match the large-scale bias of DLAs (hereafter painted models). We also introduce two simple models for LAEs, on one in EAGLE and one based on a semi-analytic prescription, both of which are calibrated to match the luminosity function of LAEs.

We now turn the discussion of our observations in the context of previous searches of DLA associations (Sect. 4.6.1), moving next to the interpretation of our observations with models (Sect. 4.6.2), and concluding with forecasts of how future searches will refine the determination of the typical properties of DLA hosts via small-scale clustering of Ly$\alpha$ emitters (Sect. 4.6.3).

### 4.6.1 Detection rates and comparison with previous studies

Previous searches for DLA host galaxies have revealed fewer than 20 spectroscopically confirmed DLA host galaxies at $z > 2$. These have been identified in more than a decade by a range of surveys with many different instruments including a highly-successful long campaign with X-Shooter (see a summary in Krogager et al., 2017). Our MUSE survey, in under 30 hours of observations, has uncovered an unprecedented large sample of associations, detecting high-purity objects in three out of six targeted DLA fields, and more systems (although with lower confidence) in five of six fields we have observed. As discussed in Sect. 4.3, it is difficult to cleanly separate the lower confidence LAEs from contaminants, but based on the their clustering in velocity space it is very likely many of these are real DLA associations.

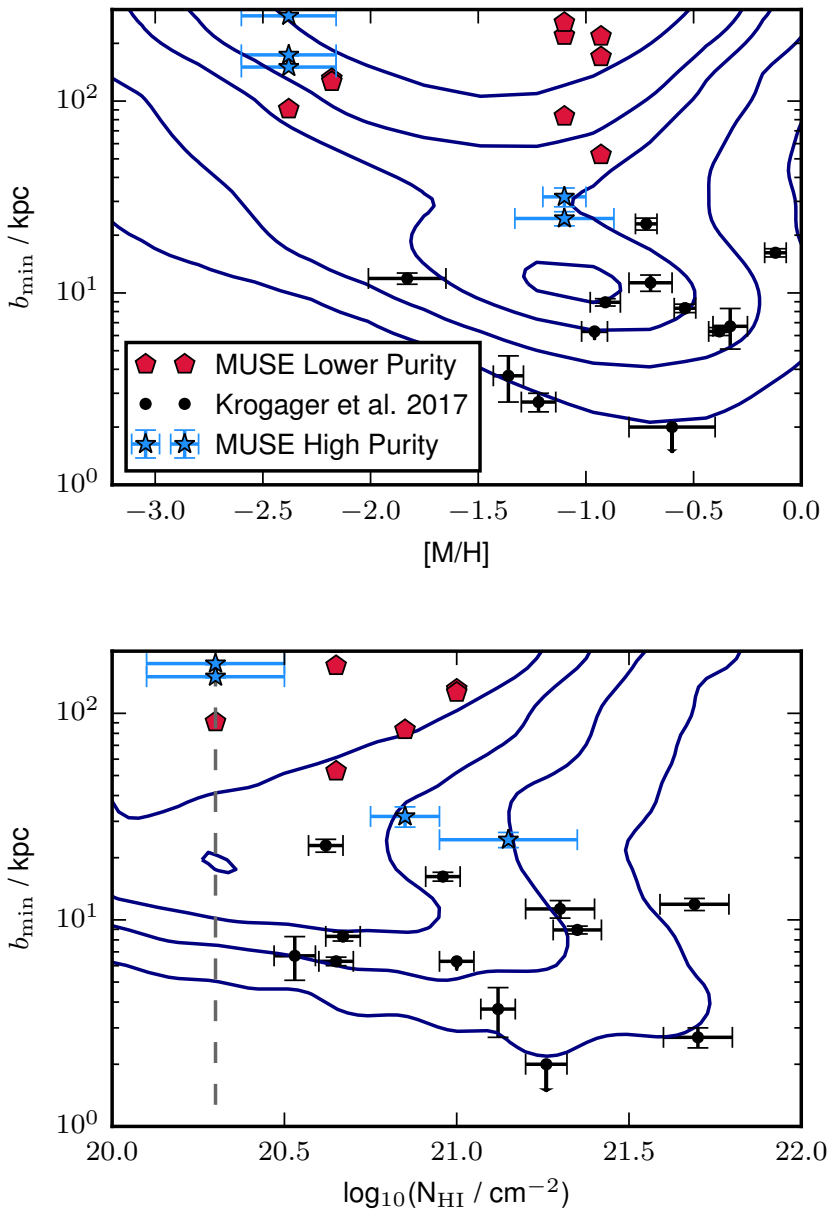


Figure 4.12: MUSE detected DLA associations in context with current samples of $z > 1.5$ DLA host galaxies. The figure shows the galaxy impact parameter (in physical kpc) as a function of the DLA metallicity (top) and column density (bottom). The comparison sample (complied in Krogager et al., 2017, black points) and MUSE high confidence detections (blue stars) are plotted. The contours show predictions of the distribution of LAEs from the EAGLE simulation using URCHIN DLAs. The simulated contours apply for a MUSE like selection and don't simulate the varied selection of the literature sample. All high and low purity detections are plotted. In the lower panel the DLA column density threshold is marked (vertical dashed line).

Taking only the high-purity detections as the lower limit, we establish a detection rate of at least 50%, which rises to 60% if one excludes the field J0818+2631 which suffered from shallower and incomplete observations. While we have been extremely successful in discovering associated galaxies, we have only detected one plausible host galaxy. This detection rate is consistent with surveys that did not pre-select targets in metallicity (Møller et al., 2002). This is considerably lower than metallicity selected samples (e.g. Krogager et al. 2017, 64%), but it enabled the study of the clustering of all DLAs.

This work therefore confirms the competitiveness of Ly$\alpha$ as a means to search for DLA host galaxies, and more generally associated galaxies Indeed, we are searching to larger impact parameters that some of the previous surveys, and have detected in some cases galaxies at sufficiently large impact parameters to make unlikely a direct association between these LAEs and the gas measured in absorption. Ly$\alpha$ also appears a powerful complementary technique to searches with ALMA at high metallicity (Neeleman et al., 2017), where obscured systems where Ly$\alpha$ would be absent due to scattering can be revealed instead via FIR and [CII] emission. While current detections with ALMA are confined to highly star forming galaxies (24 and 110 $M_\odot$ $yr^{-1}$), MUSE enables the detection of SFRs of the order of 1 $M_\odot$ $yr^{-1}$ and, hence, is sensitive to the lower SFR galaxies that, arguably, constitute the bulk of the DLA hosts.

To this end, we note how our sources with fluxes $\gtrsim 10^{-18}$ erg $s^{-1}$ $cm^{-2}$ and luminosities of $\gtrsim 10^{41}$ erg $s^{-1}$ overlap with the population of faint emitters detected in deep long-slit observations by Rauch et al. (2008). It is therefore quite plausible that our programme has in fact finally detected within quasar fields the tip-of-the-iceberg of the faint population of LAEs that host the majority of the DLA population. As argued in Rauch et al. (2008), these small proto-galactic clumps have individually only limited cross section, but are numerous enough to explain the abundance of DLAs, which in turn is consistent with the number density of faint LAEs (Drake et al., 2017).

As our search differs from many previous efforts in that we are not limited to searches at small-impact parameters due to the large MUSE field of view nor preselect fields based on absorption properties, it is interesting to compare our detected associations with the existing sample of spectroscopically confirmed host galaxies. Fig. 4.12 shows a comparison between the MUSE detected associations and the DLA host galaxies compiled in Krogager et al. (2017), plotted as impact parameter against the metallicity and column density of the DLA. In the case of the MUSE associations, the closest object is shown in analogy to previous studies, including both the high-purity and lower confidence sample.

The first observation is that indeed the MUSE detected associations are almost exclusively at larger impact parameters than the literature hosts. It is expected that $b_{min}$ will be biased to larger radii in the case of the MUSE observations, as we have plotted the closest galaxy even in the cases where it is unlikely there is a direct link with the absorbing gas. However, that does not immediately explain why all our detections would be found at larger impact parameters compared to previous searches, especially for overlapping values of metallicity and column density. We argue instead that our search, although not revealing in many case the direct host galaxies, provide a more representative view of the typical galaxy population around DLAs. Previous searches, either by virtue of the detection method or by the fact the only the closest galaxies may have been reported as DLA hosts in most instances, are likely to have yielded samples that include only the brightest and closest associations without capturing the full or even more typical distribution of galaxy properties near DLAs. When we include all detected sources the top part of the plot becomes highly populated. Bimodality can be seen in the EAGLE contours for both plots, this is the boundary between physically associated host galaxies and close associations.

From the top panel of Fig. 4.12 it there doesn't appear to be to be a strong correlation between impact parameter and metallicity, indeed it is believed that there is only a weak relation between the two properties (Christensen et al., 2014). The lower panel shows a convincing anti-correlation between impact parameter and column density.

This consistent with the predictions from simulations (Rahmati & Schaye, 2014).

### 4.6.2 Constraints on the DLA host halo mass

As discussed in the previous sections, in most cases we have no evidence that the detected associations correspond to the DLA host galaxy, but rather the detected LAEs act as tracers of the environment within which the DLAs arise. Nevertheless, within a given cosmological model, we can still constrain the properties of the host galaxies (and in particular their typical halo mass) by comparing the small-scale clustering of LAEs with predictions as a function of halo mass. This approach is similar to the DLA/LBG cross-correlation analysis by (Cooke et al., 2006), albeit on smaller scales, and further offers additional insight into the high-bias reported by (Pérez-Ràfols et al., 2018b)

To this end, we first calculate radial density profiles of LAEs around DLA sightlines. These profiles are constructed by selecting a DLA and counting the mean number of LAEs inside a velocity window and given radius $D_{LOS}$, as a function of $D_{LOS}$. This is shown in Fig. 4.13, where the MUSE observations are compared against both URCHIN DLAs and the results of the painted halo scheme, applied to both the EAGLE 100 Mpc simulation and the semi-analytic model. The painted model shown in this case has $\alpha$=0.75 using the parameters shown in Table 4.4. We use the high confidence sample for this task as there are uncertainties in the purity and completeness of the lower SNR candidates. As the high-purity LAEs extend down only to a luminosity of $\approx 10^{42}$ erg s$^{-1}$, we apply this limit also to the modelled LAEs. We empirically estimate the uncertainty in the measured profile using jackknife resampling. As at $D_{LOS} < 24$ kpc there are no detections, we place an upper limit on the profile by calculating the maximum rate at which we would expect at least one detection over six DLAs with 66.7% confidence assuming Poissonian statistics. Based on our (admittedly large) empirical uncertainties both the painted halos and URCHIN DLAs are consistent with our observations (perhaps with a preference for

the painted models), implying that models of LAEs clustered within dark matter halos are able to reproduce the observed radial profile. Here, we take the velocity window to be $\pm$650 km s$^{-1}$, as this corresponds to the length of the EAGLE 25 Mpc simulation at $z = 3.017$. The velocity window cannot be set any higher for the 25 Mpc simulation as this is the maximum separation two objects can have. However, this does not affect the MUSE observations as none of the detections were outside this smaller velocity interval ($\pm$1000 km s$^{-1}$was the original search window).

Next, to remove the geometric effect of greater volume as $D_{LOS}$ increases, we convert the number of LAEs within $D_{LOS}$ to a density measured in a cylindrical aperture. Additionally we divide these density profiles by the mean number density of LAEs, to convert the density profiles in terms of over-density of LAEs with respect to the mean density. In the case of the EAGLE simulations the mean number density is simply calculated from the model, while for our MUSE observations and the GALICS mock catalogue, the mean density is taken from integrating the Drake et al. (2017) Schechter luminosity function down to Ly$\alpha$ luminosities $10^{42}$ erg s$^{-1}$.

Fig. 4.14 (top left) shows the relative density profile for the MUSE DLA observations in comparison to both model DLAs identified with URCHIN and from the painted a halo model. The halo model used here is $\alpha$=0.75, which is the same for all three painted simulations, the GALICS mock catalogue and the EAGLE 25 and 100 Mpc boxes. We use this intermediate model as it corresponds to halo masses which are suitably converged in all three simulations. The shape of the relative density profile probes several properties of the DLA population. The initial plateau of the curves at small radii describes the extent of individual DLAs, while the tail encodes small-scale clustering information.

Inspecting this panel, the most notable difference is between the density profile from the URCHIN DLAs and those painted with the halo prescription: the density at larger scales is much less enhanced for URCHIN DLAs. This is because, within EAGLE, DLAs populate also smaller halos due to the lack of low-mass cut-off, and are thus less clustered on average. We also observe that the painted model and results from the

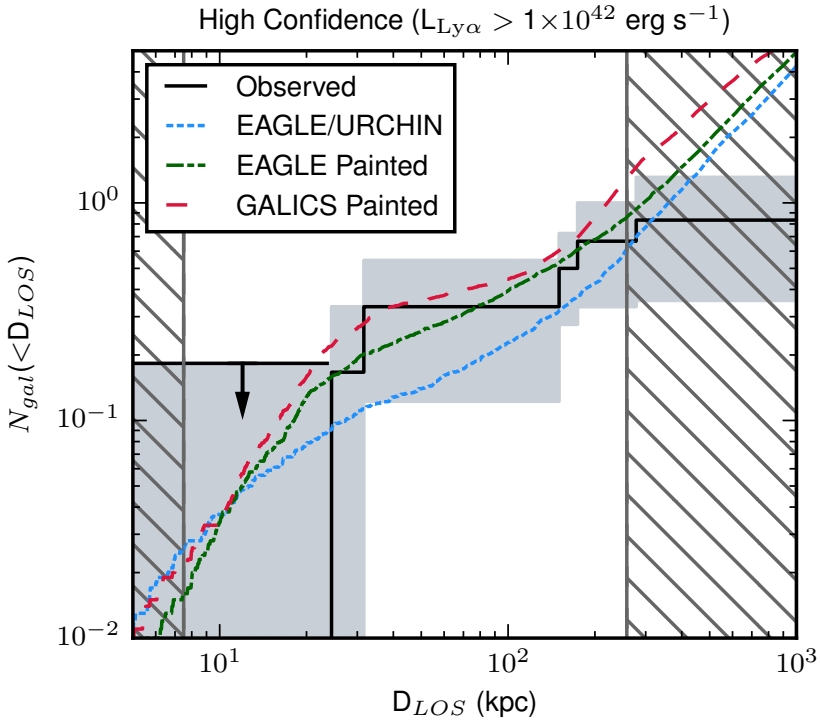


Figure 4.13: The cumulative distribution of LAEs around DLA sightlines for our MUSE observations (high confidence only, black) and two simulated datasets. $N_{gal}$ is the average number of bright LAEs within a radius ($D_{LOS}$) and a velocity window of 650 km s$^{-1}$from a DLA. The DLAs used in calculating the simulated profiles come from EAGLE with URCHIN post-processing (blue, dotted) and the painted halos applied to both the EAGLE 100 Mpc (green, dot-dashed) and GALICS (red dashed) data. The $1\sigma$ error in the profile from the observations is estimated via jackknife resampling (grey shaded), and, at small scales where there are no detections, the 66.7% upper limit is shown (black arrow). The grey hatched regions indicate the limits of scales probed in the MUSE observations, between 1 arcsec (of the order of the seeing) and 0.5 arcmin (the field of view) converted into a proper distance at the typical DLA redshift.

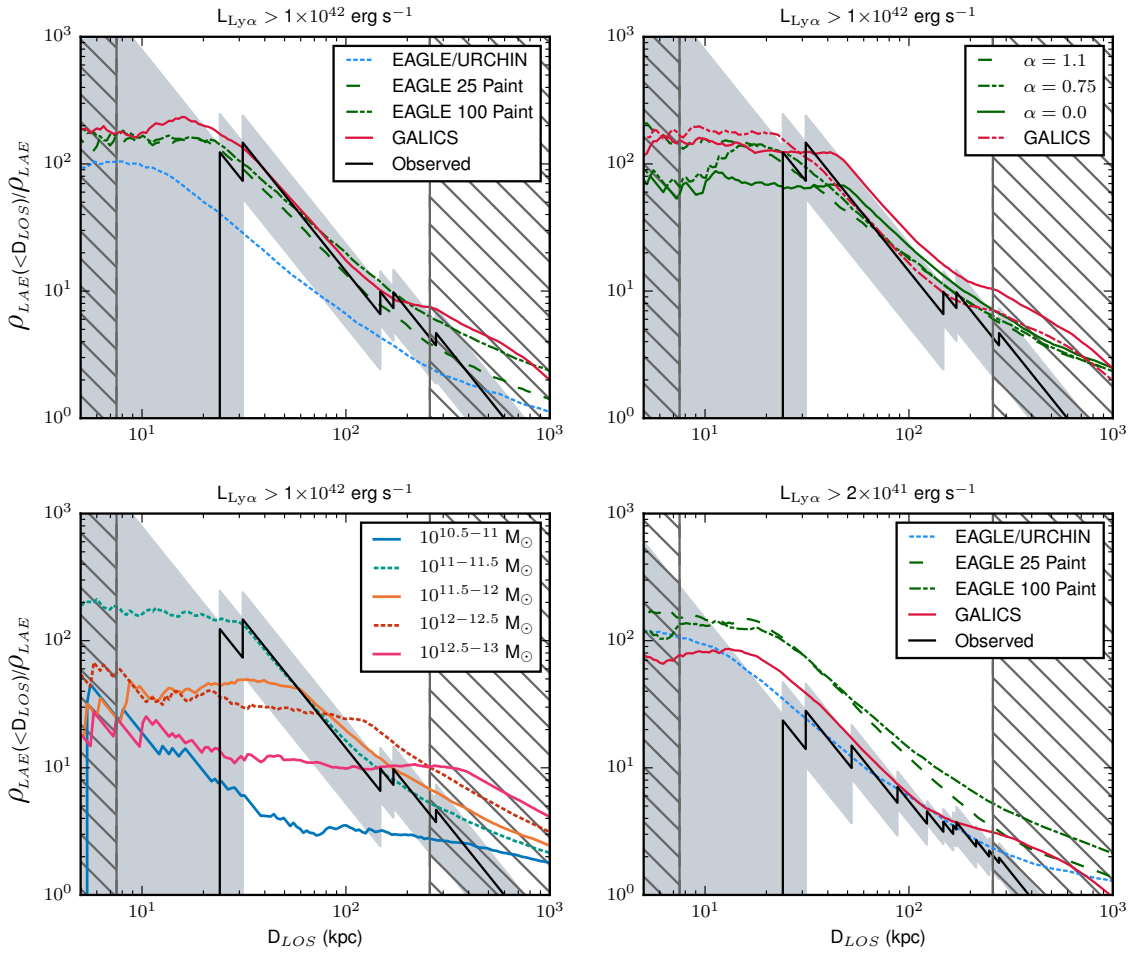


Figure 4.14: The radial over-density profile of LAEs around DLAs, computed as the density of LAEs with respect to the mean density, measured as a function of the circular aperture $D_{LOS}$ from a DLA, simulated and observed. **Top left:** Shown are our MUSE results (black, solid) in comparison to the DLAs defined by URCHIN post-processing of the EAGLE 25 Mpc box and the painting model that matches the Pérez-Ràfols et al. (2018b) results. The painted DLA are shown for both the GALICS mock catalogue (blue, solid) and the EAGLE 25 and 100 Mpc simulations (green, dot-dashed and dashed lines respectively). **Top right:** The MUSE observations are shown against results from the EAGLE 100 Mpc box with painted DLAs shown with 3 different model parameters ($\alpha = 1.1, 0.75$ and $0.0$). The profile from the GALICS lightcone is shown (red) only for $\alpha = 0.75$. **Lower left:** Halos from the EAGLE 100 Mpc box are selected by halo mass ($M_{\rm h}$) and painted with a uniform cross-section. The selected ranges are shown in the legend. The $10^{11}$-$10^{11.5}$ M$_\odot$that best matches the observed profile is similar to the characteristic halo mass estimated from DLA clustering. **Lower right:** The same plot as the top right but now using all detected LAEs down to a luminosity of $2 \times 10^{41}$ erg s$^{-1}$s. The completeness estimated for the MUSE data (for extended sources) has been applied to the simulations.

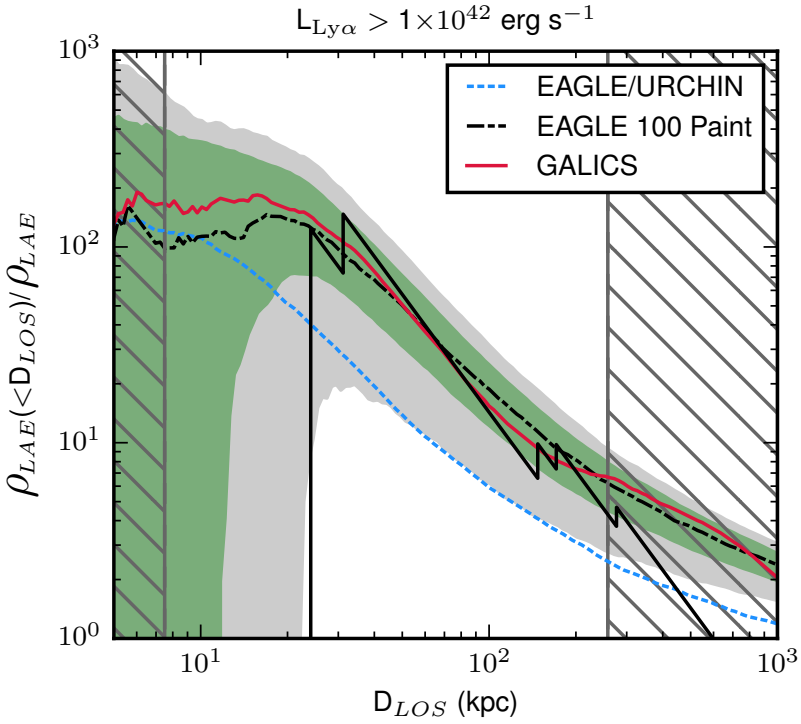


Figure 4.15: The simulated scatter of sets of 6 DLA with MUSE-like observations from the EAGLE 100 Mpc painted simulation (grey shaded). The error that would be obtained with a hypothetical 24 DLA MUSE survey is also displayed (green shaded).

URCHIN calculation in the EAGLE 25 Mpc converge at large scales, as expected by construction given that they are the same box and they should converge to the same mean density. We note that some of the suppression of URCHIN results with respect to the 100 Mpc simulation and GALICS may be due to limited volume, as the 25 Mpc painted DLAs lie below the two larger painted simulations. We investigate next whether the discrepancy with the URCHIN DLAs is significant. The good agreement between the 100 Mpc EAGLE box and the GALICS lightcone with painted DLAs strengthens these results, showing that the assumptions made about LAEs in EAGLE are sufficiently robust for this comparison. From Fig. 4.14, the most substantial conclusion is that the MUSE observations appear to be in good agreement with the halo prescription of Font-Ribera et al. (2012).

In Fig. 4.14 (top right), the relative density profiles for the three sets of parameters for the halo prescription (eq. 4.5.3, table 4.4) are shown when applied to the EAGLE 100 Mpc box (green). For $\alpha = 0.75$, the profile from the GALICS mock is also shown

(red). GALICS suffers from limited mass resolution below halo masses of $\sim 10^{10}$ $M_\odot$ and limited volume at high masses ($> 2\times10^{12}$ $M_\odot$ ) so we only apply the intermediate model which has a high enough cut-off mass but is not dominated by the high mass tail. As described above, in each case, the absolute normalisation ($\Sigma_0$) is fixed to reproduce the DLA number density, $\ell(X)$, derived with URCHIN post-processing in the 25 Mpc EAGLE simulation. At large scales ($>$ 100 kpc), the three models are indistinguishable, as it is expected given that all three sets of parameters are chosen to match the observed bias of DLAs measured at large scales. The largest difference between the different models is at scales of $\simeq$ 20 kpc, which is due to the changing distribution of DLA sizes affecting where the flat part of the curve ends. For the most extreme model ($\alpha = 0$), the DLA cross-section is fixed for all halos above $M_{min}$, and in this case the break in the profile is set by this fixed DLA radius (seen at 30 kpc for the $10^{11-11.5}$ $M_\odot$ bin).

The relative density profile is largely insensitive to these different painting models, despite the minimum mass raising from $3\times10^9$ to $2.2\times10^{11}$ $M_\odot$. This is an indication that the relative density profile in the halo prescription is mainly dominated by the effect of the massive halos ($M_h > 10^{11}$ $M_\odot$), as shown more explicitly in Fig. 4.14 (lower left). In this panel, the contribution of different halos has been split into bins of halo mass, where in each bin the halos are painted with a constant cross-section (i.e. $\alpha$=0). As before, the constant $\Sigma_0$ is calibrated in the 100 Mpc simulation to match the $\ell(X)$ predicted from the URCHIN calculation. In the lowest mass bin, the halos simply do not contain enough LAEs or cluster with them strongly. In the $10^{11} < M_h < 10^{11.5}$ $M_\odot$ bin the profile reproduces the MUSE data well. Above $10^{12}$ $M_\odot$ the profile peak declines, because the DLA cross-sections have become so large that some chosen DLAs can be very far from the halo center. Fig. 4.14 shows that, in this painting scheme, the characteristic halo mass preferred by our observations is $10^{11} < M_h < 10^{12}$ $M_\odot$ which interestingly agrees with the halo mass yielding an equivalent bias to that of inferred from DLAs ($\approx 10^{11.78}$ $M_\odot$ , Pérez-Ràfols et al. 2018b). This single mass model is clearly a simplification, and we can refine this

estimate by returning to Eq. 4.5.3 of the model rather than assuming a constant dependence with mass. Using the 100 Mpc EAGLE simulation, we compute the mean halo mass of a DLA in this scheme, finding that, for the values of $\alpha$ we tested, this characteristic halo mass falls in the same range as before, ranging from $10^{11.27}$ ($\alpha = 1.1$) to $10^{11.78}$ $M_\odot$ ($\alpha = 0$). Taken together, the convergence of these estimates on the characteristic halo mass corroborates the finding of a high bias inferred from the large-scale clustering measurement.

In the above comparisons, we have utilised only the high-purity LAE sample as the fainter detections suffer from limited completeness and may lack purity. To test the robustness of our results, we now lower the threshold to include LAEs with $L_{Ly\alpha} \geq 2 \times 10^{41}$ erg s$^{-1}$. As we are using a velocity window of $\pm 650$ km s$^{-1}$ (as explained above), we have removed the candidates outside this range. As discussed in Appendix A.1, there is limited completeness to this luminosity threshold, therefore we must apply a correction prior to comparison with simulated profiles. To do so, we have chosen to apply the estimated incompleteness function (derived from the mean completeness measured over all six fields using realistic mock sources) to the simulations, as this procedure is less noisy than weighting the data by completeness. Fig. 4.14 (lower right) shows the observations compared with the modelled results, showing that at this lower luminosity, the higher number of objects has improved our constraints on the density profile. It can be seen that when considering these fainter candidates the data now appears to match the EAGLEURCHIN and GALICS profiles more closely, as opposed to the painted EAGLE preferred by the higher luminosity data. Indeed, the data now weakly reject the painted model applied to the EAGLE 100 Mpc simulation. This discrepancy between low and high luminosities could be understood if the Ly$\alpha$ luminosity function differs around DLAs with respect to the field. However, as the disagreement extends to 200 kpc, this appears to be unlikely, as this scale is far beyond typical halos at this redshift. A much more plausible hypothesis is that the LAE completeness is lower than estimated and hence the data are biased low with respect to the models. Indeed, as detailed in Appendix A.1 our

estimate of completeness does not include the additional manual vetoing stage, which cannot be trivially reproduced in large samples. It is therefore likely the completeness measured is overestimated. This uncertainty justifies our previous choice to focus our analysis on the high confidence sample. We do emphasise, however, that our sample is still rather small, and therefore any current inference is likely to be subject to sample variance (see next section).

Fig. 4.14 also shows a divergence in model predictions with luminosity. At high luminosity ($L_{Ly\alpha} > 10^{42}$ erg s$^{-1}$) both the painted EAGLE and GALICS profiles are in very close agreement for the same model parameters. However, at the lower luminosity, the predictions start diverging despite being derived with the same prescription. Future datasets will require more precise modelling of LAE properties, and - conversely - deeper observations will be able to discriminate more among different models.

### 4.6.3 Forecasts for future searches

In comparing our results to those of simulations, we have quantified the expected uncertainty in the radial density profile using resampling. With only 6 DLAs, this may still underestimate the errors in the radial density profile. To quantify more robustly the uncertainty arising from sample variance, we drew sets of 6 DLAs from the EAGLE 100 Mpc box and use them to estimate the radial density profile. Iterating over 500 realizations of 6 DLAs each, we derive the scatter expected between sets of 6 DLAs. To quantify the uncertainty we estimate the standard deviation over all realizations. For this task, we use the EAGLE 100 Mpc box using DLAs painted onto halos with $\alpha = 0.75$. We prefer the larger simulation, as it provides many more independent realisations than the 25 Mpc box. Fig. 4.14 shows the $1\sigma$ uncertainty (grey shaded) estimated from this Monte Carlo simulation for a survey of 6 DLA fields with MUSE like observations. The simulated uncertainty is indeed larger than the errors estimated though resampling, although the two results are comparable.

Based on this, we conclude that currently our results are compatible with both models described in this paper.

Also shown in Fig. 4.14 is the $1\sigma$ uncertainty one would obtain with a sample of 24 DLA fields studied with MUSE (green shaded), showing that with future observations it may be possible to start discriminating between models. Datasets of this size are achievable in the near term with MUSE. Interestingly, if the underlying radial density profile were as low as predicted by the EAGLE simulation with URCHIN post-processing, the BOSS halo painting prescriptions could be ruled out. Some degeneracy will, however, exist between different DLA models and the clustering of the LAEs used in this cross-correlation, but it should be possible to break this degeneracy with surveys of LAEs that measure the large-scale clustering properties (e.g. MUSE-Wide and HETDEX at higher luminosities; Diener et al. 2017).

## 4.7 Summary and Conclusions

We report the results from the first MUSE spectroscopic survey of high-redshift ($z \gtrsim 3$) DLAs in six quasar fields. We have searched for associated galaxies close to six $3.2 < z < 3.8$ DLAs with deep rest-frame UV integral field spectroscopy, finding five high-purity associations of three DLAs, plus 9 additional lower-significance objects in five out of six fields. Together with a very high detection rate (up to $\approx 80\%$) in a sample not pre-selected by absorption properties, we report for the first time the detection of three to four LAEs to a very metal-poor DLA at high redshift. This DLA, in the field J1220+0921, appears to be embedded in a filamentary structure, as traced by three bright LAEs stretching across the full MUSE FoV.

We have also compared our results to the predictions of recent hydrodynamical simulations and a simple prescription in which DLAs are painted inside dark matter halos, showing that our detections of multiple associations support a picture in which DLAs arise from neutral gas in proximity to several galaxies clustered within dark matter halos. We have also explored the small-scale clustering of DLAs, by comparing

the density of LAEs detected around DLAs to cosmological models in which DLAs and galaxies populate dark matter halos. The results of our MUSE survey are consistent with the predictions from both the EAGLE simulation with radiative transfer post-processing, and a simpler scheme which paints DLA cross-sections onto dark matter halos. More quantitatively, our comparison with simulations shows that a simple halo prescription tuned to reproduce the large-scale clustering of DLAs also explain the number of LAEs associations on small-scales in our observations. Based on this model, we conclude that at least some DLA hosts have a characteristic halo mass of $10^{11} < M_{\rm h} < 10^{12}$ $M_{\odot}$.

Considering individual detections, in the field J0255+0048, MUSE has unveiled an extended 37 kpc Ly$\alpha$ structure, potentially tracing a major merger, possibly hosted in a massive halo. In the field of DLA J1220+0921 we have discovered a rich environment, with several LAEs surrounding the DLA. Moreover, in half of our sample, we identified faint LBGs in proximity to the DLAs (see appendix A.3). Taken together these lines of evidence point to a role of more massive halos in association with some significant fraction of the DLA population, and clearly shows the need for models incorporating clustering beyond the more traditional models treating individual galaxy/DLA associations. We finally showed how (deeper) MUSE observations in large sample of $\approx 25$ DLAs will provide sufficient statistical power to discriminate among different models, and refine the mass estimates of halos hosting DLAs. Given the efficiency of MUSE, it should be possible to assemble such a sample in the near future.

# Chapter 5

# Summary and Conclusions

In this thesis, we have explored the origin of the CMB Cold Spot and the connection between galaxies and neutral gas reservoirs at high redshift, in both cases using high multiplex spectroscopy. In Chapter 2, we conducted a redshift survey of $\sim 7000$ galaxies at $z < 0.5$ sampling the $5^\circ$ (radius) core of the CMB Cold Spot, in order to test claims that the Cold Spot was imprinted by a large void. In Chapters 3 and 4, we present the results of a MUSE integral field survey of six Damped Lyman $\alpha$ systems at $z \simeq 3.5$, studying their environments and searching for host galaxies. Chapter 3 presents a detailed analysis of one extraordinary system detected in the survey, which may be an ongoing merger. Chapter 4 presents the full analysis of the DLA survey and a statistical comparison between the distribution of galaxies as observed and in simulations. First, we will review the key findings of the Cold Spot survey.

## 5.1 The CMB Cold Spot

The Cold Spot is a feature in the Cosmic Microwave Background which has been claimed to represent a departure from standard cosmology (Vielva et al., 2004), however its statistical significance is debated (Cruz et al., 2006; Zhang & Huterer, 2010). Inoue & Silk (2006) proposed that a large void may have imprinted the Cold

Spot via the Integrated Sachs-Wolfe effect, and many attempts were made to search for such a structure (e.g. Rudnick et al. 2007; Granett et al. 2010; Bremer et al. 2010). In particular, Szapudi et al. (2015) claimed the detection of a large $z = 0.2$ supervoid aligned with the Cold Spot, detected using photometric redshifts. The authors claimed that the rarity of such a structure makes a causal link with the Cold Spot more credible than the chance alignment of two rare features. In order to test these claims, we conducted the 2dF Cold Spot redshift survey (2CSz) with 2dF+AAOmega, surveying 66 deg$^2$ of the core of the Cold Spot. The survey selection was a simple magnitude limit ($i < 19.2$) which enabled us to exploit the existing GAMA survey (Driver et al., 2009; Liske et al., 2015) as a control field, and is well matched to the target redshift range $z < 0.4$. By studying the distribution of galaxies along the line of sight and fitting them with an analytic form for a perturbation we make the following conclusions:

- A 34% under-density with a radius of 119 h$^{-1}$Mpc was detected with a central redshift of $z = 0.14$. This structure is much narrower in redshift than measured in photometric redshift analyses in Szapudi et al. (2015), but it is a deeper under-density. Two smaller under-densities were observed at $z \simeq 0.26$ and 0.30. The total ΛCDM ISW effect from these three voids is estimated at $-9.1 \pm 6.1\mu$K, where the decrement is dominated by the large $z = 0.14$ void. This temperature decrement is much too small to explain the CMB Cold Spot which has a core temperature of -150 $\mu$K.

- In the tail of the survey redshift distribution a candidate void was detected at $z = 0.42$ with a scale radius of 162 h$^{-1}$Mpc. Even if this void were physical, it would only have an ISW imprint of -22.6±14.7 $\mu$K, still insufficient to explain most of the Cold Spot. We conclude this feature is likely contaminated by systematic uncertainties in the survey such as systematic redshift failures and ambiguities in the modelled redshift distribution.

- The structure along the line of sight seen in 2CSz bears similarities to that

of our control field G23. In particular, G23 also has a large under-density at $z = 0.42$, lending support to the idea that these features are caused by a systematic effect. The correspondence between these features in the Cold Spot and in G23 indicates that the detected voids are not sufficiently rare to justify appeals to alternative cosmologies. G23 has no Cold Spot like feature in the CMB.

- Our redshift survey, when combined with previous data (Bremer et al., 2010), rules out the presence of a very large void capable of explaining the CMB Cold Spot under a ΛCDM ISW effect up to $z \sim 1$. At higher redshifts the ISW effect becomes less significant, as the universe tends to matter-domination at this time.

Taken together these results rule out a large contribution from a supervoid to the CMB Cold Spot under standard cosmology. Simultaneously, the similarity with G23 makes it difficult to imagine how an alternative cosmology would generate a large imprint from the Cold Spot sightline without doing so in G23. Kovács (2018) has argued for a greatly enhanced ISW amplitude for large voids, perhaps up to ten times the ΛCDM expectation. This proposal will have to be reconciled with the lack of Cold Spot like feature for G23 or other voids of this scale (Nadathur & Crittenden, 2016; Kovács et al., 2017).

Further properties of the Cold Spot voids that we did not explore in the above work are the three dimensional structure and the large-scale environment of the voids. With our sampling of ∼100 galaxies per deg$^2$ we did not deem it practical to extract the structure of the under-densities in three dimensions. For the large-scale environment it was noted when calibrating the survey that the galaxy number density measured near the Cold Spot is $\sim 8\%$ lower than near G23, measured in the VST-ATLAS imaging used for target selection. If this difference is not due to a gradient in the VST-ATLAS photometric calibration, it would indicate a large under-density in the South Galactic Cap, possibly related to the Local Hole (e.g.

Whitbourn & Shanks 2014). If this under-density extends to higher redshift, it may be that the voids differ from those in G23 by their large-scale environment, the Cold Spot voids may be embedded in a larger scale under-density while the G23 voids are not. One should, however, be cautious about searching many parameters to find something anomalous about these structures, in order to justify claims that they generated the CMB Cold Spot. Future wide-field redshift surveys in the South such as TAIPAN (Kuehn et al., 2014), WALLABY (Koribalski & Staveley-Smith, 2009) and 4MOST (de Jong et al., 2012) should provide context for the 2CSz results.

With the ISW effect largely ruled out under ΛCDM, the question of whether or not the Cold Spot represents a problem for standard cosmology remains. There are numerous proposals to explain the Cold Spot with models which could be an extension to standard cosmology such as non-standard inflation (e.g. Bueno Sánchez, 2014) or more exotic scenarios such as topological defect or an inflationary multiverse (e.g. Cruz et al., 2007; Larjo & Levi, 2010). Leaving aside possible further tests of these models, assessing the significance of the Cold Spot is problematic. The effect of *a posteriori* choices in estimating the significance is a difficult problem to overcome. To properly assess the significance of an outlier the size of the parameter space searched must be taken into account. The Cold Spot was detected when applying a Spherical-Mexican-Hat wavelet filter to CMB maps (Vielva et al., 2004), the parameter space searched in that analysis was small. But different filter radii were studied along with different distribution moments, and the scale radius was chosen to maximise the significance (at the time $\simeq 3\sigma$). Zhang & Huterer (2010) argued that the filter radius, statistic and even the wavelet were selected *a posteriori*. They showed that if one searched Gaussian CMB realisations for the most extreme statistic in each realisation, then the Cold Spot was not anomalous at all. However, this approach goes beyond the parameter space studied in Vielva et al. (2004), it included different filters and statistics not used in the original analysis. The significance of the Cold Spot is therefore difficult to estimate objectively, as it depends on one's interpretation of how it was found. Some non-standard hypotheses to explain the

Cold Spot make further testable predictions, but for now the Cold Spot does not appear to be in conflict with standard cosmology.

## 5.2 Gas and Galaxies

In Chapters 3 and 4, we presented the results from a integral field survey of Damped Lyman $\alpha$ systems (DLAs) at $z \simeq 3.5$. DLAs are high column density absorption line systems found in absorption against distant quasars, and host most of the neutral hydrogen in the universe over a wide range in redshift ($z = 0 - 5$). An understanding of the connection between DLAs and possible host galaxies has long been sought, because knowledge of how DLAs trace the galaxy population is sensitive to key galaxy evolution processes like feedback (e.g. Barnes & Haehnelt, 2014). With MUSE, we studied six DLA fields: in Chapter 3 we discuss in detail an extraordinary ∼40 kpc extended structure discovered in association with a DLA, while chapter 4 contains the results from the full sample and a detailed comparison of the distribution of galaxies near the DLAs as observed and predicted with a hydrodynamical simulation and a semi-analytic model. The key findings are as follows:

- We targeted a sample of six DLAs which spanned a range of absorption parameters with MUSE, three fields have been revealed to host Ly$\alpha$ bright galaxies at high significance within a few hundred km s$^{-1}$ of the absorption redshift. Furthermore, we identified lower signal-to-noise ratio galaxies in five of the six fields, but the purity and completeness of this sample cannot however be guaranteed. Even when considering only the high significance detections, 50% is unprecedented for an unselected survey of high redshift DLAs.

- In the first field that was studied in detail in this survey (J0255+0048), we found the DLA to lie very close to a Ly$\alpha$ nebula, extending up to 50 kpc from the DLA. The structure is dominated by two unresolved clumps which each contribute about 1/3 of the Ly$\alpha$ luminosity, and these are separated

in velocity space by $\Delta v = 140 \pm 20$ km s$^{-1}$. Embedded near the edge of the nebula is a compact continuum source, which appears to be a Lyman Break Galaxy. Aside from this source, the rest of the structure is undetected in deep continuum imaging. In absorption the system is dominated by two strong velocity components which are separated by $155 \pm 6$ km s$^{-1}$, which is consistent with the separation observed between the clumps in emission. This correlation in velocity between absorption and emission could be understood if this system is an ongoing merger, with the material at the DLA position being tidal debris. Furthermore it is difficult energetically for the LBG to power the Ly$\alpha$ luminosity of the structure, but the data are consistent with the clumps being powered by young ($< 7$ Myr) in-situ starbursts. This further supports the hypothesis of an ongoing merger, which has recently triggered an episode of star-formation.

- In the most metal-poor DLA in our sample (J1220+0921 with a metallicity of $Z/Z_{\odot} = -2.33$) we discovered the absorber to be located in a dense structure composed of 4 galaxies spanning 380 kpc and the full MUSE field of view. Counterparts to a metal-poor DLA have never been detected at high redshift. The detected galaxies all lie within $\pm 400$ km s$^{-1}$ of the absorption redshift, and their impact parameters range between 89 and 278 kpc. The brightest counterpart is an LBG with an extremely extended Ly$\alpha$ halo. With these impact parameters, it is unlikely that any of these galaxies are physically connected to the DLA material, however they trace the environment that the DLA inhabits. In this case, the detected galaxies and the DLA appear to follow a filamentary structure across the MUSE field of view. In attempting to understand this structure we drew analogues from the EAGLE simulations, finding one example that showed a simulated DLA embedded in a filament connecting two dense regions.

- We compared the observed distribution of galaxies around the DLAs to mock

observations based on the EAGLE simulation and the GALICS semi-analytic model. Our method allows us to compare modelled and observed DLA-galaxy associations without having to assume which detected counterpart is the host galaxy. In the simulated observations DLAs were identified firstly by post-processing EAGLE to calculate H I column densities and secondly by painting DLAs onto dark matter halos. Our comparison shows that the observed distribution of galaxies is consistent with the predictions from both the EAGLE galaxy formation simulations and a simple model put forward to fit the large-scale clustering of DLAs. Although there is some degeneracy with modelling the Ly$\alpha$ galaxy population, we have shown that our results hint at a characteristic DLA halo mass in the range of $10^{11} - 10^{12}$ $M_{\odot}$. Broadly, our results indicate that the perceived tension between clustering measurements and searches for DLA hosts may be over-estimated.

The role of merging galaxies in the high redshift DLA population has previously been studied in simulations. Haehnelt et al. (1998) put forward mergers as a means by which small protogalactic clumps could generate the very large velocity widths observed in some DLAs. J0255+0048 bears many indications that it is ongoing merger, the correspondence between the two components in emission and absorption and potential evidence of a recent starburst. Furthermore, in absorption, the velocity width of this DLA is in the upper 10% of velocity widths for DLAs at similar redshifts. With proper follow-up J0255+0048 may offer the first observational insight into the contribution of mergers and tidal tails to DLA properties.

The bias of DLAs has been a source of controversy since it was measured by the BOSS wide-field spectroscopic survey (Font-Ribera et al., 2012). To reconcile the observed bias with galaxy formation simulations some studies have employed strong stellar feedback in low-mass galaxies, as this suppresses the DLA cross-section of low-mass halos therefore increasing the bias (Barnes & Haehnelt, 2014; Bird et al., 2014). Another tension arose as the halo mass is similar to that of LBGs, and yet most searches failed to find LBGs in close proximity to the quasar (e.g. Fumagalli

et al. 2014). However, we have shown that our results are consistent with a model based on the bias of DLAs and a recent hydrodynamical simulation. This was possible thanks to the integral field method, with which the completeness can be well understood. With larger samples, it may be possible to use the small scale clustering to also probe the physical extent of optically-thick neutral gas in high redshift galaxies.

Optical integral field spectroscopy offers a very efficient method to probe the link between DLAs and galaxies and hence the circum-galactic medium at $z \simeq 3$. Currently most attempts to study this connection target absorption systems and then search for the galaxies around them, with the downside of this technique being that it is difficult to interpret the non-detection of galaxies, which may fall below the detection limit. We have attempted circumvent this problem by studying the small scale clustering with DLAs. Currently the power of these constraints is limited by the small number of fields, but in the future new programmes could dramatically expand this work. In order to best exploit these data we will have to more accurately model the Ly$\alpha$ galaxies in the simulations, which are used to measure the cross-correlation with DLAs.

The ability of MUSE to densely identify high redshift galaxies may open up a new approach to study the CGM: a galaxy-centred analysis where one selects a population of galaxies and then studies the absorption line systems in proximity to them. Thinking about larger scales, the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) will detect several hundred thousand Ly$\alpha$ emitters at $z = 1.9 - 3.5$ to study dark energy, but it will also provide a rich dataset to study the clustering of DLAs on intermediate scales and even bright host galaxies. Deep integral field spectroscopy has also opened up another avenue to study the CGM of high redshift galaxies, by studying their extended Ly$\alpha$ halos (e.g. Leclercq et al. 2017; Wisotzki et al. 2016). The emission mechanism is not well understood, but it offers a means to probe the CGM without the need for a bright background source. However Ly$\alpha$ emission is sensitive to the small scale processes (e.g. clumping), which

are demanding to simulate in galaxy formation simulations. Finally the James Webb Space Telescope will open up a new window on the high redshift universe, and will allow studies of the connection between neutral gas and galaxies to study both reionisation and neutral reservoirs in the CGM.

## 5.3 Future Outlook and Concluding Remarks

Highly multiplexed spectroscopic surveys are incredibly prolific and powerful tools for studying both cosmological questions and the physics of galaxy formation. Soon a new generation of wide field multi-object spectrographs will take cosmology to new heights, with the primary goal of constraining dark energy though baryon acoustic oscillations and redshift-space distortions (DESI, 4MOST, Euclid). Simultaneously high multiplex instruments are being deployed on 8 meter telescopes with multiplex factors of over a thousand (MOONS, PFS), opening up vastly larger samples at high redshift. MUSE has transformed the study of faint emission line galaxies and diffuse Ly$\alpha$ emission, revealing structure beyond the reach of narrowband imaging. In the near future, MUSE will be used to observe deep fields for up to 150 hours, reaching uncharted depths. There is also a $z < 3$ frontier currently probed by KCWI, and a concept design for a Blue-MUSE covering $\lambda \simeq 3600 - 5600$ Å is further poised to produce significant advancements in cosmology and galaxy evolution. With recent advances in instrumentation and near future projects the outlook for studying cosmology and galaxies with high multiplex spectroscopy is promising.

Returning to consider the future of the ISW effect and dark energy, the wide field cosmology surveys that will be conducted in the next decade will be particularly powerful. The galaxy-CMB cross-correlation method for studying the ISW effect will be best applied to LSST and Euclid, which will provide wide and deep imaging with excellent photometric calibration. Within $\Lambda$CDM no significant ISW effect is predicted at $z > 1$, however some modified gravity models predict an effect. Future surveys will much better sample this redshift range, in particular deep continuum

surveys with SKA. Simultaneously, wide field spectroscopic surveys will offer unprecedented volume for stacking analyses of the ISW effect, enabling much more precise identification and measurement of structures. These surveys will put claims of an enhanced ISW effect for photometric redshift voids to the test, on a statistical level rather than for the single sightline demonstrated in Chapter 2. The primary purpose of many of these surveys is to constrain the equation of state of dark energy, these constraints will be greatly advanced in the next decade.

The study of the Circum-Galactic Medium and the relationship between gas and galaxies has much to gain in the coming decades. HETDEX, DESI and WEAVE will refine the clustering measurements of DLAs at $z > 2$. The construction of the Extremely Large Telescopes offers an enormous leap in sensitivity, particularly so in the near infrared with adaptive optics. The ELTs not only offer the ability to study much fainter DLA host galaxies, but resolve them with adaptive optics and study their detailed properties and dynamics. MUSE made substantial progress in studying the diffuse Ly$\alpha$ emission of the CGM, the ELTs potentially posses the sensitivity to study other emission lines from the CGM or perhaps fluorescent emission from the IGM. The SKA will enable the study of H I though the 21 cm line out to much larger distances than is currently possible, at high redshift it will enable intensity mapping of the large scale distribution of neutral hydrogen. The connection between neutral gas and galaxies was redefined during the epoch of reionisation. Reionization is becoming a major frontier at the interface between cosmology and galaxy evolution. The long awaited JWST will open up an unprecedented view of the early universe and reionisation, offering multiplexed spectroscopy of high redshift galaxies. Reaching back to a time when primitive galaxies ionised the previously neutral IGM, after which DLAs emerged as the dominant reservoirs of neutral gas.

In the course of this thesis we have studied two long standing questions in cosmology and galaxy evolution. Both areas have tremendous opportunities in datasets that will become available in the next two decades.

# Appendix A

# Chapter 4 Additional Material

## A.1 Additional tests on the robustness of the LAE sample

The significance of the detected LAEs is calculated from the rescaled variance of the pixels that enter the segmentation map. A possible concern is that, due to the fact that detector pixels are resampled in the final datacubes, we may be overestimating the SNR by neglecting correlation in the noise between neighbouring pixels, as commonly noted in imaging (Fumagalli et al., 2015). We test for this possibility in Fig. A.1, where we compare the flux distribution in the first and second neighbouring pixels to the theoretical value assuming uncorrelated noise. We also present the correlation matrix estimated from the datacube, computed from the standard deviation of pixels added to themselves and shifted along two of the axis of the datacube.

Both tests show that following the resampling procedure some degree of correlation is introduced in the data, but this effect is limited to the nearest neighbours and thus it is unlikely to alter our estimate of the SNR, which is believed to be accurate following the rescaling described in the Sect. 4.3.1. This result is in line with the analysis of MUSE data presented by Bacon et al. (2017).

We have also tested the completeness of our search to emission line objects. To this

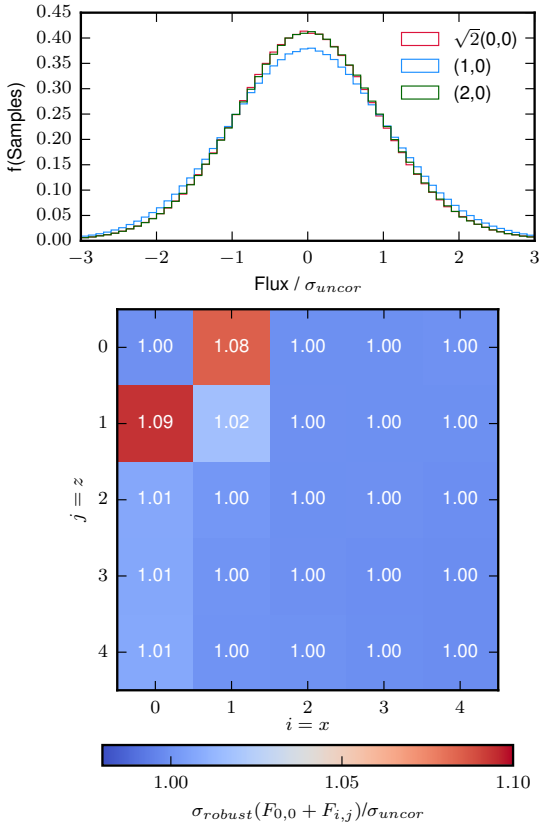


Figure A.1: An example of the correlated noise in field J2351+1600 at the wavelengths corresponding to that of the search window. *Top:* The distribution of fluxes in pairs of neighbouring pixels normalized by the expected uncorrelated noise, showing the low level of correlated noise in the datacube. The normalized flux distribution for pixels that are immediate neighbours in the $x$ direction are shown in blue (1,0), as are the pixels separated by an additional 1 pixel in green (2,0). The expectation given uncorrelated noise, i.e. the flux distribution of single pixels multiplied by $\sqrt{2}$, is also shown in red ($\sqrt{2}$(0,0)). *Lower:* The correlation matrix estimated from the datacube. The matrix shows the robust standard deviation obtained from the flux distribution from adding the cube to itself shifted in the x dimension by $i$ and the z direction by $j$. Each cell is coloured by the $\sigma$ and labelled with the value.

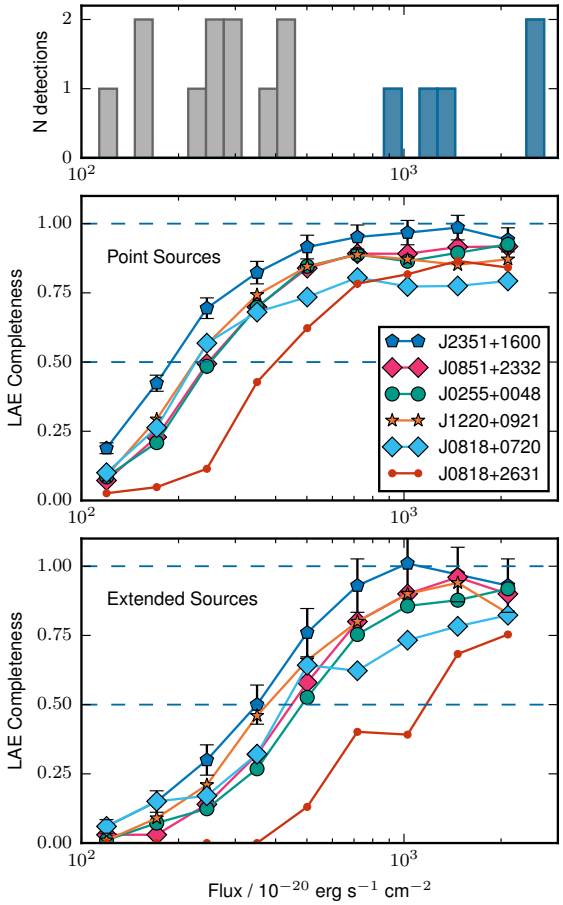


Figure A.2: LAE completeness as a function of flux. Errors are shown for field J2351+1600. **Middle:** the LAE completeness for different fields (black) shown against the Ly$\alpha$ flux, based on injecting point sources. 50 and 100 % completeness are indicated with blue dashed lines. In the case of field J2351+1600, the uncertainty in the completeness is also shown. **Bottom:** The same as the middle panel but now for the realistic extended mock sources. **Top:** The flux distribution of the high-confidence LAEs (blue) and lower significance candidates (grey).

end, we inject mock emission lines in the detection datacubes, assuming a Gaussian profile both in the spectral (with 2 channels full-width at half-maximum) and spatial directions (full-width at half-maximum of 3.5 pixels). This simulates unresolved point sources convolved with the seeing. We then run the CubExtractor with the same settings adopted in our main search, and count the number of recovered mock lines with SNR $\geq$ 7 and $N_{voxels} \geq 40$, spanning more than 3 wavelength channels. Ten mock lines are injected at a time to limit crowding effects and possible blends, and the process is repeated 50 times for each cube and in different flux bins. Additionally, we studied the effect of LAEs having extended profiles on the completeness. Following Drake et al. (2017), we injected sources drawn from Leclercq et al. (2017) which are rescaled in flux. This provides a more realistic estimate of the completeness, by using source parameters from a MUSE LAE sample. The sample presented in Leclercq et al. (2017) had their Ly$\alpha$ surface brightness profiles fitted with two exponential profiles, representing an extended halo and a component which follows the UV continuum of the galaxies. The continuum-like components are typically unresolved by MUSE, so they are injected as Gaussian profiles with a FWHM matching the seeing. The flux ratio between components and the spectral FWHM are applied to the mock sources. We repeated this for ten iterations with ten mock sources per run. We believe this is a more accurate estimate of the completeness.

The estimated completeness as a function of line flux is shown in Fig. A.2 for the 6 fields. This test shows that we are $\simeq 50\%$ complete for $L_{Ly\alpha} = 10^{41.5}$ erg s$^{-1}$ for point sources and $10^{41.8}$ erg s$^{-1}$ for realistic sources. This analysis, however, does not include the rejection of sources associated with the visual inspection, and is therefore a slightly optimistic estimate. From Fig. A.2, however, we conclude that the high-confidence LAEs (i.e. SNR = 12) do not suffer from incompleteness and form a highly-complete sample. The lower SNR candidates fall instead at lower completeness, although the exact value is uncertain due to the unknown effects introduced by the visual inspection.

## A.2 Notes on the fields with lower significance detections

In the following sections, we individually discuss the detected LAEs in each DLA field where lower SNR objects have been found. The results from fields J0851+2332, J0255+0048 and J1220+0921 (containing high-confidence associations) are described in Sec. 4.4. We also comment on candidate hosts identified in some fields from deep broadband imaging (Fumagalli et al., 2014) which we extract spectra for.

#### J2351+1600

Three low confidence candidates are detected in the field J2351+1600. These objects are offset in velocity from the DLA by 187.9, -32.3, -252.0 km $s^{-1}$ in order from highest SNR, to the lowest. These candidates have impact parameters ranging from 130 to 166 kpc, which is likely too large to be directly associated with the gas giving rise to the damped absorber. This DLA was noted in Fumagalli et al. (2015) as one of a 6 DLAs with candidate broadband hosts, however MUSE observations of this object revealed it to be a $z = 0.3886$ star-forming galaxy.

#### J0255+0048

Fumagalli et al. (2015) noted that in this field there is apparent emission at the position of the DLA, as measured in B band photometry. However, contamination from leakage of the quasar though the "blocking" Lyman Limit System could not be excluded. We do not detect Ly$\alpha$ emission at the position of the quasar, however if there were a continuum object without Ly$\alpha$ emission at such small separation we could not resolve it from the quasar. The LBG in this field, detected here by MUSE, was not detected in the B band imagery for this field, likely due to the higher seeing and to the fact that that the B band is probing shorter wavelengths than the Lyman break of this galaxy.

**J0818+0720**

No high confidence or candidate LAEs are detected in association to DLA J0818+0720. In Fumagalli et al. (2015) broadband host candidates to DLA J0818+0720 were detected in the WFC3/UVIS F390W imaging. These objects fall very close to the quasars ($0.7 \pm 0.1$ arcsec and $1.5 \pm 0.1$ arcsec), they appear to be compact, possibly stellar, in the WFC3 imaging. Even with the excellent seeing of the MUSE data these objects are too faint and too close to the quasar to have their spectra extracted in a clean way.

**J0818+2631**

Two LAE candidates are detected in proximity to DLA J0818+2631, but there are no high confidence detections in this field. Id91 is a promising candidate for a direct association with the DLA, as it has a velocity offset from the DLA of just +290.5 km s$^{-1}$ and an impact parameter of 7.0 arcsec (52.5 kpc). The other candidate LAE in this field, id87, has a similar position in velocity space (+231.5 km s$^{-1}$) but has a larger impact parameter (170.1 kpc). Due to only 2 of 6 exposures for this field being taken the data quality is poor, and some objects are rejected because they fell in regions where one of the two exposures is masked, preventing us from confirming the robustness of these detections.

The candidate host galaxy identified Fumagalli et al. (2015) for this DLA is very close to the quasar line of sight (2.05 arcsec), but with MUSE this galaxy was confirmed to be a low redshift interloper at $z = 0.554$, highlighting once more the need for spectroscopic searches for DLA host galaxies.

## A.3 Continuum Spectra of LBGs

In this Section, we report the spectra of galaxies with detected continuum counterparts at the depth of our MUSE observations.

For the galaxies in the fields J0255+0048 and J1220+0921, the masks used to extract the LBGs spectra are taken from the SExtractor segmentation masks as described above. In the case of the detection in field J0255+0048, the continuum counterpart is only $\simeq$ 2.5 arcsec from the quasar and hence, is not deblended by SExtractor. For this object, the mask to extract the spectrum is generated manually, using a circular aperture with a diameter of 1.2 arcsec. Due to its proximity to the quasar, the spectrum is contaminated, and requires careful sky subtraction. An annual sky aperture around the quasar, but excluding the continuum counterpart is used to extract a sky spectrum, which is then used to subtract the quasar contamination from the spectrum.

## A.4 Properties of the candidate LAEs

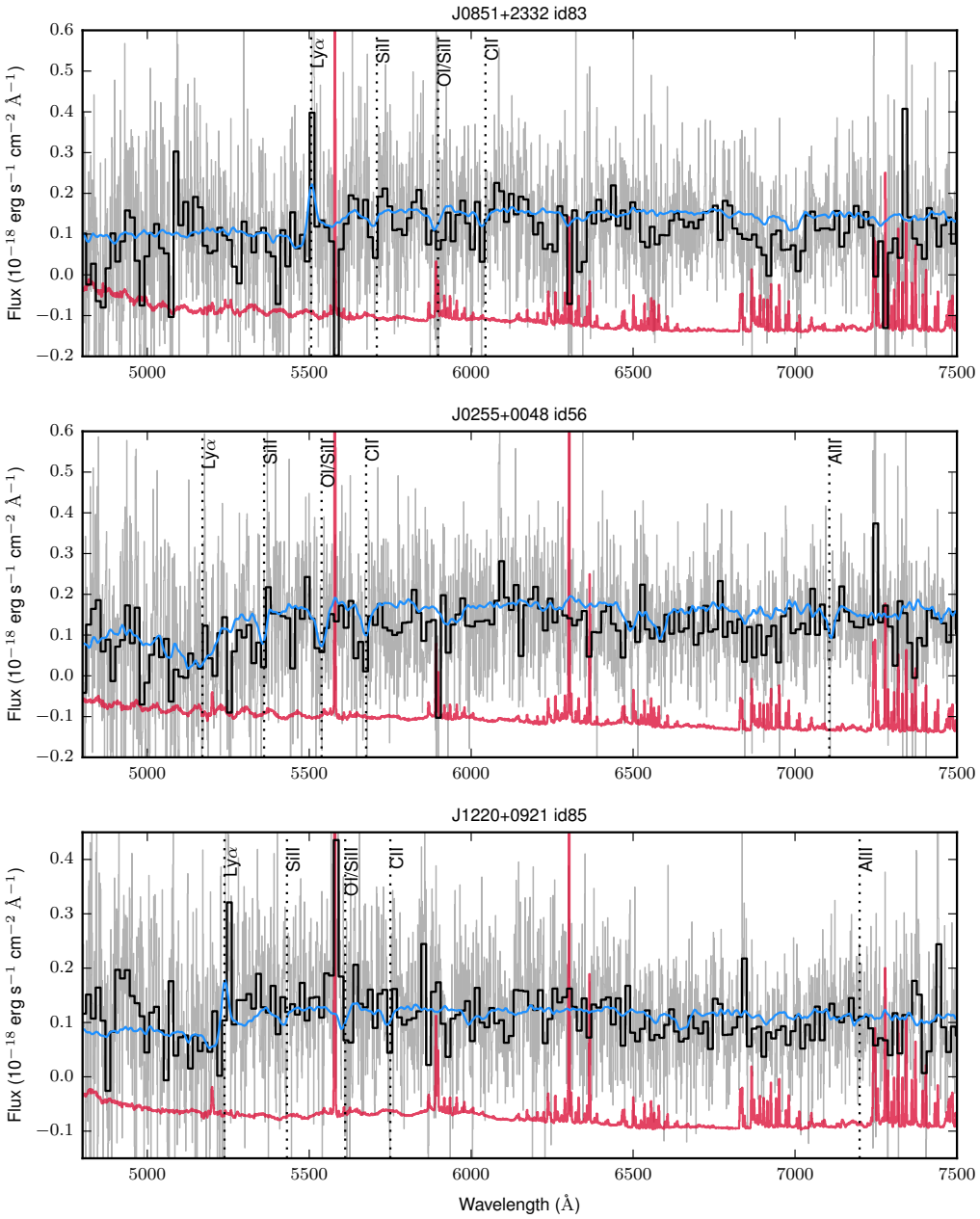


Figure A.3: MUSE spectra of the detected continuum counterparts to LAEs in DLA fields. In each panel the 1D MUSE spectrum integrated over the object is shown (grey) over-plotted with a re-binned version of the same spectrum to suppress the noise (black). The 1 $\sigma$ flux error in the unsmoothed spectrum is shown with shifted down for clarity, with zero at the bottom of each plot (red). With each object a template LBG spectrum is plotted (blue) redshifted to match the sources, the templates shown were selected to roughly match the presence of Ly$\alpha$ emission or absorption. The wavelengths of lines commonly observed in LBG spectra are also labelled (vertical dotted lines).

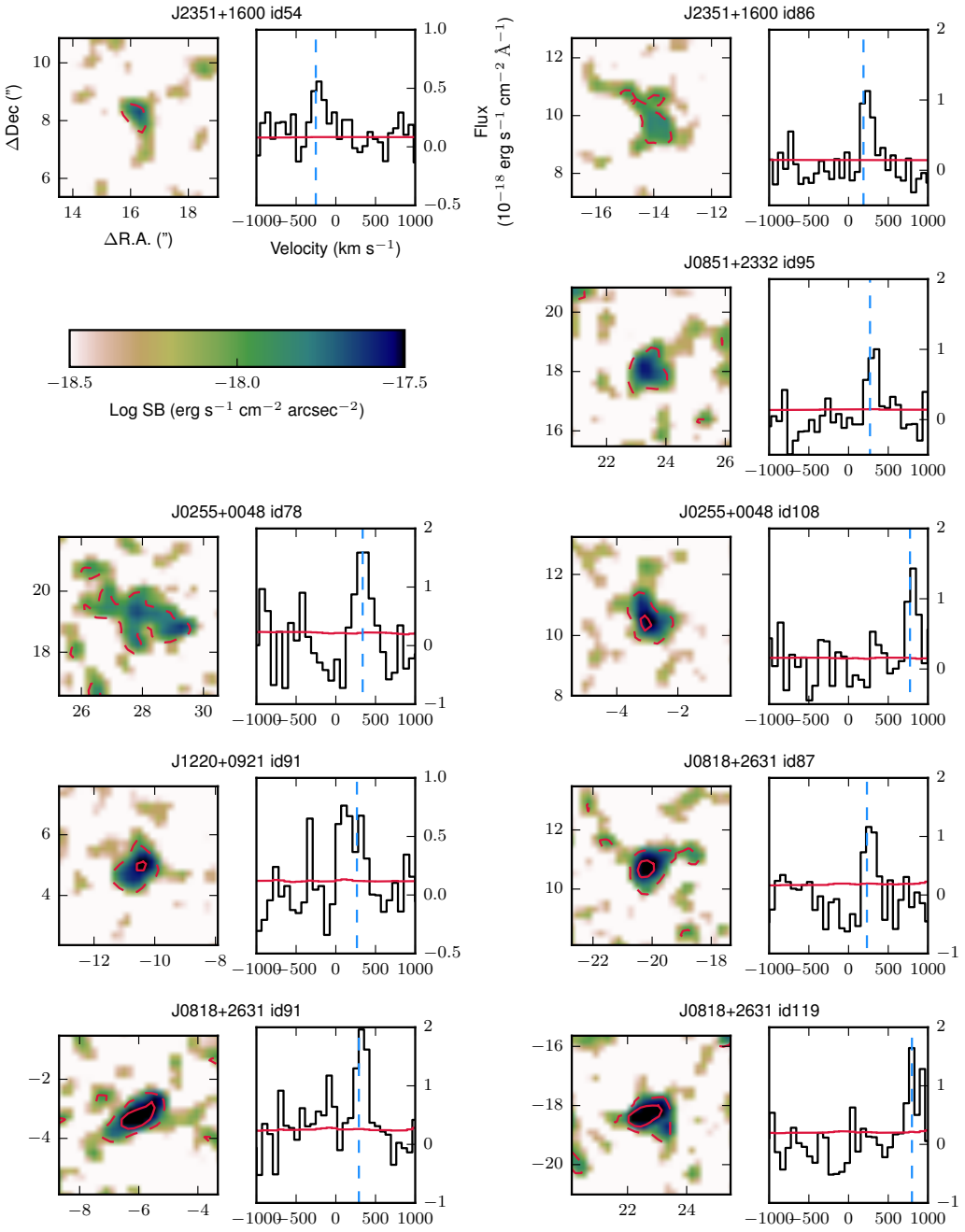


Figure A.4: Cutouts of the lower SNR LAE candidates. Each object has an optimally extracted Ly$\alpha$ surface brightness map (left) and an 1D spectrum of the detected emission line. The images have Ly$\alpha$ SB contours (red), drawn at at $10^{-18}$ (dashed) and $10^{-17.5}$ (solid) erg s$^{-1}$ cm$^{-2}$ arcsec$^{-2}$. The Ly$\alpha$ SB map has been smoothed with a Gaussian kernel ($\sigma$ = 1 pixel) and the axes are in physical kpc from the DLA. For the spectra the flux (black), 1$\sigma$ flux error (red) and line central velocity (vertical dashed blue line) and shown plotted as a velocity offset from the DLA redshift.

# Bibliography


Aarseth S. J., 1963, MNRAS, 126, 223

Alcock C., et al., 2000, ApJ, 542, 281

Altay G., Theuns T., 2013, MNRAS, 434, 748

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046

Artale M. C., et al., 2017, MNRAS, 470, 1771

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Astropy Collaboration et al., 2013, A&A, 558, A33

Babcock H. W., 1939, Lick Observatory Bulletin, 19, 41

Bacon R., et al., 2010, in Ground-based and Airborne Instrumentation for Astronomy III. p. 773508, doi:10.1117/12.856027

Bacon R., et al., 2017, A&A, 608, A1

Balbi A., et al., 2001, ApJ, 558, L145

Barnes L. A., Haehnelt M. G., 2009, MNRAS, 397, 511

Barnes L. A., Haehnelt M. G., 2014, MNRAS, 440, 2313

Barreira A., Li B., Baugh C. M., Pascoli S., 2014a, J. Cosmology Astropart. Phys., 8, 059

Barreira A., Li B., Baugh C. M., Pascoli S., 2014b, J. Cosmology Astropart. Phys., 8, 59

Beaver E. A., Burbidge E. M., McIlwain C. E., Epps H. W., Strittmatter P. A., 1972, ApJ, 178, 95

Beck R., Csabai I., Rácz G., Szapudi I., 2018, MNRAS, 479, 3582

Bennett C. L., et al., 2011, ApJS, 192, 17

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bielby R. M., et al., 2011, MNRAS, 414, 2

Bielby R., et al., 2013, MNRAS, 430, 425

Bielby R., Crighton N. H. M., Fumagalli M., Morris S. L., Stott J. P., Tejos N., Cantalupo S., 2017, MNRAS, 468, 1373

Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313

Bird S., Haehnelt M., Neeleman M., Genel S., Vogelsberger M., Hernquist L., 2015, MNRAS, 447, 1834

Birkinshaw M., Gull S. F., Hardebeck H., 1984, Nature, 309, 34

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Borisova E., et al., 2016, ApJ, 831, 39

Bremer M. N., Silk J., Davies L. J. M., Lehnert M. D., 2010, MNRAS, 404, L69

Bueno Sánchez J. C., 2014, Physics Letters B, 739, 269

Busswell G. S., Shanks T., Frith W. J., Outram P. J., Metcalfe N., Fong R., 2004, MNRAS, 354, 991

Cabré A., Gaztañaga E., Manera M., Fosalba P., Castander F., 2006, MNRAS, 372, L23

Cai Y.-C., Cole S., Jenkins A., Frenk C. S., 2010, MNRAS, 407, 201

Cai Y.-C., Li B., Cole S., Frenk C. S., Neyrinck M., 2014a, MNRAS, 439, 2978

Cai Y.-C., Neyrinck M. C., Szapudi I., Cole S., Frenk C. S., 2014b, ApJ, 786, 110

Cai Y.-C., Neyrinck M., Mao Q., Peacock J. A., Szapudi I., Berlind A. A., 2017a, MNRAS, 466, 3364

Cai Z., et al., 2017b, ApJ, 837, 71

Cannon R., et al., 2006, MNRAS, 372, 425

Cantalupo S., Lilly S. J., Haehnelt M. G., 2012, MNRAS, 425, 1992

Carroll S. M., 2001, Living Reviews in Relativity, 4, 1

Carswell R. F., Webb J. K., 2014, VPFIT: Voigt profile fitting program, Astrophysics Source Code Library (ascl:1408.015)

Carswell R. F., Hilliard R. L., Strittmatter P. A., Taylor D. J., Weymann R. J., 1975, ApJ, 196, 351

Chang S., Kleban M., Levi T. S., 2009, J. Cosmology Astropart. Phys., 4, 025

Charlot S., Fall S. M., 1993, ApJ, 415, 580

Christensen L., Wisotzki L., Roth M. M., Sánchez S. F., Kelz A., Jahnke K., 2007, A&A, 468, 587

Christensen L., Møller P., Fynbo J. P. U., Zafar T., 2014, MNRAS, 445, 225

Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, ApJ, 648, L109

Colbert J. W., Malkan M. A., 2002, ApJ, 566, 51

Cole S., Kaiser N., 1989, MNRAS, 237, 1127

Cole S., et al., 2005, MNRAS, 362, 505

Colless M., et al., 2001, MNRAS, 328, 1039

Conklin E. K., 1969, Nature, 222, 971

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

Cooke J., O'Meara J. M., 2015, ApJ, 812, L27

Cooke J., Wolfe A. M., Gawiser E., Prochaska J. X., 2006, ApJ, 636, L9

Cooke R. J., Pettini M., Steidel C. C., 2017, MNRAS, 467, 802

Copi C. J., Schramm D. N., Turner M. S., 1995, Science, 267, 192

Crain R. A., et al., 2015, MNRAS, 450, 1937

Crain R. A., et al., 2017, MNRAS, 464, 4204

Crittenden R. G., Turok N., 1996, Physical Review Letters, 76, 575

Croom S., Saunders W., Heald R., 2004, Anglo-Australian Observatory Epping Newsletter, 106, 12

Cruz M., Martínez-González E., Vielva P., Cayón L., 2005, MNRAS, 356, 29

Cruz M., Tucci M., Martínez-González E., Vielva P., 2006, MNRAS, 369, 57

Cruz M., Turok N., Vielva P., Martínez-González E., Hobson M., 2007, Science, 318, 1612

Cucciati O., et al., 2014, A&A, 570, A16

D'Odorico V., et al., 2016, MNRAS, 463, 2690

Danforth C. W., et al., 2014, preprint, (arXiv:1402.2655)

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

De Cia A., Ledoux C., Mattsson L., Petitjean P., Srianand R., Gavignaud I., Jenkins E. B., 2016, A&A, 596, A97

Dekel A., Silk J., 1986, ApJ, 303, 39

Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604

Diener C., et al., 2017, MNRAS, 471, 3186

Dijkstra M., Westra E., 2010, MNRAS, 401, 2343

Dobos L., Csabai I., Yip C.-W., Budavári T., Wild V., Szalay A. S., 2012, MNRAS, 420, 1217

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Drake A. B., et al., 2017, A&A, 608, A6

Driver S. P., et al., 2009, Astronomy and Geophysics, 50, 5.12

Driver S. P., et al., 2011, MNRAS, 413, 971

Efstathiou G., Sutherland W. J., Maddox S. J., 1990, Nature, 348, 705

Eisenstein D. J., et al., 2001, AJ, 122, 2267

Eisenstein D. J., et al., 2005, ApJ, 633, 560

Eisenstein D. J., et al., 2011, AJ, 142, 72

Erb D. K., et al., 2014, ApJ, 795, 33

Farrow D. J., et al., 2015, MNRAS, 454, 2120

Faucher-Giguère C.-A., Kereš D., Dijkstra M., Hernquist L., Zaldarriaga M., 2010, ApJ, 725, 633

Finelli F., García-Bellido J., Kovács A., Paci F., Szapudi I., 2016, MNRAS, 455, 1246

Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, ApJ, 473, 576

Foltz C. B., Chaffee Jr. F. H., Weymann R. J., 1986, AJ, 92, 247

Font-Ribera A., et al., 2012, J. Cosmology Astropart. Phys., 11, 059

Fosalba P., Gaztañaga E., Castander F. J., 2003, ApJ, 597, L89

Fox A. J., Ledoux C., Petitjean P., Srianand R., 2007, A&A, 473, 791

Francis C. L., Peacock J. A., 2010, MNRAS, 406, 14

Friedmann A., 1922, Zeitschrift fur Physik, 10, 377

Fumagalli M., O'Meara J. M., Prochaska J. X., Kanekar N., 2010, MNRAS, 408, 362

Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796

Fumagalli M., O'Meara J. M., Prochaska J. X., Kanekar N., Wolfe A. M., 2014, MNRAS, 444, 1282

Fumagalli M., O'Meara J. M., Prochaska J. X., Rafelski M., Kanekar N., 2015, MNRAS, 446, 3178

Fumagalli M., Cantalupo S., Dekel A., Morris S. L., O'Meara J. M., Prochaska J. X., Theuns T., 2016, MNRAS, 462, 1978

Fumagalli M., Haardt F., Theuns T., Morris S. L., Cantalupo S., Madau P., Fossati M., 2017a, MNRAS, 467, 4802

Fumagalli M., et al., 2017b, MNRAS, 471, 3686

Furlanetto S. R., Schaye J., Springel V., Hernquist L., 2005, ApJ, 622, 7

Furlong M., et al., 2015, MNRAS, 450, 4486

Furlong M., et al., 2017, MNRAS, 465, 722

Fynbo J. U., Møller P., Warren S. J., 1999, MNRAS, 305, 849

Fynbo J. P. U., Prochaska J. X., Sommer-Larsen J., Dessauges-Zavadsky M., Møller P., 2008, ApJ, 683, 321

Fynbo J. P. U., et al., 2010, MNRAS, 408, 2128

Fynbo J. P. U., et al., 2013, MNRAS, 436, 361

Garel T., Blaizot J., Guiderdoni B., Schaerer D., Verhamme A., Hayes M., 2012, MNRAS, 422, 310

Garel T., Blaizot J., Guiderdoni B., Michel-Dansac L., Hayes M., Verhamme A., 2015, MNRAS, 450, 1279

Garel T., Guiderdoni B., Blaizot J., 2016, MNRAS, 455, 3436

Gawiser E., et al., 2007, ApJ, 671, 278

Geller M. J., Huchra J. P., 1989, Science, 246, 897

Giannantonio T., Scranton R., Crittenden R. G., Nichol R. C., Boughn S. P., Myers A. D., Richards G. T., 2008, Phys. Rev. D, 77, 123520

Giannantonio T., Crittenden R., Nichol R., Ross A. J., 2012, MNRAS, 426, 2581

Granett B. R., Neyrinck M. C., Szapudi I., 2008, ApJ, 683, L99

Granett B. R., Szapudi I., Neyrinck M. C., 2010, ApJ, 714, 825

Gurzadyan V. G., Kashin A. L., Khachatryan H., Poghosian E., Sargsyan S., Yegorian G., 2014, A&A, 566, A135

Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. p. 64 (arXiv:astro-ph/0106018)

Haehnelt M. G., Steinmetz M., Rauch M., 1998, ApJ, 495, 647

Hawkins E., et al., 2003, MNRAS, 346, 78

Henry P. S., 1971, Nature, 231, 516

High F. W., Stubbs C. W., Rest A., Stalder B., Challis P., 2009, AJ, 138, 110

Ho S., Hirata C., Padmanabhan N., Seljak U., Bahcall N., 2008, Phys. Rev. D, 78, 043519

Hopkins A. M., et al., 2013, MNRAS, 430, 2047

Hotchkiss S., Nadathur S., Gottlöber S., Iliev I. T., Knebe A., Watson W. A., Yepes G., 2015, MNRAS, 446, 1321

Hubble E. P., 1925, The Observatory, 48, 139

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Inami H., et al., 2017, A&A, 608, A2

Inoue K. T., Silk J., 2006, ApJ, 648, 23

Jaffe T. R., Banday A. J., Eriksen H. K., Górski K. M., Hansen F. K., 2005, ApJ, 629, L1

Jorgenson R. A., Murphy M. T., Thompson R., 2013, MNRAS, 435, 482

Joshi R., Srianand R., Noterdaeme P., Petitjean P., 2017, MNRAS, 465, 701

Kaiser N., 1984, ApJ, 284, L9

Kanekar N., et al., 2018, ApJ, 856, L23

Kantowski R., Chen B., Dai X., 2015, Phys. Rev. D, 91, 083004

Kapteyn J. C., 1922, ApJ, 55, 302

Kashikawa N., Misawa T., Minowa Y., Okoshi K., Hattori T., Toshikawa J., Ishikawa S., Onoue M., 2014, ApJ, 780, 116

Kennicutt Jr. R. C., 1998, ApJ, 498, 541

Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349

Kim T.-S., Partl A. M., Carswell R. F., Müller V., 2013, A&A, 552, A77

Koribalski B., Staveley-Smith L., 2009.

Kovács A., 2018, MNRAS, 475, 1777

Kovács A., García-Bellido J., 2016, MNRAS, 462, 1882

Kovács A., Granett B. R., 2015, MNRAS, 452, 1295

Kovács A., et al., 2017, MNRAS, 465, 4166

Krogager J.-K., Møller P., Fynbo J. P. U., Noterdaeme P., 2017, MNRAS, 469, 2959

Krumholz M. R., Ellison S. L., Prochaska J. X., Tumlinson J., 2009, ApJ, 701, L12

Kuehn K., et al., 2014, in Ground-based and Airborne Instrumentation for Astronomy V. p. 914710, doi:10.1117/12.2055677

Lacey C. G., et al., 2016, MNRAS, 462, 3854

Lagarde N., Decressin T., Charbonnel C., Eggenberger P., Ekström S., Palacios A., 2012, A&A, 543, A108

Lagos C. d. P., et al., 2015, MNRAS, 452, 3815

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Lanzetta K. M., Wolfe A. M., Turnshek D. A., Lu L., McMahon R. G., Hazard C., 1991, ApJS, 77, 1

Lanzetta K. M., Wolfe A. M., Turnshek D. A., 1995, ApJ, 440, 435

Larjo K., Levi T. S., 2010, J. Cosmology Astropart. Phys., 8, 034

Le Fèvre O., et al., 2005, A&A, 439, 845

Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1

Leclercq F., et al., 2017, A&A, 608, A8

Ledoux C., Petitjean P., Fynbo J. P. U., Møller P., Srianand R., 2006, A&A, 457, 71

Leitherer C., et al., 1999, ApJS, 123, 3

Lemaître G., 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Liske J., et al., 2015, MNRAS, 452, 2087

López S., et al., 2016, A&A, 594, A91

Maiolino R., et al., 2008, A&A, 488, 463

Mannucci F., et al., 2009, MNRAS, 398, 1915

Marinoni C., et al., 2005, A&A, 442, 801

Mather J. C., et al., 1994, ApJ, 420, 439

McAlpine S., et al., 2016, Astronomy and Computing, 15, 72

Metcalfe N., Shanks T., Campos A., McCracken H. J., Fong R., 2001, MNRAS, 323, 795

Møller P., Warren S. J., 1993, A&A, 270, 43

Møller P., Warren S. J., Fall S. M., Fynbo J. U., Jakobsen P., 2002, ApJ, 574, 51

Møller P., Fynbo J. P. U., Fall S. M., 2004, A&A, 422, L33

Møller P., Fynbo J. P. U., Ledoux C., Nilsson K. K., 2013, MNRAS, 430, 2680

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903

Nadathur S., Crittenden R., 2016, ApJ, 830, L19

Nadathur S., Lavinto M., Hotchkiss S., Räsänen S., 2014, Phys. Rev. D, 90, 103510

Naidoo K., Benoit-Lévy A., Lahav O., 2016, MNRAS, 459, L71

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Neeleman M., Wolfe A. M., Prochaska J. X., Rafelski M., 2013, ApJ, 769, 54

Neeleman M., Prochaska J. X., Wolfe A. M., 2015, ApJ, 800, 7

Neeleman M., Kanekar N., Prochaska J. X., Rafelski M., Carilli C. L., Wolfe A. M., 2017, Science, 355, 1285

Norberg P., et al., 2002, MNRAS, 336, 907

Noterdaeme P., Ledoux C., Petitjean P., Le Petit F., Srianand R., Smette A., 2007, A&A, 474, 393

Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, A&A, 505, 1087

Noterdaeme P., et al., 2012, A&A, 547, L1

Noterdaeme P., Srianand R., Rahmani H., Petitjean P., Pâris I., Ledoux C., Gupta N., López S., 2015, A&A, 577, A24

O'Meara J. M., Chen H.-W., Kaplan D. L., 2006, ApJ, 642, L9

Oort J. H., 1932, Bull. Astron. Inst. Netherlands, 6, 249

Ostriker J. P., Peebles P. J. E., 1973, ApJ, 186, 467

Peacock J. A., et al., 2001, Nature, 410, 169

Pearson S., et al., 2016, MNRAS, 459, 1827

Peebles P. J. E., 1980, The large-scale structure of the universe

Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Pérez-Ràfols I., Miralda-Escudé J., Arinyo-i-Prats A., Mas-Ribas L., 2018a, preprint, (arXiv:1805.00943)

Pérez-Ràfols I., et al., 2018b, MNRAS, 473, 3019

Perlmutter S., et al., 1999, ApJ, 517, 565

Peroux C., 2002, in Combes F., Barret D., eds, SF2A-2002: Semaine de l'Astrophysique Francaise. p. 359 (arXiv:astro-ph/0209463)

Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2012, MNRAS, 419, 3060

Péroux C., et al., 2017, MNRAS, 464, 2053

Pezzotta A., et al., 2017, A&A, 604, A33

Planck Collaboration et al., 2014a, A&A, 571, A1

Planck Collaboration et al., 2014b, A&A, 571, A11

Planck Collaboration et al., 2014c, A&A, 571, A16

Planck Collaboration et al., 2014d, A&A, 571, A19

Planck Collaboration et al., 2014e, A&A, 571, A23

Planck Collaboration et al., 2014f, A&A, 571, A27

Planck Collaboration et al., 2016a, A&A, 594, A1

Planck Collaboration et al., 2016b, A&A, 594, A13

Planck Collaboration et al., 2016c, A&A, 594, A16

Planck Collaboration et al., 2016d, A&A, 594, A21

Pontzen A., et al., 2008, MNRAS, 390, 1349

Porciani C., Norberg P., 2006, MNRAS, 371, 1824

Prochaska J. X., Herbert-Fort S., 2004, PASP, 116, 622

Prochaska J. X., Wolfe A. M., 1998, ApJ, 507, 113

Prochaska J. X., Wolfe A. M., 2009, ApJ, 696, 1543

Prochaska J. X., et al., 2001, ApJS, 137, 21

Prochaska J. X., Gawiser E., Wolfe A. M., Castro S., Djorgovski S. G., 2003, ApJ, 595, L9

Prochaska J. X., O'Meara J. M., Worseck G., 2010, ApJ, 718, 392

Rafelski M., Wolfe A. M., Chen H.-W., 2011, ApJ, 736, 48

Rafelski M., Wolfe A. M., Prochaska J. X., Neeleman M., Mendez A. J., 2012, ApJ, 755, 89

Rafelski M., Neeleman M., Fumagalli M., Wolfe A. M., Prochaska J. X., 2014, ApJ, 782, L29

Rafelski M., Gardner J. P., Fumagalli M., Neeleman M., Teplitz H. I., Grogin N., Koekemoer A. M., Scarlata C., 2016, ApJ, 825, 87

Rahmati A., Schaye J., 2014, MNRAS, 438, 529

Rahmati A., Schaye J., Pawlik A. H., Raičević M., 2013, MNRAS, 431, 2261

Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, MNRAS, 452, 2034

Rakic O., Schaye J., Steidel C. C., Rudie G. C., 2011, MNRAS, 414, 3265

Rauch M., et al., 2008, ApJ, 681, 856

Rauch M., Becker G. D., Haehnelt M. G., Gauthier J.-R., Sargent W. L. W., 2013, MNRAS, 429, 429

Rees M. J., Sciama D. W., 1968, Nature, 217, 511

Ribeiro B., et al., 2017, A&A, 608, A16

Riess A. G., et al., 1998, AJ, 116, 1009

Riess A. G., et al., 2016, ApJ, 826, 56

Roberts M. S., Whitehurst R. N., 1975, ApJ, 201, 327

Rogstad D. H., Shostak G. S., 1972, ApJ, 176, 315

Rosdahl J., Blaizot J., 2012, MNRAS, 423, 344

Ross N. P., Shanks T., Cannon R. D., Wake D. A., Sharp R. G., Croom S. M., Peacock J. A., 2008, MNRAS, 387, 1323

Rubin V. C., Ford Jr. W. K., Thonnard N., 1980, ApJ, 238, 471

Rudie G. C., et al., 2012, ApJ, 750, 67

Rudnick L., Brown S., Williams L. R., 2007, ApJ, 671, 40

Sachs R. K., Wolfe A. M., 1967, ApJ, 147, 73

Salmon B., et al., 2015, ApJ, 799, 183

Sánchez-Ramírez R., et al., 2016, MNRAS, 456, 4488

Sawangwit U., Shanks T., Cannon R. D., Croom S. M., Ross N. P., Wake D. A., 2010a, MNRAS, 402, 2228

Sawangwit U., Shanks T., Cannon R. D., Croom S. M., Ross N. P., Wake D. A., 2010b, MNRAS, 402, 2228

Schaerer D., Hayes M., Verhamme A., Teyssier R., 2011, A&A, 531, A12

Schaller M., Dalla Vecchia C., Schaye J., Bower R. G., Theuns T., Crain R. A., Furlong M., McCarthy I. G., 2015, MNRAS, 454, 2277

Schaye J., Aguirre A., Kim T.-S., Theuns T., Rauch M., Sargent W. L. W., 2003, ApJ, 596, 768

Schaye J., et al., 2010, MNRAS, 402, 1536

Schaye J., et al., 2015, MNRAS, 446, 521

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Shanks T., et al., 2015, MNRAS, 451, 4238

Sharp R., Parkinson H., 2010, MNRAS, 408, 2495

Shectman S. A., Landy S. D., Oemler A., Tucker D. L., Lin H., Kirshner R. P., Schechter P. L., 1996, ApJ, 470, 172

Sheth R. K., Tormen G., 1999, MNRAS, 308, 119

Silk J., 2011, in Carignan C., Combes F., Freeman K. C., eds, IAU Symposium Vol. 277, Tracing the Ancestry of Galaxies. pp 273–281 (`arXiv:1102.0283`), doi:10.1017/S1743921311022939

Simcoe R. A., Sargent W. L. W., Rauch M., 2004, ApJ, 606, 92

Slipher V. M., 1913, Lowell Observatory Bulletin, 2, 56

Slipher V. M., 1915, Popular Astronomy, 23, 21

Smith K. M., Huterer D., 2010, MNRAS, 403, 2

Smith H. E., Cohen R. D., Burns J. E., Moore D. J., Uchida B. A., 1989, ApJ, 347, 87

Smith K. M., Zahn O., Doré O., 2007, Phys. Rev. D, 76, 043510

Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, MNRAS, 470, 4646

Smoot G. F., et al., 1992, ApJ, 396, L1

Sobral D., Matthee J., Darvish B., Schaerer D., Mobasher B., Röttgering H. J. A., Santos S., Hemmati S., 2015, ApJ, 808, 139

Sotiriou T. P., Faraoni V., 2010, Reviews of Modern Physics, 82, 451

Spergel D. N., et al., 2007, ApJS, 170, 377

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

Springel V., Yoshida N., White S. D. M., 2001a, New Astron., 6, 79

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001b, MNRAS, 328, 726

Springel V., et al., 2005, Nature, 435, 629

Srianand R., Hussain T., Noterdaeme P., Petitjean P., Krühler T., Japelj J., Pâris I., Kashikawa N., 2016, MNRAS, 460, 634

Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17

Steidel C. C., Adelberger K. L., Shapley A. E., Pettini M., Dickinson M., Giavalisco M., 2003, ApJ, 592, 728

Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289

Steidel C. C., Bogosavljević M., Shapley A. E., Kollmeier J. A., Reddy N. A., Erb D. K., Pettini M., 2011, ApJ, 736, 160

Steinhardt P. J., 1983, in Gibbons G. W., Hawking S. W., Siklos S. T. C., eds, Very Early Universe. pp 251–266

Stierwalt S., Besla G., Patton D., Johnson K., Kallivayalil N., Putman M., Privon G., Ross G., 2015, ApJ, 805, 2

Szapudi I., et al., 2015, MNRAS, 450, 288

Telfer R. C., Zheng W., Kriss G. A., Davidsen A. F., 2002, ApJ, 565, 773

Trayford J. W., et al., 2017, MNRAS, 470, 771

Tremonti C. A., et al., 2004, ApJ, 613, 898

Tsujikawa S., 2013, Classical and Quantum Gravity, 30, 214003

Vanden Berk D. E., et al., 2001, AJ, 122, 549

Verhamme A., Schaerer D., Maselli A., 2006, A&A, 460, 397

Viegas S. M., 1995, MNRAS, 276, 268

Vielva P., 2010, Advances in Astronomy, 2010, 592094

Vielva P., Martínez-González E., Barreiro R. B., Sanz J. L., Cayón L., 2004, ApJ, 609, 22

Vilenkin A., 1983, Phys. Rev. D, 27, 2848

Warren S. J., Møller P., 1996, A&A, 311, 25

Warren S. J., Møller P., Fall S. M., Jakobsen P., 2001, MNRAS, 326, 759

Weatherley S. J., Warren S. J., Møller P., Fall S. M., Fynbo J. U., Croom S. M., 2005, MNRAS, 358, 985

Weilbacher P. M., Streicher O., Urrutia T., Pécontal-Rousset A., Jarno A., Bacon R., 2014, in Manset N., Forshay P., eds, Astronomical Society of the Pacific Conference Series Vol. 485, Astronomical Data Analysis Software and Systems XXIII. p. 451 (arXiv:1507.00034)

Weinberg S., 1989, Reviews of Modern Physics, 61, 1

Whitaker K. E., et al., 2014, ApJ, 795, 104

Whitbourn J. R., Shanks T., 2014, MNRAS, 437, 2146

Whitbourn J. R., Shanks T., 2016, MNRAS, 459, 496

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Wiersma R. P. C., Schaye J., Smith B. D., 2009, MNRAS, 393, 99

Wisotzki L., et al., 2016, A&A, 587, A98

Wolfe A. M., 1986, Philosophical Transactions of the Royal Society of London Series A, 320, 503

Wolfe A. M., Chen H.-W., 2006, ApJ, 652, 981

Wolfe A. M., Davis M. M., 1979, AJ, 84, 699

Wolfe A. M., Turnshek D. A., Smith H. E., Cohen R. D., 1986, ApJS, 61, 249

Wolfe A. M., Fan X.-M., Tytler D., Vogt S. S., Keane M. J., Lanzetta K. M., 1994, ApJ, 435, L101

Wolfe A. M., Prochaska J. X., Gawiser E., 2003, ApJ, 593, 215

Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861

Wright E. L., et al., 1991, ApJ, 381, 200

York D. G., et al., 2000, AJ, 120, 1579

Zafar T., et al., 2011, A&A, 532, A51

Zafar T., Péroux C., Popping A., Milliard B., Deharveng J.-M., Frank S., 2013, A&A, 556, A141

Zehavi I., et al., 2011, ApJ, 736, 59

Zhang R., Huterer D., 2010, Astroparticle Physics, 33, 69

Zhao W., 2013, MNRAS, 433, 3498

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

de Jong R. S., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460T (`arXiv:1206.6885`), doi:10.1117/12.926239

van de Voort F., Schaye J., Altay G., Theuns T., 2012, MNRAS, 421, 2809