The abundance and physical properties of O VII and O VIII X-ray absorption systems in the EAGLE simulations

Nastasha A. Wijers [1]⋆ Joop Schaye [1], Benjamin D. Oppenheimer,[2,3]
Robert A. Crain [4] and Fabrizio Nicastro[3,5]

[1]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[2]*CASA, Department of Astrophysical and Planetary Sciences, University of Colorado, 389 UCB, Boulder, CO 80309, USA*
[3]*Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA*
[4]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[5]*INAF Osservatorio Astronomico di Roma, Via Frascati 30, I-00078, Monte Porzio Catone (RM), Italy*

ABSTRACT

We use the EAGLE cosmological, hydrodynamical simulations to predict the column density and equivalent width distributions of intergalactic O VII ($E = 574$ eV) and O VIII ($E = 654$ eV) absorbers at low redshift. These two ions are predicted to account for 40 per cent of the gas-phase oxygen, which implies that they are key tracers of cosmic metals. We find that their column density distributions evolve little at observable column densities from redshift 1 to 0, and that they are sensitive to active galactic nucleus feedback, which strongly reduces the number of strong (column density $N \gtrsim 10^{16}$ cm^{-2}) absorbers. The distributions have a break at $N \sim 10^{16}$ cm^{-2}, corresponding to overdensities of $\sim 10^2$, likely caused by the transition from sheet/filament to halo gas. Absorption systems with $N \gtrsim 10^{16}$ cm^{-2} are dominated by collisionally ionized O VII and O VIII, while the ionization state of oxygen at lower column densities is also influenced by photoionization. At these high column densities, O VII and O VIII arising in the same structures probe systematically different gas temperatures, meaning their line ratio does not translate into a simple estimate of temperature. While O VII and O VIII column densities and covering fractions correlate poorly with the H I column density at $N_{\rm HI} \gtrsim 10^{15}$ cm^{-2}, O VII and O VIII column densities are higher in this regime than at the more common, lower H I column densities. The column densities of O VI and especially Ne VIII, which have strong absorption lines in the UV, are good predictors of the strengths of O VII and O VIII absorption and can hence aid in the detection of the X-ray lines.

Key words: galaxies: evolution – galaxies: haloes – intergalactic medium – quasars: absorption lines.

1 INTRODUCTION

Within extragalactic astronomy, the missing baryon problem is well-established. We know from cosmic microwave background measurements (e.g. Planck Collaboration I 2014) and big bang nucleosynthesis (see the review by Cyburt et al. 2016) how many baryons there were in the very early Universe. However, at low redshift, we cannot account for all of these baryons by adding up observed populations: stars, gas observed in the interstellar, circumgalactic, and intra-cluster media, and photoionized absorbers in the intergalactic medium (IGM). According to a census presented by Shull, Smith & Danforth (2012), about 30 per cent of baryons are observationally missing at low redshift.

Cosmological simulations predict that missing baryons reside in shock-heated regions of the IGM. Cen & Ostriker (1999) already found that the missing gas is mainly hot and diffuse. They called this gas the warm-hot intergalactic medium (WHIM), defined as gas with temperatures in the range 10^5–10^7 K. Some of this gas has already been detected. The cooler component is traced largely by O VI absorption: $T \sim 10^{5.5}$ K in collisional ionization equilibrium (CIE), which applies to high-density gas. O VI absorption has been studied observationally by many groups (e.g. Tumlinson et al. 2011; Johnson, Chen & Mulchaey 2013), often using the Cosmic Origins Spectrograph on the *Hubble Space Telescope* (*HST*). Others have investigated O VI absorption in galaxy and cosmological simulations (e.g. Cen & Fang 2006; Tepper-García et al. 2011; Cen 2012; Shull et al. 2012; Oppenheimer et al. 2016, 2018; Rahmati et al. 2016; Nelson et al. 2018). Predictions for the hotter part of the WHIM have been made with simulations (e.g. Cen & Fang 2006; Branchini

⋆ E-mail: wijers@strw.leidenuniv.nl

et al. 2009; Bertone et al. 2010a; Cen 2012; Nelson et al. 2018), typically, but not exclusively, focussing on the O VII and O VIII X-ray lines ($T \sim 10^{5.4}$–$10^{6.5}$ and $\sim 10^{6.1}$–$10^{6.8}$ K in CIE, respectively). Others have made predictions for the WHIM gas using analytical models (Perna & Loeb 1998), or using a combination of analytical models and simulations (Fang, Bryan & Canizares 2002; Furlanetto, Phillips & Kamionkowski 2005).

Besides being a major baryon reservoir, the WHIM also provides an important way to understand accretion and feedback processes in galaxy formation. As gas collapses onto a galaxy, some of it forms stars or is accreted by the central supermassive black hole. This creates feedback, where these stars (through, for example, supernova explosions) and active galactic nuclei (AGN) inject energy and momentum into their surrounding gas, slowing, stopping, or preventing further gas accretion. Current-generation cosmological simulations like EAGLE (Schaye et al. 2015), IllustrisTNG (Pillepich et al. 2018), and Horizon-AGN (Dubois et al. 2014) include star formation and AGN feedback, and metal enrichment from stars. The feedback in EAGLE and IllustrisTNG, and the AGN feedback in Horizon-AGN, is calibrated to reproduce galaxy (stellar and black hole) properties, because individual stars, supernova explosions, and black hole accretion discs are too small to be resolved in these simulations. That means that these simulations require a model of the effect of this feedback on scales they can resolve.

The effect of this feedback on the circumgalactic medium (CGM) is not as well-constrained. For example, feedback in BAHAMAS (McCarthy et al. 2017) and IllustrisTNG is calibrated to halo gas fractions, but this only applies to high-mass haloes where observations are available ($M_{500c} > 10^{13}$ M$_\odot$). One way to constrain the effects of feedback on the CGM is to study how far metals, which are created in galaxies, spread outside their haloes. In other words, we can investigate what fraction of the metals ends up in the IGM, which impacts how many metal absorbers we expect to find in quasar sightlines. This is not the only possible effect, as feedback also impacts the temperature, density, and kinematics of gas.

The main way we expect to find the hot WHIM gas is through absorption in e.g. quasar spectra (e.g. Brenneman et al. 2016; Nicastro et al. 2018; Kovács et al. 2019). This is because this WHIM gas typically has low densities: since emission scales with the density squared and absorption with the density along the line of sight, this makes absorption more readily detectable than emission in most of the WHIM. Observationally, absorption from these ions has been found around the Milky Way, as described by e.g. Bregman (2007), but claims of extragalactic O VII and O VIII are rare (e.g. Nicastro et al. 2005) and often disputed (e.g. Kaastra et al. 2006) or uncertain (e.g. Bonamente et al. 2016). Nicastro et al. (2017) review these WHIM searches in absorption. Nicastro et al. (2018) recently found two extragalactic O VII absorbers, using very long observations with the *XMM–Newton* RGS of the spectrum of the brightest X-ray blazar. These are consistent with current predictions, though the small number of absorbers means uncertainties on the total absorber budget are still large.

Besides these oxygen ions, other ions are also useful for studying the WHIM. For example, Ne VIII traces WHIM gas that is somewhat hotter than O VI traces ($T \sim 10^{5.8}$ K in CIE). Theoretically, Tepper-García, Richter & Schaye (2013) investigated its absorption in the OWLS cosmological, hydrodynamical simulations (Schaye et al. 2010) and Rahmati et al. (2016) did so with EAGLE. Observationally, e.g. Meiring et al. (2013) studied Ne VIII absorption, and recently, Burchett et al. (2019) found Ne VIII absorption associated with the CGM. A different tracer of the WHIM is broad Ly α absorption ($T \gtrsim 10^5$ K), which traces the cooler component of the

WHIM, similar to O VI but without the need for metals (e.g. Shull et al. 2012; Tepper-García et al. 2012). These ions are useful because they trace the WHIM themselves, but they can also be used to find O VII and O VIII absorbers. Recently, Kovács et al. (2019) did this: they found extragalactic O VII absorption by stacking quasar spectra using the known redshifts of 17 Ly α absorbers associated with massive galaxies.

Mroczkowski et al. (2019) give an overview of how the Sunyaev–Zeldovich (SZ) effect has been used to search for the WHIM. This effect probes line-of-sight integrated pressure (thermal SZ) or free electron bulk motion (kinetic SZ), which both have the same density dependence as absorption and are therefore less biased towards high-density gas than emission is. Attempts to detect the IGM by this method have focused on SZ measurements of cluster pairs to detect filaments between them, and cross-correlations with other tracers of large-scale structure (Mroczkowski et al. 2019 and references therein). For example, de Graaff et al. (2019) used stacked massive galaxy pairs (mean stellar mass $10^{11.3}$ M$_\odot$, virial mass $\sim 10^{13}$ M$_\odot$ h^{-1}) to detect a filamentary thermal SZ signal. Tanimura et al. (2019) similarly stacked luminous red galaxy pairs with stellar masses $> 10^{11.3}$ M$_\odot$, and found a filamentary thermal SZ signal larger than, but consistent with, predictions from the BAHAMAS simulations (McCarthy et al. 2017). Using the EAGLE simulations, Davies et al. (2019) made theoretical predictions for the thermal SZ effect from the CGM at different halo masses.

Another way to find hotter gas is through X-ray emission, but this scales with the density squared, and is therefore generally best for studying dense gas. Bregman (2007) reviews some X-ray observations of extragalactic gas, though this is limited to the denser parts of the intracluster medium (ICM) and CGM. Following earlier work on WHIM X-ray emission in simulations by e.g. Pierre, Bryan & Gastaud (2000) and Yoshikawa et al. (2003), Bertone et al. (2010a) studied soft X-ray line emission, and found that this emission should indeed be tracing mostly denser and metal-rich gas, i.e. ICM and CGM. Tumlinson, Peeples & Werk (2017) discuss some more recent results on X-ray line emission from the CGM. Instead of lines, Davies et al. (2019) investigated broad-band soft X-ray emission from the CGM in EAGLE, and found that it can be used as a proxy for the CGM gas fraction.

One way to search for the missing baryons in absorption is by doing blind surveys, where observers look for absorbers in the lines of sight to suitable background sources. There are a number of upcoming and proposed X-ray telescopes for which there are plans to carry out such surveys, including Athena (Barret et al. 2016), Arcus (Brenneman et al. 2016; Smith et al. 2016), and Lynx (The Lynx Team 2018). The O VIII and especially O VII ions tend to be the focus of such plans, since oxygen is a relatively abundant element (Allende Prieto, Lambert & Asplund 2001) and these ions have lines with large oscillator strengths (Verner, Verner & Ferland 1996), making these lines more readily detectable in the hot WHIM.

In simulations, O VII and O VIII absorption has been studied, e.g. Branchini et al. (2009) and Cen & Fang (2006) predicted the equivalent width (EW) distributions for blind surveys using an earlier generation of simulations and mock spectra generated from these. More recently, Nelson et al. (2018) studied absorption by these ions in the IllustrisTNG simulations, but they did not study EWs.

Here, we will study the column density and EW distributions of O VII (21.60 Å) and O VIII (18.967, 18.973 Å) in the EAGLE simulations (Crain et al. 2015; Schaye et al. 2015), to predict which absorption systems may be detected by future missions, and to establish the physical conditions these absorption systems probe.

EAGLE has been used to predict column density distributions that agree reasonably well with observations at a range of redshifts, for a variety of ions (e.g. Schaye et al. 2015; Rahmati et al. 2016). However, these studies all focused on ions with ionization energies lower than O VII (739 eV) and O VIII (871 eV) (Lide 2003).

This paper is organized as follows. In Section 2, we discuss our methods: the EAGLE simulations themselves (Section 2.1), and how we extract column densities (Section 2.2) and EWs (Section 2.3) from them. We present the column density distributions we find in Section 3.1, and discuss our mock spectra (Section 3.2) and the EW distributions we infer from these (Section 3.3). We discuss the origin of the shape of the column density distribution (Section 3.4) and how it probes AGN feedback (Section 3.5). We discuss the physical properties of O VII and O VIII absorption systems in Section 3.6. In Section 3.7, we discuss how absorption by these two ions correlates with three ions with UV lines: H I, O VI, and Ne VIII. In Section 3.8, we compare the gas traced by different ions along the same sightlines. In Section 4, we outline what our results predict for three planned or proposed missions in Section 4.1, and what some limitations of our work may be (Section 4.2). Finally, we summarize our results in Section 5.

2 METHODS

We study predictions for O VII and O VIII absorption in the EAGLE simulations (Crain et al. 2015; Schaye et al. 2015). We use tabulated ion fractions as a function of temperature, density, and redshift, as well as the element abundances from the simulation, to calculate the number density of ions at different positions in the simulation. By projecting the number of ions in thick slices through these simulations onto two-dimensional column density maps, we obtain column density distributions from the simulations. To calculate EWs for some of these columns, we generated synthetic absorption spectra at sightlines through their centres.

Oxygen abundances in this paper are given in solar units. Here, the solar oxygen mass fraction is 0.005 86 (Allende Prieto et al. 2001). This is simply used as a unit, and should not be updated to more recent solar abundance measurements in further work. Length units include p ('proper') or c ('comoving'). The exception is the centimetre we use in (column) densities: centimetres are always proper units.

2.1 The EAGLE simulations

In this section, we provide a short summary of the EAGLE (Evolution and Assembly of GaLaxies and their Environments) simulations. More details can be found in Schaye et al. (2015), the paper presenting the simulations calibrated to observations, Crain et al. (2015), which describes the calibration of these simulations, and McAlpine et al. (2016), which describes the data release of the EAGLE galaxy and halo data.

The code used is a modified version of GADGET3, last described by Springel (2005), using the TREE-PM gravitational force calculation. The modifications include an implementation of smoothed particle hydrodynamics (SPH) known as ANARCHY (Schaye et al. 2015, appendix A; Schaller et al. 2015). A Λ cold dark matter cosmogony is used with the parameters $(\Omega_m, \Omega_\Lambda, \Omega_b, h, \sigma_8, n_s, Y) =$ (0.307, 0.693, 0.04825, 0.6777, 0.8288, 0.9611, 0.248) (Planck Collaboration I 2014).

Gas cooling is implemented following Wiersma, Schaye & Smith (2009a), using the tracked abundances of 11 elements and

cooling rates for each. Collisional and photoionization equilibrium is assumed, with photoionization from the Haardt & Madau (2001) UV/X-ray background model. With those assumptions, and the element abundances as tracked in the simulation, we calculate the number of ions in a column through the simulated box. To do this, we use ion fraction tables from Bertone et al. (2010a,b), who investigated line emission with tables computed under the same assumptions as the EAGLE gas cooling. The ion fractions were computed using Cloudy, version c07.02.00[1] (Ferland et al. 1998). These tables give the ion fraction (fraction of nuclei of a given element that are part of a particular ion) as a function of $\log_{10}$ hydrogen number density, $\log_{10}$ temperature, and redshift. We interpolated these tables linearly, in log space for temperature and density, to obtain ion balances for each SPH particle.

These assumptions mean we ignore both local ionization sources, such as stars and AGNs, and non-equilibrium ionization. This also means we ignore the effect of flickering AGNs. These could boost the abundances of highly ionized species, such as those we are interested in, even if the AGN is not 'active' when we observe it (Oppenheimer & Schaye 2013; Segers et al. 2017; Oppenheimer et al. 2018). This boost is caused by species being ionized by the AGN radiation, and then not having time to recombine before observations are made. For typical AGN duty cycles and ion recombination times, the ions can be out of ionization equilibrium like this for a large fraction of systems. We discuss the impact of these assumptions in more detail in Section 4.2.

Since the resolution of the simulations is too low to resolve individual events of star formation and feedback (stellar winds and supernovae), or accretion discs in AGNs, the star formation and stellar and AGN feedback on simulated scales are implemented using subgrid models, with model feedback parameters calibrated to reproduce the $z = 0.1$ galaxy stellar mass function, the relation between black hole mass and galaxy mass, and reasonable galaxy disc sizes. Star formation occurs where the local gas density is high enough, with an additional metallicity dependence following Schaye (2004). The rate itself depends on the local pressure, in a way that reproduces the Kennicutt–Schmidt relation (Schaye & Dalla Vecchia 2008). Stars lose mass to surrounding gas particles as they evolve, enriching them with metals. This is modelled according to Wiersma et al. (2009b).

A major problem in galaxy formation simulations at EAGLE-like resolution is that if reasonable amounts of stellar and AGN feedback energy are injected into gas surrounding a single-age population of stars or a black hole at each time-step, the energy is radiated away before it can do any work. This causes gas in galaxies to form too many stars. Following Booth & Schaye (2009) and Dalla Vecchia & Schaye (2012), the solution for this problem used in EAGLE is to statistically 'save up' the energy released by these particles, until it is enough to heat neighbouring particles by a fixed temperature increment of $10^{7.5}$ K (stellar feedback) or $10^{8.5}$ K (AGN feedback). Which particles are heated and when is determined stochastically. The expectation value depends on the local gas density and metallicity (Crain et al. 2015).

Table 1 gives an overview of the different EAGLE simulations we use in this work. The reference feedback model was calibrated for the standard EAGLE resolution, as used in e.g. Ref-L100N1504. Simulation Recal-L025N0752 was calibrated in the same way as the reference model, but at an eight times higher mass resolution. This is used to test the 'weak convergence' of the simulations,

---

[1]This older version is consistent with the EAGLE cooling rate calculations.

Table 1. The simulations used in this work. The names consist of three parts, in the format `<name>-L<size>N<particles>`. The name is the name or abbreviation for the stellar and AGN feedback model, as explained in the text. The size is the size of the simulation box in comoving Mpc, and the last part is the cube root of the number of dark matter particles (equal to the initial number of gas particles) used in the simulation. The table lists the dark matter particle mass (m_{DM}), the initial gas particle mass $m_{gas,init}$, and the Plummer-equivalent gravitational softening length at low redshift (l_{soft}).

Simulation name	m_{DM} (M$_\odot$)	$m_{gas,init}$ (M$_\odot$)	l_{soft} (pkpc)
Ref-L100N1504	9.70×10^6	1.81×10^6	0.70
Ref-L050N0752	9.70×10^6	1.81×10^6	0.70
Ref-L025N0376	9.70×10^6	1.81×10^6	0.70
Ref-L025N0752	1.21×10^6	2.26×10^5	0.35
Recal-L025N0752	1.21×10^6	2.26×10^5	0.35
NoAGN-L050N0752	9.70×10^6	1.81×10^6	0.70

in the language of Schaye et al. (2015), compared to strong convergence tests, which use the same feedback parameters at different resolutions (Appendix A). The idea behind this is that the parameters for subgrid feedback will generally depend on which scale is considered subgrid, so the subgrid model parameters are expected to be resolution-dependent. Finally, NoAGN-L050N0752 is a variation of Ref-L050N0752 without AGN feedback, which was not described in Schaye et al. (2015) or Crain et al. (2015). Except for Appendix A, where we test resolution and box size convergence, we will only use the Ref-L100N1504 (reference), Ref-L050N0752 (50 cMpc reference), and NoAGN-L050N0752 (no AGN) simulations.

2.2 Column density calculation

To obtain column densities from the EAGLE simulations, we calculate the number of ions in columns (elongated rectangular boxes) of finite area and fixed length. We 'slice' the simulation along the Z-axis, then divide each slice along the X- and Y-directions into narrow three-dimensional columns, which, when projected, become the pixels of a (two-dimensional) column density map. We make such a column density map for each slice along the Z-direction. We use 32 000 columns along both 100 cMpc sides of the simulation box, meaning each column has an area of 3.125^2 ckpc^2 (comoving kpc) and each map has 1.024×10^9 pixels. We use 16 slices, each 6.25 cMpc thick. In Appendix A, we verify that the column density statistics are converged at this pixel size, and examine the effect of slice thickness. These columns are an approximation of what is done observationally, where absorbers are defined by regions of statistically significant absorption in spectra of nearly point-like sources (typically quasars), and column densities are obtained by fitting Voigt profiles to these regions.

To project the ions of each SPH particle onto a grid, we need to know the shape of the gas distribution that individual particles model. There is no unique function for this, since anything smaller than a single SPH particle is, by definition, unresolved. However, there is a sensible choice. In SPH, a similar assumption about the gas distribution must be made, which is used to evolve the gas particles' motions and thermodynamic properties. The function describing this is known as a kernel. We use this same function

in projection: the C2-kernel[2] (Wendland 1995). Along the Z-axis, we simply place each particle in the slice that contains its centre. We have verified that the results are insensitive to the chosen kernel: the difference in the column density distribution function (CDDF, the main statistic we are interested in) compared to using a different kernel (the Gadget kernel) is $\lesssim 0.05$ dex for 10^{11} cm^{-2} < $N_{O\,VII,\,VIII}$ < $10^{16.5}$ cm^{-2}, which covers the column densities we are interested in. We discuss the kernels in more details in Appendix A.

An issue that arises in EAGLE is that for cold, dense gas (star-forming gas), the temperature is limited by a fixed equation of state, used to prevent artificial fragmentation during the simulation (Schaye & Dalla Vecchia 2008). This means that the temperature of this gas does not represent the thermodynamic state we would actually expect the gas to be in, and ion balance calculations using this temperature are unreliable. Following e.g. Rahmati et al. (2016), we therefore fix the temperature of all star-forming gas to 10^4 K, typical for the warm phase of the ISM. However, we found that the treatment of this gas has virtually no impact, as O VII and O VIII are high-energy ions, and therefore mainly exist in hot and/or low-density gas (see Section 3.6). The difference between CDDFs calculated using our standard method, using the temperatures in the simulation output, and excluding star-forming gas altogether is < 0.01 dex at 10^{11} cm^{-2} < N < $10^{16.5}$ cm^{-2}.

The CDDF is defined as

$$f(N, z) \equiv \frac{\partial^2 n}{\partial N \partial X}, \qquad (1)$$

where X is the dimensionless absorption length, n is the number of absorption systems, and N is the ion column density. Here,

$$dX = dz \ [H_0/H(z)] (1 + z)^2, \qquad (2)$$

which means that at $z = 0$, $dX = dz$ (Bahcall & Peebles 1969). At higher redshifts, using dX accounts for changes in the CDDF due to the uniform expansion of the Universe for absorption systems of fixed proper size. We calculate dz from the thickness of the spatial region we project (i.e. the length along the projection axis), through the Hubble flow, using the redshift of the snapshot and the cosmological parameters used in the simulation.

2.3 Spectra and equivalent widths

We calculated the EWs along EAGLE sightlines using mock spectra obtained using a program called SPECWIZARD. It was developed by Joop Schaye, Craig M. Booth and Tom Theuns, and is described in Tepper-García et al. (2011, section 3.1). A given line of sight is divided into (one-dimensional) pixels in position space. First, the number of ions is calculated for each SPH particle intersecting the line of sight. Then the column density in each pixel is calculated by integrating the particles' assumed ion distribution (defined by the SPH kernel) over the extent of the pixel. This guarantees that the column density is correct at any resolution. For the ion distribution, we assume a three-dimensional Gaussian, since this is easy to integrate. We have verified that the difference with column densities and EWs obtained with a different kernel is negligible. The ion-number-weighted temperature and peculiar velocity along the line of sight are also calculated in each pixel.

Once this real space column density spectrum has been calculated, it is used to obtain the absorption spectrum in velocity space.

---

[2]In the simulation, the function depends on the 3D distance to the particle position; for the projection, we input the 2D distance (impact parameter) instead.

For each pixel in the column density spectrum, the optical depth distribution in velocity space is calculated as for a single absorption line. The absorption is centred based on the position of the pixel and the Hubble flow at the simulation output redshift, together with the peculiar velocity of the pixel. The thermal line width (b parameter) is calculated from the temperature. We have not modelled any subgrid/unresolved turbulence, and neglect Lorentz broadening in our calculations. We also use these 'ideal' spectra directly, and do not model continuum estimation, noise, line detection, or a detecting instrument in our EW calculations.

We calculate rest-frame equivalent widths, EW, from these mock spectra by integrating the entire spectrum:

$$\text{EW} = \left(1 - \sum_{i=0}^{N-1} \frac{F_i}{N}\right) \lambda_{\text{rest}} \frac{H(z)\, l_{\text{sl,com}}}{c\,(1+z)}, \qquad (3)$$

where N is the number of pixels, F_i is the flux in pixel i, normalized to the continuum, λ_{rest} is the rest-frame wavelength of the absorption line, z is the redshift, $H(z)$ is the local Hubble flow, c is the speed of light, and $l_{\text{sl,com}}$ is the comoving box size. The (observer-frame) velocity difference across the full 100 cMpc box is 7113 km s^{-1} ($\Delta z = 0.02373$) at redshift 0.1, and 12 026 km s^{-1} ($\Delta z = 0.040\,11$) at redshift 1 with the EAGLE cosmological parameters. The spectra we generate inherit the periodic boundary conditions of the simulation box, and therefore probe velocity differences of at most half these values.

We use Verner et al. (1996) oscillator strengths and wavelengths for the 18.97 Å O VIII doublet: $f_{\text{osc}} = 0.277, 0.139$ and $\lambda = 18.9671, 18.9725$ Å, respectively. For the O VII resonant line, we use values consistent with theirs: $\lambda = 21.6019$ Å and $f_{\text{osc}} = 0.696$, but ours come from a data compilation by Kaastra (private communication). The other O VII He-like lines (forbidden and intercombination) have wavelengths sufficiently far from the resonant line (e.g. Bertone et al. 2010a) that these should be clearly separated from each other at the resolutions achieved by instruments aboard Arcus (Brenneman et al. 2016; Smith et al. 2016), Athena (Lumb, den Herder & the Athena Science Team 2017; Barret et al. 2018), and Lynx (The Lynx Team 2018) (see Sections 3.2 and 4.1 for further discussion of the telescopes and instruments). Therefore, we only discuss the resonant line here, and note that for lines at the detection limits of these instruments, the lines are not blended. Also, these lines have much weaker oscillator strengths, by a factor of at least ~6000, so although they can be important in emission studies, the forbidden and intercombination lines will not be important in absorption.

Since we compute the EW by integrating the entire spectrum, if there is more than one absorption system along the line of sight, we will only recover the total EW. However, the EWs we are interested in are the potentially detectable ones, i.e. the rarer, larger values, corresponding to the highest column densities. These will not generally be 'hidden' in projection by even larger EW absorption systems, though we find that multiple systems of similar strength along such lines of sight are not uncommon. We will discuss this issue further in Section 3.3, where we examine some spectra and the conversion between column density and EW.

To determine the relation between the (projected) column density and EW, we obtained mock spectra along 16 384 lines of sight through the reference simulation (Ref-L100N1504) output at redshift 0.1. Since we are mainly interested in the systems with large EWs, we did not choose random lines of sight. Instead, we selected sightlines using the column density maps for O VI, O VII, and O VIII: for each ion, we selected samples randomly from

evenly spaced log column density bins. We used column densities over the full 100 cMpc box depth, along the Z-direction, for this selection.

The matching of EWs to projected column densities is the reason we calculate the EWs along 100 cMpc sightlines, despite the possibility of combining two absorption systems this way. The peculiar velocities of the absorbers are often ~ 500 km s^{-1}, and can reach 1000 km s^{-1} or more. Compared to a Hubble flow across the redshift 0.1 box of ~ 7000 km s^{-1}, this means that matching absorption in the spectrum to one of 16 slices along the line of sight would not be a straightforward process. A mistake here could have major implications for the reconstructed EW distribution, since the highest column density absorption systems are orders of magnitude more rare than more typical ones. This means that mismatches of absorption systems with large EWs to lower column densities could lead to a significant overestimation of the abundance of these rare absorption systems. For 100 cMpc, there is no possibility of mismatches along the line of sight.

With this sample of column densities and EWs, we make a two-dimensional histogram of column density and EW. By normalizing it by the number of absorbers at each column density, we obtain a matrix from this histogram that we use to convert column density distributions to EW distributions. Though we make these matrices for 100 cMpc sightlines, we will also apply them to column density distributions obtained using 6.25 cMpc slices of the simulation. That is to say, we assume that the relation between column density and EW does not depend significantly on the slice thickness (sightline length).

3 RESULTS

3.1 Column density distributions

The starting points for our column density distributions are column density maps. Fig. 1 shows a map of column densities calculated as we described in Section 2.2. These column densities are for columns through the full depth of the 100 cMpc box, using 400² pixels, and therefore do not produce the converged CDDFs we use in the rest of the paper. They do demonstrate that O VII and O VIII trace the large-scale structure in the box, as indicated by the total gas surface density map. We will investigate the spatial distribution of these ions (column densities around haloes of different masses) in more detail in an upcoming paper.

Fig. 2 shows the O VII and O VIII CDDFs (solid lines) we obtained from the reference simulation (Ref-L100N1504). The main feature visible here is the 'knee' or break in the distributions for both ions at column densities around 10^{16} cm^{-2}. We will investigate this feature in detail in Section 3.4, where we find that it marks the transition from absorption arising in extrahalo and intrahalo gas. We show in appendix A that these distributions are converged with pixel size in the column density range we show here (up to 10$^{16.5}$ cm^{-2}), and reasonably converged with the size and resolution of the simulations. We have verified that the effects of some technical choices in the calculation of the column densities are small.

The dotted lines indicate the distribution we would get if all gas had an oxygen abundance 0.1 times the solar value. This demonstrates the impact of non-uniform metal enrichment: it causes the CDDF to be much shallower, extending to larger maximum column densities. We have verified that hot or high-density absorbers are most enriched with metals, while the less dense IGM (especially the cooler part) is less enriched. This means the O VII and O VIII CDDFs are sensitive to the distribution of oxygen.
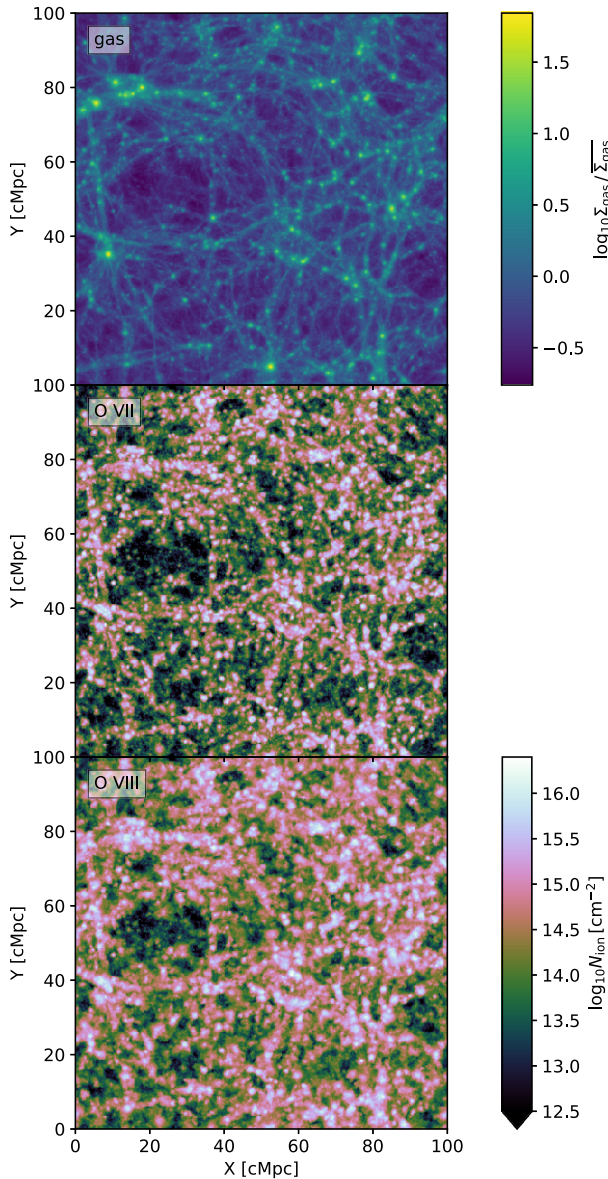
Figure 1. A map of the O VII (middle) and O VIII (bottom) column densities in the 100 cMpc EAGLE `Ref-L100N1504` simulation at $z = 0.1$, at a resolution of 400^2 pixels and with a column depth of 100 cMpc. The full column density range is not shown. The top panel shows the corresponding gas surface overdensities. O VII and O VIII trace large-scale structures.

From a different perspective, 40 per cent of gas-phase oxygen and 27 per cent of total oxygen[3] is in these two ions at $z = 0.1$ in the reference simulation (`Ref-L100N1504`). That means that absorption from these ions is also important in determining where the bulk of the metals produced in galaxies go.

We also compare our O VII and O VIII CDDFs to those of Nelson et al. (2018). They obtained column density distributions for these

---

[3]The total amount of oxygen here is the amount of oxygen released by stars, and does not count oxygen that remains in stellar remnants (a significant quantity) or is swallowed by black holes (a relatively insignificant quantity). Compared to gas-phase oxygen, the total also includes oxygen that was in gas, but is now in stars that formed from that gas.
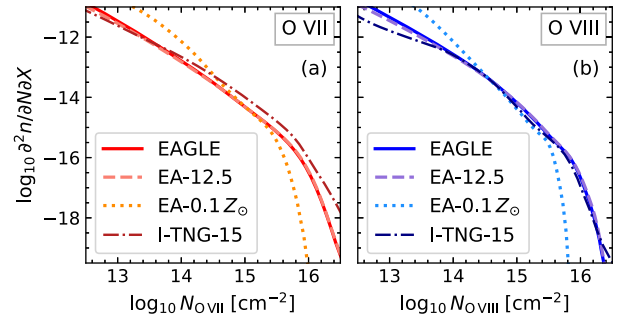
Figure 2. The CDDF, as described in equation (1), of O VII (a) and O VIII (b) in the 100 cMpc EAGLE reference simulation at $z = 0.0$. The solid lines (EAGLE) show our standard CDDF: it uses the simulation oxygen abundances, and column densities are measured in 6.25 cMpc long columns. The dotted lines (EA-0.1 $Z_\odot$) use 0.1 times the solar oxygen abundance for all gas instead. All the EAGLE column densities are calculated from $32\,000^2$ columns in each slice along the line-of-sight direction; each column has an area of 3.125^2 ckpc^2. We also compare to the IllustrisTNG 100-1 CDDF (I-TNG-15, Nelson et al. 2018), shown with dot–dashed lines. Since these column densities were measured in 15 cMpc columns, we also show the CDDF we obtain using 12.5 cMpc columns for comparison (dashed lines, EA-12.5). This shows that EAGLE and IllutrisTNG-100 predict similar column density distributions for these ions, and that using realistic metallicities is crucial in determining the column density distribution.

ions from the IllustrisTNG 100-1 simulation, using a similar method to ours. Nelson et al. (2018) use metallicity-dependent ion balances to calculate their column densities, but remark that the dependence of ion fractions on metallicity is minimal. They use 15 cMpc long columns to calculate their column densities, but we find that this makes almost no difference for the CDDF in the column density range shown here. They also use a different UV/X-ray background than adopted here.

The IllustrisTNG 100-1 simulation uses a similar cosmology, volume, and resolution as EAGLE, but a different hydrodynamics solver (moving mesh instead of SPH). While it also includes star formation and stellar and AGN feedback, it models these processes using different subgrid prescriptions.

Fig. 2 shows that the CDDFs from the EAGLE reference simulation (`Ref-L100N1504`) and IllustrisTNG-100-1 agree remarkably well; the differences are small compared to the effect of assuming a constant metallicity. Note that the EAGLE and IllustrisTNG-100-1 absorber numbers differ by a factor ≈ 2 at column densities above the break for O VII, though the differences are small compared to the dynamical range shown. A comparison at fixed metallicity (not shown) did not decrease the differences, meaning they are not dominated by different metal distributions in the two simulations.

Fig. 3 shows that the CDDF evolves little between redshifts 0 and 1. The shape of the CDDF does not change much, but there is some evolution: the incidence of high column densities decreases with time, while that of low column densities increases. The distribution evolves the least around its break. We have verified that these changes are not simply the result of the changing pixel area and slice thickness in physical units at the fixed comoving sizes we used, by comparing these changes to the effects of changing column dimensions described in appendix A. The evolution we find in EAGLE is similar to that shown in fig. 5 of Nelson et al. (2018).
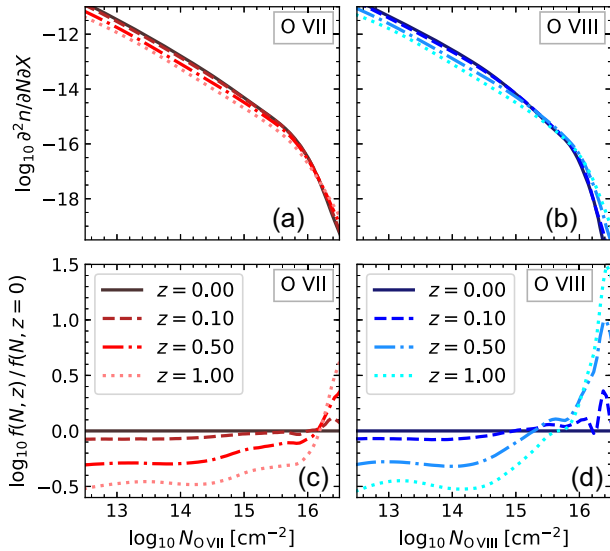
Figure 3. The evolution of the CDDF of O VII (a, c) and O VIII (b, d) in the 100 cMpc EAGLE reference simulation from $z = 1$ to 0. The top panels show the CDDFs themselves, while the bottom panels show the CDDFs relative to the $z = 0$ CDDF. This shows that the CDDF evolves only mildly between redshifts 0 and 1.

3.2 Absorption spectra

We move on to an examination of a few of the spectra we will obtain equivalent widths (EWs) from. Three example spectra are shown in Fig. 4. We show our ideal spectra (resolution of $2\,\mathrm{km\,s^{-1}}$, no noise), and mock spectra with spectral resolutions and effective areas corresponding to the Athena X-IFU, Arcus,[4] and Lynx X-ray Grating Spectrometer (XGS) instruments for a 100 ks observation of a blazar sightline. For the blazar, we use a source flux of $1 \times 10^{-11}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ between 2 and 10 keV with a photon spectral index of $\Gamma = 1.8$, as specified for the Athena 5σ weak line sensitivity limit by Lumb et al. (2017). We model the Poisson noise, but no other sources of error/uncertainty, and determine the unabsorbed number of photons per bin from the blazar spectrum and the redshifted line energy. We do not account for the slope of the blazar spectrum across our sightline. The spectra are periodic, like the simulations we derived them from, which is why we see absorption at low line of sight velocities in panel (f). We selected these spectra to be roughly representative of a larger sample we examined: 10 spectra at column densities of 10^{14}–$10^{17}\,\mathrm{cm^{-2}}$, in 0.5 dex intervals, for each ion.

The ideal spectra illustrate that most absorption systems have more than one component. Single-component systems do occur, though. Athena is more sensitive than Arcus, but it cannot distinguish close components as easily. Line widths, in Athena X-IFU and even in Arcus spectra, may not provide very stringent limits on absorber temperatures, since these widths are inferred mostly from the component structure. With Lynx, it may be more feasible to separate the different components.

We note that it is not uncommon to find multiple absorption systems along a single line of sight. This should not be a major problem for our 6.25 cMpc CDDFs, but it could affect the relation between column density and EW, since the column densities and

EWs of the different systems simply add up, even when the stronger system's absorption is saturated. However, we show in Appendix B that the EW distributions obtained using different conversions between column density and EW are quite similar around the expected detection thresholds of Athena, Arcus, and Lynx, so we do not expect this to have a significant impact on our survey predictions.

We see clear differences between the spectra obtained with the simulated instruments: Arcus has much higher resolution than Athena, enabling clearer measurements of line shapes and detections of different components, while Athena's much higher effective area enables it to detect weaker lines in the same observing time. Note that the science requirements for these two instruments specify different observing times for the proposed blind WHIM surveys. The Lynx XGS is planned to have the highest spectral resolution of the three and a large effective area, meaning it recovers these lines best. We will discuss these instruments in more detail in Section 4.1.

Finally, the ideal spectra confirm that we need to model the O VIII doublet as two blended lines. They are intrinsically blended, but the offset between the lines is large enough that it might affect the line saturation (panel a), so modelling the two lines as one may also be inadequate. The EW of the detected line will be determined by both doublet components together.

3.3 Equivalent widths

As described in Section 2.3, we calculated the rest-frame EW distribution in the reference simulation (Ref-L100N1504) by first obtaining the relation between the column density from the projection and the total EW along the same line of sight for a subset of sightlines with relatively high column densities. This relation includes the scatter in EW at fixed column density. We then used this relation to obtain the EW distribution from the column density distribution. This approximation is necessary because computing absorption spectra corresponding to the full $32\,000^2$ grid of pixels is too computationally expensive. We did not fit any models to this relation, but simply binned the distribution of EWs as a function of column density, and used it as a conversion matrix. Below the minimum column density of $10^{13}\,\mathrm{cm^{-2}}$ for which we selected sightlines, we used the linear curve of growth to make the conversion. We discuss the calculation of the EW distributions and the effect of including scatter in more detail in Appendix B.

Fig. 5 shows the rest-frame EW as obtained by integrating the absorption spectra, as a function of the column density computed with the same code (SPECWIZARD) used to generate the absorption spectra.[5] This shows that for the highest column densities and

---

[4]The effective collecting area for Arcus is $500\,\mathrm{cm^{-2}}$, but the spectrometer only uses half of this to improve spectral resolution (subaperturing).

[5]Here, for each ion, we only show the sightlines selected uniformly in column density for that particular ion, while we use the very similar distribution of all sightlines we have spectra for when converting column densities to EWs. If we plot the EWs against the column densities from our column density maps, there is slightly more scatter in the relation, which is clear at lower column densities. This is due to mismatches between the column densities calculated using the two different methods. These differences are generally small. At the highest column densities, they are due to non-convergence of the highest column densities with projection resolution (see Appendix A). More generally, the spectra and 2D projections assume slightly different gas distributions for a single SPH particle, and the 2D projections deal with SPH particles smaller than the projection column size in ways that the spectra do not. We account for these small differences when we calculate the EW distributions: we use the relation
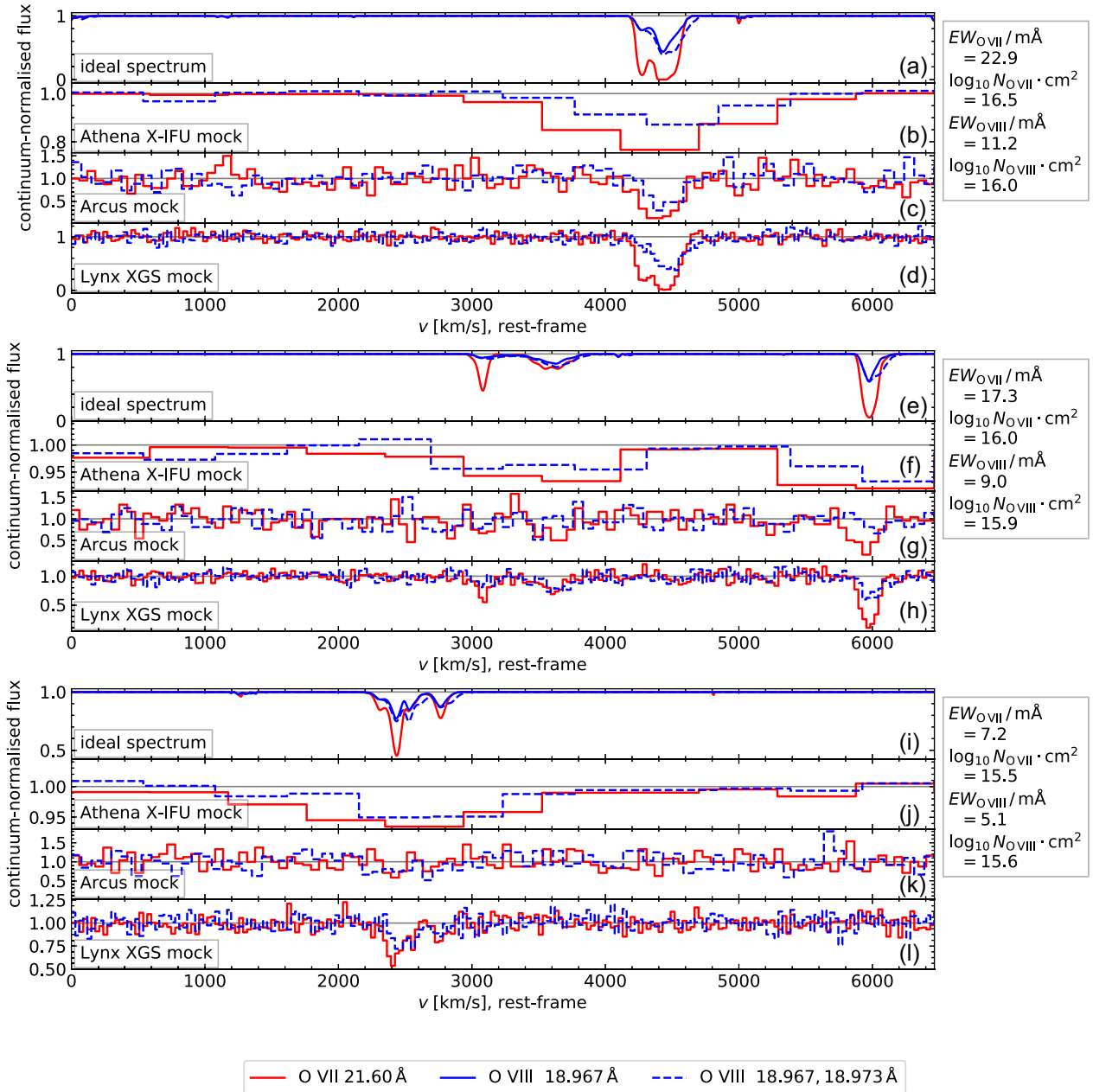
Figure 4. Example mock spectra from the EAGLE reference simulation (Ref-L100N1504) at $z = 0.1$. On the x-axis, we show rest-frame velocity, on the y-axis, we show the flux, normalized to the continuum. The (a), (e), and (i) panels show the ideal spectra the other panels are derived from. These have a resolution of 2 km s^{-1} and are noise-free. In the other panels, we show mock spectra for the Athena X-IFU (b, f, j), Arcus (c, g, k), and Lynx XGS (d, h, l) effective areas and resolutions. These spectra have Poisson noise added to them, but no other error or uncertainty sources. We show the spectra for the O VII resonant line and the O VIII doublet. In the ideal spectra, we also show the spectrum using only the stronger of the two O VIII doublet lines to gauge the effect of the weaker one. The three panel groups (a, b, c, d), (e, f, g, h), and (i, j, k, l) represent three different sightlines, selected to illustrate some general trends. We chose sightlines with O VII column densities of $10^{16.5}$, $10^{16.0}$, and $10^{15.5}$ cm^{-2}. The column densities and EWs measured over the whole sightline, for the ideal spectra, are shown to the right of the spectra. For Athena X-IFU mock spectra, we use a (full-width half-maximum) resolution of 2.1 eV and an effective area (specified for soft X-rays) of 1.05 m^2 (Lumb et al. 2017; Barret et al. 2018). For Arcus, we used a spectral resolving power $E/\Delta E$ (where ΔE is the full-width at half-maximum and E is the observed energy) of 2000 below 21.60 Å (observed wavelength), and 2500 at longer wavelengths, with an effective area (average at 16–28 Å) of 250 cm^2 (Smith et al. 2016). We used $E/\Delta E = 5000$ and an effective area (specified for 0.3–0.7 keV) of 4000 cm^2 for the Lynx XGS (The Lynx Team 2018). The background (blazar) source flux was 1×10^{-11} erg cm^{-2} s^{-1} between 2 and 10 keV with a photon spectral index of $\Gamma = 1.8$ as specified for the Athena 5σ weak line sensitivity limit by Lumb et al. (2017), and an observing time of 100 ks. The spectra, like the simulations they come from, are periodic. We bin the instrument mock spectra to two pixels per full width at half-maximum.
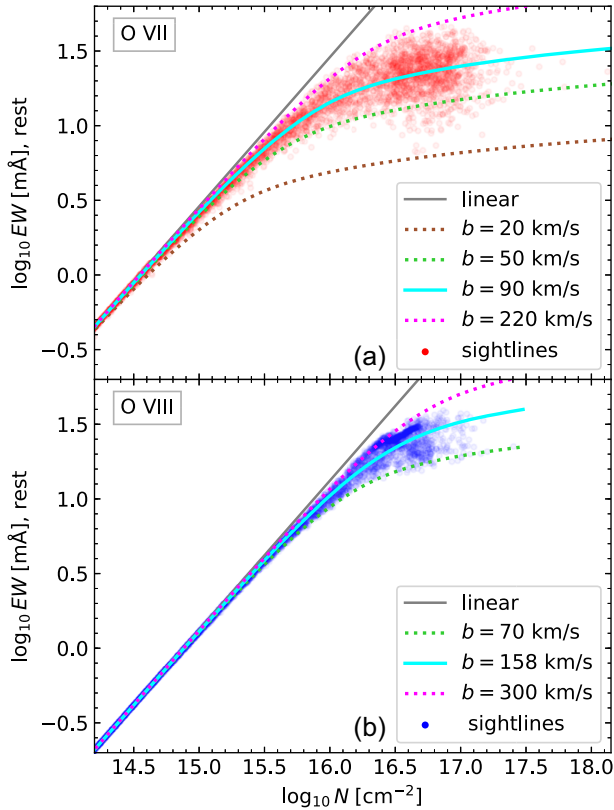
Figure 5. The curve of growth for the O VII resonant line (a) and the O VIII doublet (b) in the reference (`Ref-L100N1504`) simulation at redshift 0.1. We plot the rest-frame EW calculated over the full 100 cMpc sightlines against the column density along those same sightlines. The solid cyan lines are for the best-fitting *b* parameters to the log EW, using only the sightlines selected for the ion in question. The solid grey lines indicate the linear column density–EW relation, which applies to optically thin gas. The dotted lines roughly indicate the upper and lower range of *b* parameters, with the brown line indicating the bare minimum for O VII while the green lines indicate values below which absorbers are rare. This demonstrates our sampling of the column density–EW relation, and the range of single-absorber-EWs associated with it.

EWs ($\gtrsim 10^{15}$ cm^{-2} for O VII, $\gtrsim 10^{15.5}$ cm^{-2} for O VIII) , the relation between the two is no longer linear. Indeed, inspection of the mock spectra (examples in Fig. 4) shows the lines at these column densities are typically saturated. Furthermore, the scatter in the relation at the highest column densities is large, meaning a single curve of growth cannot be used to accurately convert between column density and EW. We discuss the effect of scatter on the EW distributions in Appendix B.

We characterize this relation using a *b*-parameter dependent column density–EW relation. This *b* is a measure of the line width: a model Gaussian absorption line profile has a shape $1 - \exp(-\tau(\Delta v))$, with $\tau(\Delta v) \propto Nb^{-1}\exp(-(\Delta v\, b^{-1})^2)$. Here Δv is the distance from the line centre in rest-frame velocity units, τ is the optical depth, and *N* is the column density. The constant of proportionality depends on the atomic physics of the transition producing the absorption. The solid and dotted lines in Fig. 5 show
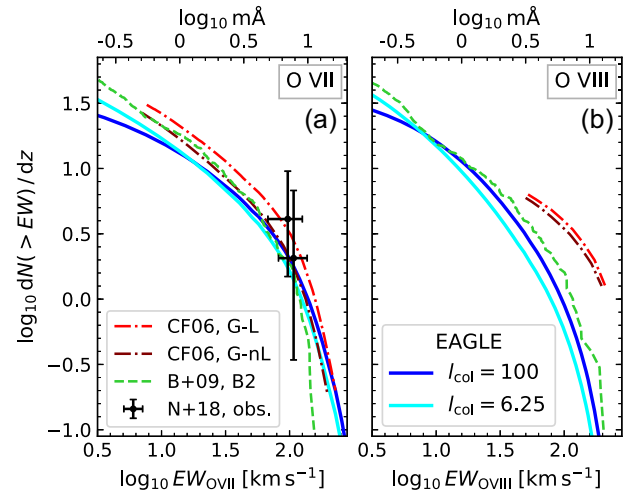
Figure 6. Comparison between the O VII (a) and O VIII (b) rest-frame EW distributions in the EAGLE reference simulation (`Ref-L100N1504`) at $z = 0.1$, the GSW model of Cen & Fang (2006) (CF06) at $z \sim 0$ with (G–L) and without (G–nL) assuming ionization equilibrium, the Branchini et al. (2009) (B+09) B2 model at $z \leq 1$, and observations by Nicastro et al. (2018) (N+18). Note that the units differ from those in Fig. 5: the EWs are in km s^{-1} (still rest-frame), and the number of absorption systems per unit redshift z is counted, rather than the number per unit absorption length *X*. The EAGLE distributions were computed using the 100 and 6.25 cMpc CDDFs and the 100 cMpc column density–EW relation calibrated to all the sightlines in our sample. This shows the rough agreement between the EW distribution we predict for O VII and that predicted in other work, and an agreement with the available observations. The O VIII distributions have larger differences, and there are no extant observational constraints we are aware of.

curves of growth for different *b* parameters. There is no clear trend of *b* parameter with column density.

In Fig. 6, we show the EW distributions we obtain from the CDDFs measured over 6.25 and 100 cMpc sightlines, and compare them to simulation predictions from other groups and a recent measurement. First, comparing our two distributions, the differences are as expected: when we group together multiple absorption systems along a longer sightline, we measure a higher number density of high EWs, while lower-EW absorption systems are 'swallowed up' and therefore less common. The differences are larger for O VIII, matching what we see in the CDDF when we vary the slice thickness in Appendix A.

We also compare our distributions to predictions made by Cen & Fang (2006) and Branchini et al. (2009), and to recent observations by Nicastro et al. (2018), including the update for the EW of the more distant absorber given by Nicastro (2018). Nicastro (2018) measured two O VII resonant line absorption systems at redshifts 0.43 and 0.36 in the spectrum of a very bright blazar, and used that to calculate the EW distribution for such absorbers. These measurements are consistent with all the predictions made so far, in part because the distribution measured from two systems is still quite uncertain.[6] Still, this agreement is encouraging. Fig. 3 of Nicastro et al. (2018) shows a similar agreement. The EAGLE EW distribution shown there was obtained from the column density

between column densities from our grid projections and EWs calculated at matching sightlines.

---

[6]The 68 per cent error bars in the figure account for this. Considering the overlapping error bars, we do not interpret these two data points as pointing to a steep slope in the distribution function.

distribution using a fixed *b*-parameter, and the 100 cMpc sightline CDDF.

The Cen & Fang (2006) predictions we compare to are based on the Cen & Ostriker (2006) simulations in a 123 cMpc box with 1024^3 cells, and a dark matter particle mass of $3.9 \times 10^8 \, M_\odot \, h^{-1}$. They use somewhat different cosmological parameters than EAGLE. Since these simulations solve the hydrodynamics equations on a fixed Eulerian grid, galaxies and the effects of their feedback are much less well-resolved in these simulations than in EAGLE. The results we show here are from their simulations with galactic feedback ('GSW'). The 'G-nL' curves are derived from simulations tracking ion abundances without assuming ionization equilibrium, while the 'G-L' curves were obtained by using equilibrium ion abundances instead. For the photoionization in this model, they use the $z = 0$ radiation field from the simulation, which is consistent with observations. They measured the EW distribution by making mock spectra with a resolution of $19 \, \text{km s}^{-1}$ (Fang et al. 2002) for random sightlines. They then identified and counted absorption lines in those. Their predictions are specified for redshift 0.

In Fig. 6, we also compare our EW distributions to the B2 model of Branchini et al. (2009, fig. 4). In this work, they present three models, of which they consider B2, the model shown here, to be the most realistic. It is also the most optimistic about what we can detect. The B2 model uses simulations by Borgani et al. (2004). This is an SPH simulation using 2×480^3 particles (gas + dark matter), in a 192 cMpc h^{-1} box and a different set of cosmological parameters from Schaye et al. (2015) or Cen & Ostriker (2006). The mass resolution is three orders of magnitude lower than for EAGLE. It includes star formation, supernova feedback, and radiative cooling/heating (for primordial gas). Branchini et al. (2009) impose a density–metallicity relation to get the oxygen abundance for each SPH particle. In model B2, the relation from the simulations of Cen & Ostriker (1999) is imposed, including scatter. They calculate ion abundances with CLOUDY, assuming ionization equilibrium. They measure the EW distribution from these sightlines by constructing lightcones, from which they obtain mock spectra at a resolution of $9 \, \text{km s}^{-1}$, and then identify lines in these. Branchini et al. (2009) plot observed EWs in their figure, for absorbers at redshifts 0–0.5. We show the rest-frame distribution obtained by assuming all their absorbers are at redshift 0.25.

For O VII, both the Branchini et al. (2009) and the Cen & Fang (2006) models agree reasonably well with our EW distribution, especially at higher column densities. For O VIII, the differences are much larger. The Branchini et al. (2009) B2 model is roughly consistent with EAGLE in this respect, while the latter predicts many fewer absorption systems at high column densities than Cen & Fang (2006).

3.4 The physical origin of the break

We now focus on the main feature of the O VII and O VIII column density distributions: the break ('knee') at large column densities (just below $10^{16} \, \text{cm}^{-2}$, see Fig. 2). We look into the temperatures, densities, and metallicities of absorption systems at different column densities to investigate this.

Figs 7, 8, and 9 show what type of gas dominates the CDDF at different column densities. Fig. 7 shows the contribution of absorption systems with different ion-weighted overdensities $\delta = \rho / \overline{\rho_b} - 1$. The shape of the total distribution (black) is different to that shown in some other figures because it shows the number of absorption systems per unit *log* column density (per unit absorption length), instead of per unit column density (per unit absorption length). We
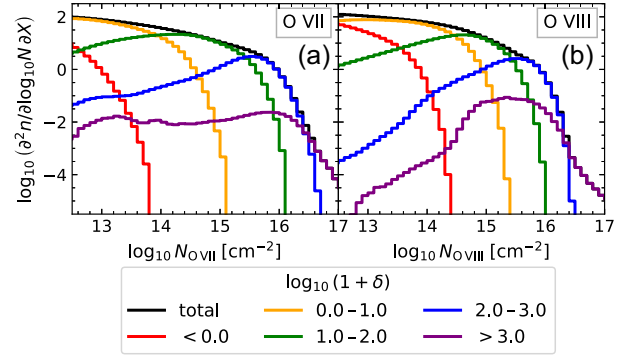
Figure 7. The CDDFs for O VII (a) and O VIII (b) absorption systems with ion-weighted overdensities in the ranges indicated in the legend. These distributions are for 6.25 cMpc sightlines in the reference (Ref-L100N1504) simulation at redshift 0. This shows that the break in the distribution occurs at overdensities $\sim 10^2$ for both ions, and therefore likely derives from the transition from absorption tracing sheets and filaments to absorption tracing haloes.
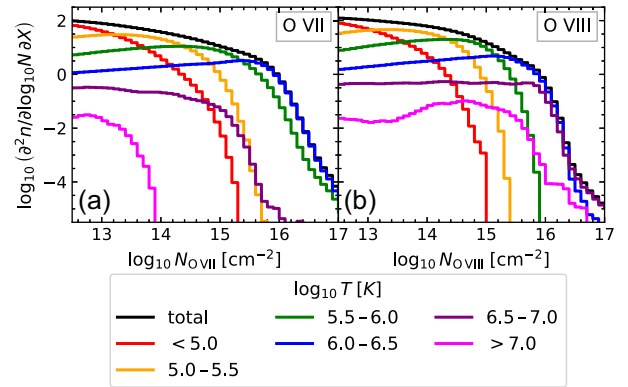
Figure 8. The CDDFs for O VII (a) and O VIII (b) decomposed into absorption systems with ion-weighted temperatures in the ranges indicated in the legend. This shows that WHIM gas dominates the distributions at all column densities likely to be observable in the foreseeable future, and that absorbers with column densities beyond the break are at temperatures associated with collisional ionization.
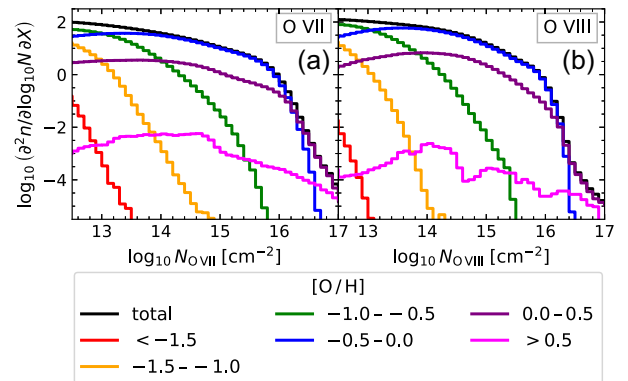
Figure 9. The CDDFs for O VII (a) and O VIII (b) decomposed into absorption systems with different ion-weighted SPH-smoothed oxygen abundances. The values indicated in the legend are ranges in the log oxygen mass fraction, relative to the solar oxygen mass fraction. This shows that most absorption systems we might detect in the foreseeable future should have oxygen-ion-weighted oxygen abundances around the solar value. (Note that mass- and volume-weighted metallicities may well be smaller.)

show column densities up to 10^{17} cm^{-2} here, unlike in most figures, to show what happens at the highest column densities. However, since the CDDFs are not converged above $10^{16.5}$ cm^{-2}, the values here should be taken as indicative.

We see in Fig. 7 that higher densities tend to produce higher column densities. At the break in the CDDFs, the ion-weighted overdensities are typically $\sim 10^2$. This points towards a cause for the break: gas at $\delta \gtrsim 10^2$ is typically within dark matter haloes, which have lower covering fractions than partially collapsed intergalactic structures (filaments, sheets, and voids) found at lower overdensities. The denser gas in haloes is also more likely to produce high column densities if it has the right temperature. This picture is broadly confirmed by visual inspection of the column density maps in Fig. 1,[7] which show that the highest column densities ($\gtrsim 10^{16}$ cm^{-2}) mostly occur in nodes in the cosmic web. This result is also roughly consistent with the results of e.g. Fang et al. (2002), who compared O VII and O VIII CDDFs from their simulation to analytical models and found they could explain the CDDF at high column densities as coming from collapsed haloes.

Fig. 8 shows the contribution of absorption systems with different ion-weighted temperatures to the O VII and O VIII CDDFs. The black line shows the total CDDF. Gas with temperatures of 10^5–10^7 K is canonically referred to as the WHIM, where e.g. Cen & Ostriker (1999) found most of the 'missing baryons' in their simulations. We see that absorbers tracing this gas dominate the CDDF at column densities likely to be observable with planned future missions. For O VII, we see that absorbers around the break and above tend to trace gas somewhat above $\sim 10^6$ K, while for O VIII, absorbers around the break tend to trace gas at $\sim 10^{6.5}$ K, with the temperature continuing to increase somewhat with column density. These temperature ranges are close to the temperatures where each ion attains its peak ion fraction in CIE. As we will show in more detail in Section 3.6, the absorption systems with column densities above the break do indeed tend to be primarily collisionally ionized while below the break, lower temperatures dominate and photoionization is important.

Finally, in Fig. 9, we turn to oxygen abundances. It shows that over a wide range in column densities, absorbers tend to trace gas with oxygen-ion-weighted metallicities at, or somewhat below, solar, while oxygen abundances become solar or larger at the largest column densities ($N \gtrsim 10^{16.3}$ cm^{-2}). Oxygen abundances above ≈ 3 times the solar value do occur, but they are rare. It is important to note that what we show here is the mean oxygen abundance weighted by the contribution of each gas element to the oxygen ion column density. Mass-weighted and volume-weighted mean metallicities can be much lower.

3.5 The effect of AGN feedback

To investigate the effect of AGN feedback on the CDDF, we compare the column density distributions we found in Section 3.1 to a recent EAGLE simulation not described by Schaye et al. (2015) or Crain et al. (2015): `NoAGN-L050N0752`. As the name suggests, this simulation does not include AGN feedback, while the rest of the (subgrid) physics is the same as in the reference model. This means it does not produce a realistic universe: AGN feedback is needed to quench star formation in high-mass galaxies and their progenitors in the EAGLE model, and to regulate the gas fractions of haloes.

[7]Recall that this figure shows column densities at a much lower resolution than we use for the CDDFs, and measures column densities in 100 cMpc columns.
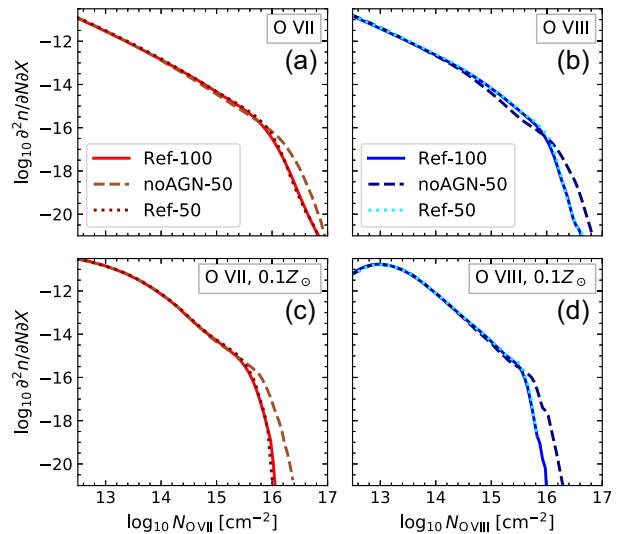

Figure 10. The top panels (a, b) compare the reference and no AGN models, to show the effect of AGN feedback on the $z = 0.1$ O VII (left-hand panels) and O VIII (right-hand panels) CDDFs. The column density distribution in the no AGN simulation (`NoAGN-L050N0752`, 'noAGN-50') is compared to the reference simulation (`Ref-L100N1504`, 'Ref-100'). We also show the 50 cMpc reference simulation (`Ref-L050N0752`, 'Ref-50') results to demonstrate that differences between the reference and no AGN simulation distributions are not due to the size of the simulation domain. To gauge how much of the differences are due to differences in gas oxygen content, we show CDDFs made assuming all gas has a metallicity of 0.1 $Z_\odot$ in the bottom panels. This shows that beyond the break in the CDDF, a boost in metal enrichment due to AGN feedback partly offsets its other effects.

Fig. 10 also shows the CDDFs we get from the reference model in a 50 cMpc box (`Ref-L050N0752`) to verify that any differences are not due to the difference in box size.

The top panels of Fig. 10 show that, when AGN feedback is disabled, the CDDF is barely affected at the lowest column densities, intermediate column densities are slightly less common, and the highest column densities occur more frequently. The decrease at intermediate column densities is larger for O VIII than for O VII. At column densities $\gtrsim 10^{16.5}$ cm^{-2}, the CDDFs are not converged with pixel size in the column density maps (see Appendix A). We show larger column densities in this plot mainly to show the difference with the no AGN simulation at high column densities more clearly, but relative differences in this range may not be reliable.

Since massive galaxies in the no AGN simulations produce too many stars, but too weak outflows, we might expect that some of this difference, especially at higher column densities probing halo gas, might be due to differences in gas metallicity. To check this, we do a similar comparison in the bottom panels of Fig. 10, except we assume all gas has a metallicity of 0.1 times the solar value when we calculate the number of ions in our columns.

We then note that the differences below the break in the CDDF (almost) disappear when comparing constant metallicity results, while the differences above the CDDF break increase somewhat. We interpret the causes of these differences as follows. Within the haloes, we see larger column densities in the absence of AGN feedback. This could be due to higher densities, metallicities, or mass of the hot gas responsible for the absorption at these high column densities. The effect of AGN feedback at higher column densities increases, but only a bit, if we assume a fixed metallicity. This indicates that the main effect of AGN feedback is to decrease
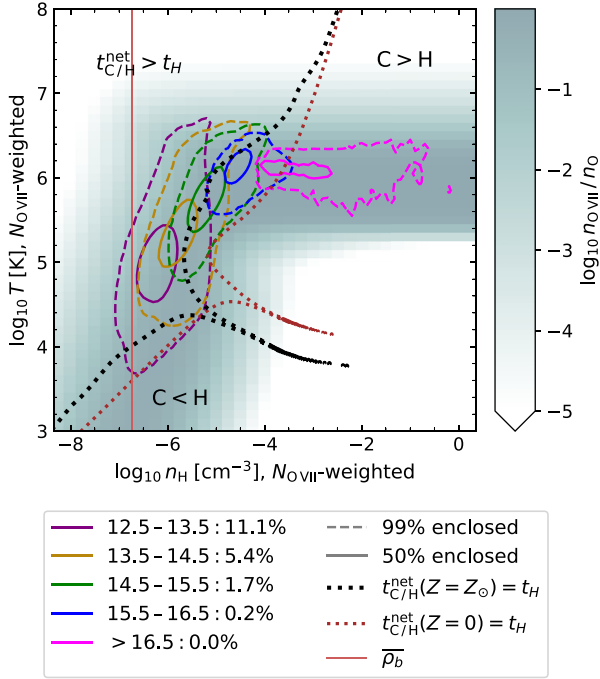
Figure 11. Phase diagram of the O VII absorption system density and temperature in different column density bins in `Ref-L100N1504` at $z = 0$, using 6.25 cMpc long sightlines. The rainbow-coloured contours indicate the distribution of ion-weighted average temperatures and densities for absorbers with column densities in different bins. The legend indicates the bins by their values of $\log_{10} N_{\text{O VII}} / \text{cm}^{-2}$; the percentages indicate what fraction of columns in the simulation (including those with zero column density) is in each bin. Solid and dashed contours enclose 50 and 99 per cent of the columns in each bin, respectively. The grey-scale in the background indicates which fraction of oxygen nuclei are O VII ions. We indicate gas at fixed heating and cooling time-scales using the dotted pairs of contours. They indicate constant net radiative cooling (upper contours) and heating (lower contours) time-scales $t_{\text{C/H}}^{\text{net}}$ equal to the Hubble time t_H for gas with solar metallicity and with primordial abundances, using a Haardt & Madau (2001) background. The vertical brown line indicates the $z = 0$ average baryon density.

the density of the hot gas, while it also increases its metallicity. These effects partially cancel out, but the density effect dominates. The enhancement of metal ejection by AGN feedback also explains the increase in the CDDF below the break.

This picture is supported by the work of Davies et al. (2019). They showed that for haloes with masses $\gtrsim 10^{12} \, \text{M}_\odot$ at $z = 0$ in EAGLE, the amount of AGN feedback injected into these haloes is a good predictor of the halo gas fraction at fixed halo mass, and that this is due to AGN ejecting gas from the CGM. In a follow-up paper, Oppenheimer et al. (2019) showed that this feedback also decreases ion column densities around galaxies, though they discuss H I, C IV, and O VI.

3.6 Physical properties of the absorbers

To investigate the physical properties of the absorption systems, we first look at the temperature and density of gas traced by the O VII and O VIII ions. In Figs 11 and 12, contours in different colours show the distribution of absorption system temperatures and densities in different column density ranges for O VII and O VIII, respectively. The grey-scale shows the fraction of oxygen ions in
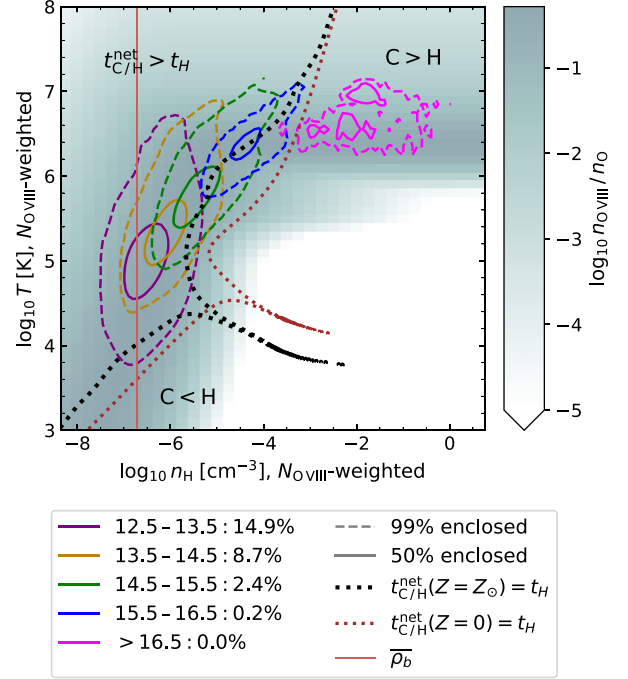
Figure 12. As Fig. 11, but for O VIII.

the ionization state of interest, and the dotted contours show lines of constant net radiative cooling or heating time-scale. The hydrogen number density displayed along the axis, used to calculate the ion balance as well as the cooling times, was calculated from the mass density assuming the primordial hydrogen mass fraction 0.752, but the difference with a solar hydrogen mass fraction conversion is negligible.

Comparison of the coloured contours and the grey-scale shows that the absorbers reside in regions of the phase diagram where the fraction of oxygen ions in that state is high, as expected. These ion fractions account for photoionization by the UV/X-ray background (Haardt & Madau 2001). At high densities, the ion fraction becomes density-independent, as the influence of the ionizing radiation becomes negligible and the ion balance asymptotes to CIE.

As expected, the column density increases with the density of the absorbing gas, with the lowest column densities often coming from (almost) underdense gas. The percentages in the legend indicate the fraction of sightlines (including those with temperatures and densities outside the plotted range) in that column density range. The highest column densities are rare, accounting for less than 0.1 per cent of sightlines. Furthermore, as the ion fraction colouring shows, most of the sightlines have their absorption coming mainly from photoionized gas: the gas at lower densities where the ion balance is density-dependent. However, at the highest column densities, especially where $N_{\text{O VII}} \gtrsim 10^{15.5} \, \text{cm}^{-2}$, the absorption comes mainly from collisionally ionized gas.

For ionization models of observed absorbers, it is important to know what are reasonable assumptions when trying to establish what sort of gas an absorber traces. For both O VII and O VIII, photoionized gas dominates the absorption at lower column densities. For O VII, gas at the CDDF break ($N_{\text{O VII}} \sim 10^{16} \, \text{cm}^{-2}$) is clearly dominated by gas in the CIE peak temperature range (Fig. 8). For O VIII, WHIM gas is also clearly dominant there, but CIE temperature gas is only just becoming dominant. At the CDDF break ($\sim 10^{16} \, \text{cm}^{-2}$), CIE gas is important, but the absorption

systems here have sufficiently low densities that photoionization still influences the ion fractions. For a given temperature, the ion fractions at $n_H \sim 10^{-5}$ cm^{-3} [the value used by Nicastro et al. (2018) to model the absorbers they found] can differ from the CIE values by factors $\gtrsim 2$ within the CIE temperature range (i.e. at temperatures where, in CIE, the ion fraction is at least half the maximum value). Therefore, in modelling these absorption systems, a temperature consistent with CIE does not necessarily imply that CIE ion fractions will be accurate.

To quantify what fractions of gas are affected by this, we divide the temperature–density plane into four regions, defined by a temperature and a density cut.[8] For the temperature cut, we use the lowest temperature at which the ion fraction in CIE reaches half its maximum value: $T \approx 10^{5.5}$, $10^{6.2}$ K for O VII and O VIII, respectively. For the density cut, we focus on the same temperature ranges, where, in CIE, the ion fraction is at least half the maximum value. We consider the gas to be in CIE when it has a temperature above the cut, and a density above a minimum value. This minimum is the lowest density where the ion fractions in the CIE temperature range defined above, differ from those at the highest density we have data for (10^3 cm^{-3}) by less than a factor 1.5. This means $n_H \approx 10^{-4.75}$, $10^{-4.25}$ cm^{-3} for O VII and O VIII, respectively. For O VII and O VIII, ≈ 32 and 8 per cent of the ion mass is in CIE, respectively. Gas at lower densities, but with a temperature above the cut, is the gas that may be mistaken for gas in CIE based on temperature diagnostics, but where ionization modelling based on this assumption may cause errors. This is ≈ 45 and 19 per cent of the O VII and O VIII mass, respectively. Gas below the density and the temperature cuts is not necessarily in pure PIE: the ion fractions are still influenced by collisional processes at the higher temperatures in this regime. This gas accounts for ≈ 23 and 72 per cent of the O VII and O VIII mass, respectively. A small fraction of the O VII and O VIII mass, ≤ 0.5 per cent, is at densities above the CIE threshold, but below the temperature cut. For these density thresholds and ions, there is more gas at CIE temperatures that is not actually in CIE than there is CIE gas. However, note that this depends on choices we made. If we choose our density thresholds based on a maximum factor 2 difference between the ion fraction and full CIE ion fraction, both ions have more gas in CIE than in the error-prone regime.

We look into the effect of temperature further by comparing the b parameter range of Fig. 5 with the thermal broadening for different temperature ranges we see in Figs 11 and 12. Note that those b parameters come from comparing total EW and column density, not line fitting, and do not account for instrumental broadening.

The thermal contribution to the line widths $b_{th} = \sqrt{2kT\,m^{-1}}$ is equal to 16–57 km s^{-1} for gas with temperatures $T = 10^{5.4}$–$10^{6.5}$ K. This range dominates the O VII CDDF at the column densities $\gtrsim 10^{15}$ cm^{-2}, where Fig. 5 shows we can estimate the line width from the column density and EW. For O VIII, we find $b_{th} = 36$–81 km s^{-1} for temperatures around the O VIII collisional ionization peak ($T = 10^{6.1}$–$10^{6.8}$ K). For the gas at $T = 10^{7.0}$ K, reached by some of the highest-column-density O VIII, we find $b_{th} = 102$ km s^{-1}. For both ions, the lowest EWs we find are consistent with thermal broadening, but the typical b parameters for these absorption systems are larger than predicted by their ion-

---

[8]Note that we use the intrinsic temperature and density of gas in EAGLE (not the ion-weighted values in projection), along with its ion content, for these calculations. This means we make cuts in the temperature–density plane of Fig. 13.
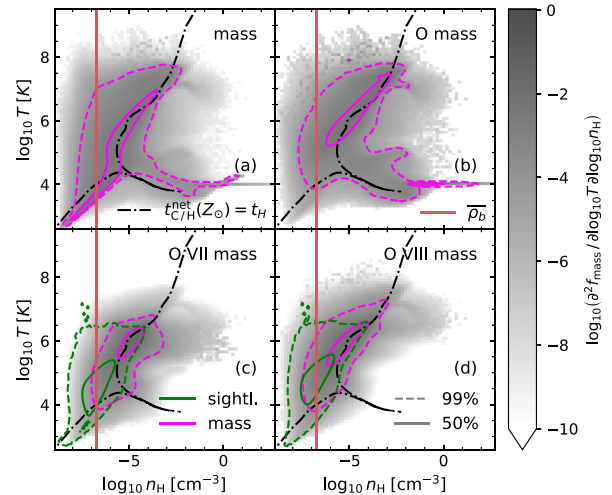

Figure 13. Phase diagrams for the total gas mass (a), oxygen mass (b), and O VII (c) and O VIII (d) masses in `Ref-L100N1504` at $z = 0$. The solid and dashed fuchsia contours enclose 50 and 99 per cent of the total mass. The green contours show the distribution of ion-weighted temperatures and densities that were computed for 6.25 cMpc long columns. This is similar to the distributions shown in Figs 11 and 12, but for all non-zero ion column densities. The black, dot–dashed contours indicate a constant radiative cooling or heating time-scale equal to the Hubble time for gas with solar abundances. The vertical brown lines indicate the $z = 0$ average baryon density.

weighted temperatures. This strengthens our conclusion from visual inspection of virtual spectra that velocity structure is important for line widths, especially when the spectral resolution is too low to resolve components in absorption systems. This also means that measured line widths for these ions may not provide meaningful constraints on absorber temperatures.

We show contours indicating where the radiative net cooling (or heating) time equals the Hubble time for gas with primordial and solar abundances in brown and black, respectively. We calculated these time-scales using the tables of Wiersma et al. (2009a), which were also used for radiative cooling in the simulations. These include the effects of the evolving Haardt & Madau (2001) UV/X-ray background. For each metallicity, the lower-temperature contour indicates where the net heating time-scale is equal to the Hubble time , with net heating within a Hubble time at lower temperatures (C < H). The higher-temperature contours indicate net cooling time-scales, with net cooling within a Hubble time at higher densities (C > H). The pairs of contours converge in 'peaks' to the right at $\sim 10^4$ K, where radiative cooling and heating balance each other.

As column densities increase, more of the absorbing gas can cool within a Hubble time. For gas at column densities $\gtrsim 10^{16}$ cm^{-2}, which we noted previously is dominated by halo-density gas, this means that the absorbing gas may be a part of cooling flows, where hot gas cools radiatively as it flows from the CGM to the central galaxy. As we saw in Fig. 9, gas with such column densities typically has roughly solar or larger oxygen abundances. However, some of it may also be gas that has recently been shock-heated in outflows.

We explore the effect of the gas phase distribution on the absorption system properties with Fig. 13. The panels show the distribution of gas for mass, oxygen mass, and O VII and O VIII mass. The horizontal lines for high densities at 10^4 K in the upper panels are the result of setting all star-forming gas to that temperature. The green contours (sightline distribution, for 6.25 cMpc long sightlines,

as in Figs 11 and 12) and fuchsia contours (mass distribution) seem to be similar, as expected, but the sightline distribution prefers lower-density gas. This is due to its higher volume-filling fraction, hence higher covering fraction, relative to its mass fraction. We also show the radiative heating/cooling time contours at solar abundances as in Figs 11 and 12. This demonstrates how the cooling and heating time-scales shape gas conditions: most gas is either roughly at the temperature set by radiative heating over a Hubble time, or is too hot or diffuse to cool within about a Hubble time.

Comparing the ion mass distributions to the mass and oxygen distributions clearly shows the effect of ionization state fractions and metallicities. The low-density, low-temperature gas associated with the cool IGM contains few ions because it contains little oxygen. Comparing the ion mass distribution to the total oxygen mass distribution shows that the low-temperature, high-density gas (associated with haloes and galaxies) is not traced by these ions because neither collisional nor photoionization ionizes much of the oxygen to O VII or O VIII in this regime.

3.7 Correlations between O VIII, O VII, and Ne VIII, O VI, and H I absorption

We now look at how O VII and O VIII absorption correlate, and also how this absorption correlates with neutral hydrogen, O VI and Ne VIII. This is shown in Figs 14 and 15. The neutral hydrogen we show is H I and H_2 together, where we count the total number of hydrogen atoms. We use the prescription of Rahmati et al. (2013) to obtain the hydrogen neutral fraction, since the ionization balance tables we use for the other ions do not account for gas self-shielding against ionizing radiation, which is important for H I fractions in cool, dense gas. For all but the highest column densities, the neutral fraction will be dominated by H I. For neutral hydrogen, we see that over a very large range of column densities ($N \sim 10^{15}$–10^{21} cm^{-2}), the trend between hydrogen column density and O VII and O VIII column density is flat, with a lot of scatter. However, we also see that in this range ~ 50 per cent of the hydrogen absorbers have O VII or O VIII counterparts with column densities $\gtrsim 10^{15}$ cm^{-2}, which is potentially observable, especially when searching at a specific redshift.

In fact, Kovács et al. (2019) have already reported extra-galactic O VII absorption by stacking quasar spectra using the known redshifts of 17 Lyα absorbers. They did make an extra selection, using only absorbers associated with massive galaxies (stellar mass $\gtrsim 10^{10}$ M$_\odot$). The average EW of the Lyα absorbers was 174.4 mÅ (corresponding to a column density $\sim 10^{14}$ cm^{-2} for $b \approx 20$ km s^{-1}), and they found an O VII column density of $1.4 \pm 0.4 \times 10^{15}$ cm^{-2}. Depending on the line width, this EW may correspond to a column density close to or in the flat part of Fig. 14(a), and the O VII column density they measure (an average for their stacked spectra) is similar to the median we predict for that column density range.

The other ions correlate better with O VII and O VIII along 6.25 cMpc sightlines. Fig. 14 also shows how good O VI and Ne VIII are at predicting strong absorption in our two X-ray lines, and what sort of O VIII counterparts to expect for an O VII line. The tightest correlation is between Ne VIII and O VII, with a maximum 1–99 per cent scatter of 1.1 dex at $N_{\rm Ne VIII} > 10^{11}$ cm^{-2}. (The maximum scatter at fixed O VII column density, as shown in Fig. 15, is the same.) The tighter correlation between Ne VIII and O VII than between O VI and O VII is likely because Ne VIII is a higher-energy ion than O VI, with an ion fraction that peaks in almost the same band in the density–temperature plane as O VII, but with a narrower

peak. We also notice from both figures in this section that O VII is better correlated with both the UV ions than O VIII.

We can see from Fig. 14 that a survey targeted at sightlines with observed high O VI or Ne VIII column densities has a greater chance of detecting O VII and O VIII systems than a blind survey. As an example, we showed in Section 3.3 that EWs of ~ 1 mÅ require column densities of $\sim 10^{14.5}$ cm^{-2} for O VII and $\sim 10^{15}$ cm^{-2} for O VIII. For those threshold column densities, we predict a 50 per cent chance of detecting an O VII (O VIII) counterpart to an O VI absorber at $N_{\rm O VI} \approx 10^{13.4}$ cm^{-2} ($10^{14.1}$). For Ne VIII, the corresponding column density is $N_{\rm Ne VIII} \approx 10^{12.9}$ cm^{-2} ($10^{13.3}$). These column densities are close to the breaks in the CDDFs $f(N, z)$ (see equation 1) of the FUV ions, but not beyond them. We find $f_{\rm O VII}(10^{13.4}, z = 0.1) = 3 \times 10^{-13}$ cm^2 and $f_{\rm Ne VIII}(10^{12.9}, z = 0.1) = 1.3 \times 10^{-12}$ cm^2 at the O VII threshold (1.7×10^{-14}, 3×10^{-13} cm^2, respectively, for O VIII).

Fig. 15 is useful for checking whether an observed absorption system matches our expectations: given an O VII or O VIII column density, it shows what counterparts to expect. For neutral hydrogen, the scatter is very large at O VII or O VIII column densities $\gtrsim 10^{15}$ cm^{-2}, but for the other ions, the correlations are tighter.

3.8 Correlation between absorber properties

Coincident absorption from more than one ion can give powerful constraints on the temperature and/or density of the gas, particularly if the ions belong to the same element. However, this only works if the same gas is causing the different absorption lines. Since gas clouds do not have constant densities and temperatures, it is possible that different absorption lines arise in different parts of the same clouds. Since in planned blind X-ray absorption surveys with Athena, Arcus, and Lynx, O VII and O VIII are the main targets (Brenneman et al. 2016; Smith et al. 2016; Lumb et al. 2017; The Lynx Team 2018), it is useful to investigate how closely the absorption system properties for these ions match. To do this, we compare temperature and density weighted by O VII and O VIII along the same 6.25 cMpc sightlines. We also compare to O VI, which has a strong doublet in the FUV range that may provide further constraints without element abundance ratio uncertainties, and to Ne VIII, which we found in Section 3.7 correlates very well with O VII and O VIII. We do not compare to neutral hydrogen gas properties, since we expect neutral hydrogen to reside in cooler, denser gas than the other ions we considered in Section 3.7. In addition, any temperature or density constraints from hydrogen to oxygen ion column density ratios would depend on the (generally unknown) metallicity of the gas.

First, we explore density differences in Fig. 16. This shows the distribution of O VI, Ne VIII, O VII, and O VIII-weighted densities[9] in the 6.25 cMpc columns we use to measure column densities. Contours show the distribution of sightlines in different O VII column density ranges. The light blue lines indicate where the densities are equal. We can see that at low column densities, where the gas is photoionized ($n_{\rm H} \lesssim 10^{-5}$ cm^{-3}), O VII tends to trace slightly higher-density gas than O VIII. This makes sense, since lower-density gas in the same radiation field is more highly ionized. The differences here are largely (≥ 98 per cent of sightlines) below 0.6 dex in all column density ranges and for all pairs of ions

---

[9]The axes show density as hydrogen number density, but the densities used for the plot are mass densities, converted to hydrogen number densities assuming a primordial hydrogen mass fraction of 0.752.
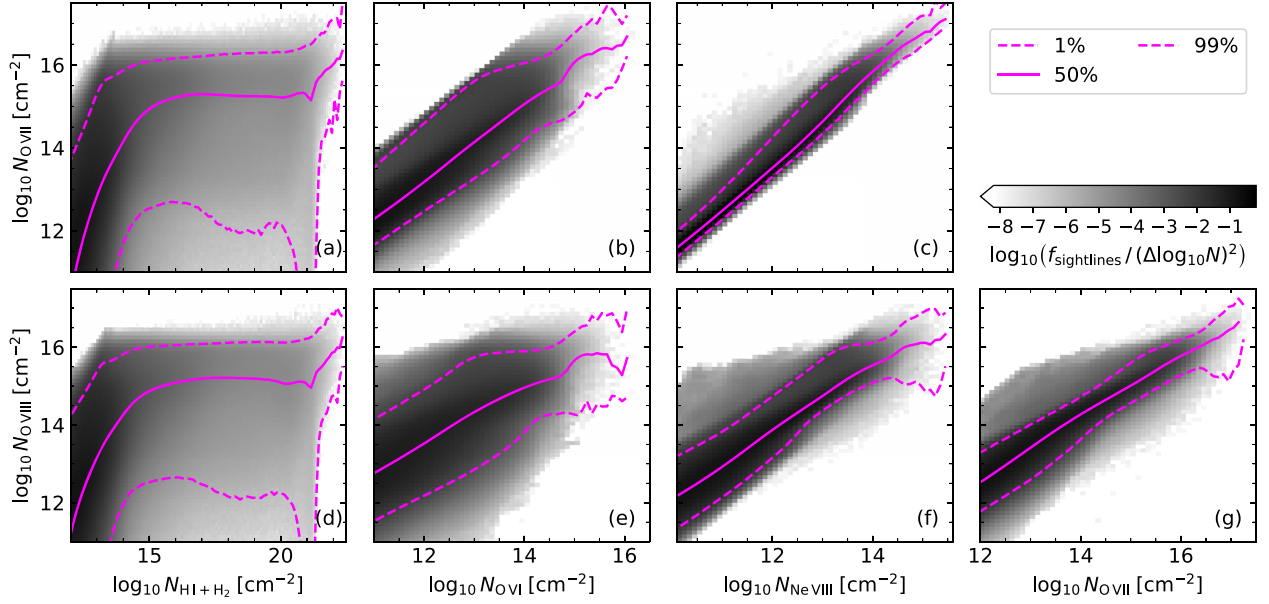
Figure 14. The range of counterpart column densities for O VII (a, b, c) and O VIII (d, e, f, g), given the column densities of neutral hydrogen (H I and H₂; a, d), O VI (b, e), Ne VIII (c, f), and O VII (g). We show the correlations for 6.25 cMpc sightlines in the reference simulation (`Ref-L100N1504`) at $z = 0.1$. The grey-scale shows the distribution of sightlines in the column density of each pair of ions, while the lines show percentiles in the distribution of the ion on the *y*-axis at fixed column density on the *x*-axis. Large H I column densities ($N_{\rm HI+H_2} \gtrsim 10^{15}\,{\rm cm^{-2}}$) predict a ~ 50 per cent chance of finding a strong ($N \gtrsim 10^{15}\,{\rm cm^{-2}}$) O VII or O VIII counterpart. Strong O VI or Ne VIII absorption is a good predictor for strong O VII and O VIII absorption.

we show. The median differences are typically small, $\lesssim 0.1$ dex, except at $N_{\rm OVII} > 10^{16.5}\,{\rm cm^{-2}}$ for O VII and O VIII, where the median difference is still < 0.2 dex. The differences between other combinations of these ions are similar.

Next, we look into temperature, shown in Fig. 17, similar to Fig. 16, but for ion-weighted temperatures. Between O VII and O VI or O VIII, differences are typically larger than for density, and clearly systematic: the higher-energy O VIII ions prefer hotter gas than O VII at all (column) densities, while O VII traces hotter gas than O VI. The median differences at lower column densities ($N_{\rm OVII} \lesssim 10^{15.5}\,{\rm cm^{-2}}$) are $\lesssim 0.2$ dex, but at higher column densities, the median temperatures differ by up to ≈ 0.3 dex for O VII and O VI or O VIII. This occurs at column densities tracing collisionally ionized gas, where the ions both tend to trace their own CIE peak temperature gas. For Ne VIII and O VII, we see a hint of a similar CIE peak preference in the narrowing of the range of absorber temperatures as the O VII column density increases, but because these two ions have similar peak temperatures, their median differences remain $\lesssim 0.1$ dex.

We also compared temperatures and densities traced by O VI and O VIII along the same 6.25 cMpc sightlines. We found similar trends in the differences to those we found between O VI and O VII and between O VII and O VIII, except that both the scatter and systematic differences were larger.

The differences in gas temperature and density probed by different ions in the same structures may be large enough to be a potential issue when modelling the absorption as coming from the same gas. In future work, we intend to test this by applying ionization models to column densities measured from virtual absorption spectra. This will also enable us to test whether enforcing different lines to have identical redshifts would help reduce the systematic errors given the energy resolution expected for upcoming missions.
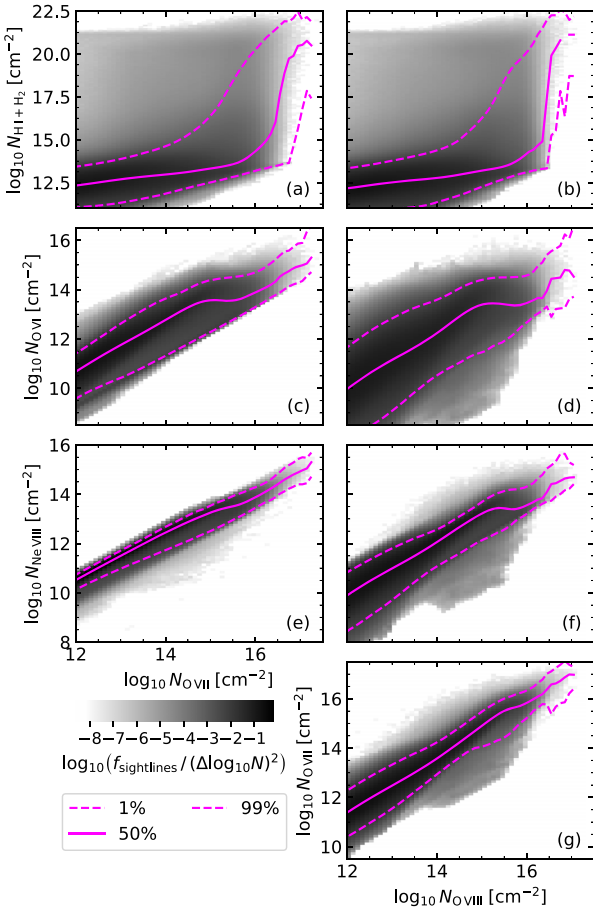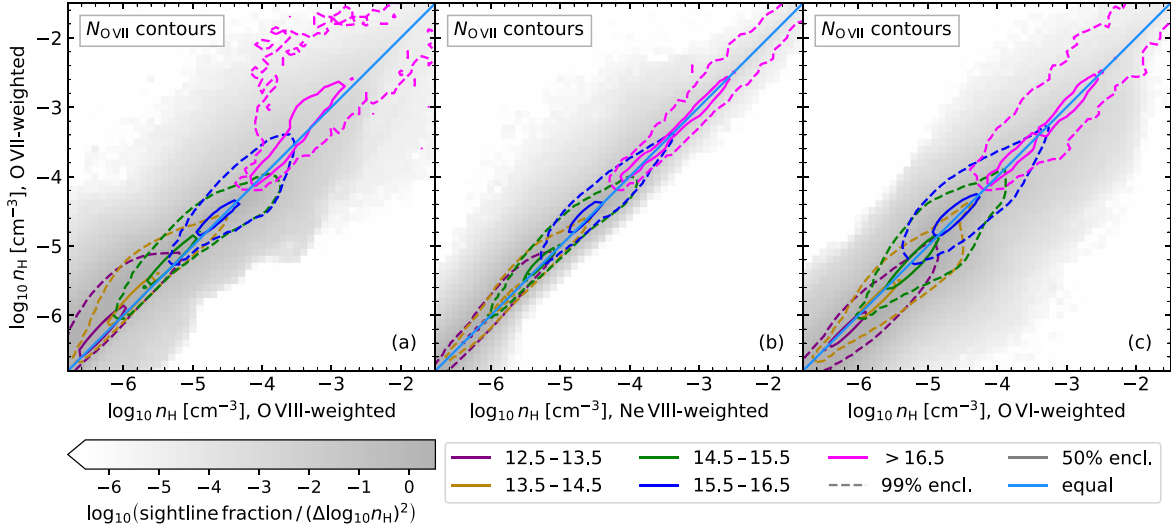
Figure 15. As Fig. 14, but with the axes reversed. For a given O VII or O VIII column density, this indicates whether a counterpart agrees with what we find in EAGLE.

Figure 16. The O VII-weighted density (vertical) of absorption systems plotted against the O VIII-weighted density (a), the Ne VIII-weighted density (b), and the O VI-weighted density (c) along the same sightline. These sightlines are our 6.25 cMpc long columns, and come from $32\,000^2$ pixel projections of the reference (Ref-L100N1504) simulation at redshift 0. The grey-scale shows the total sightline distribution, while the contours show the density distributions for different O VII column density bins, matching those in Fig. 11. These ranges are given in units of $\log_{10}\mathrm{cm}^{-2}$. The light blue lines indicate where the densities are equal. This indicates that these ions tend to trace gas with similar densities along the same sightline, though there are small systematic differences at low column densities.
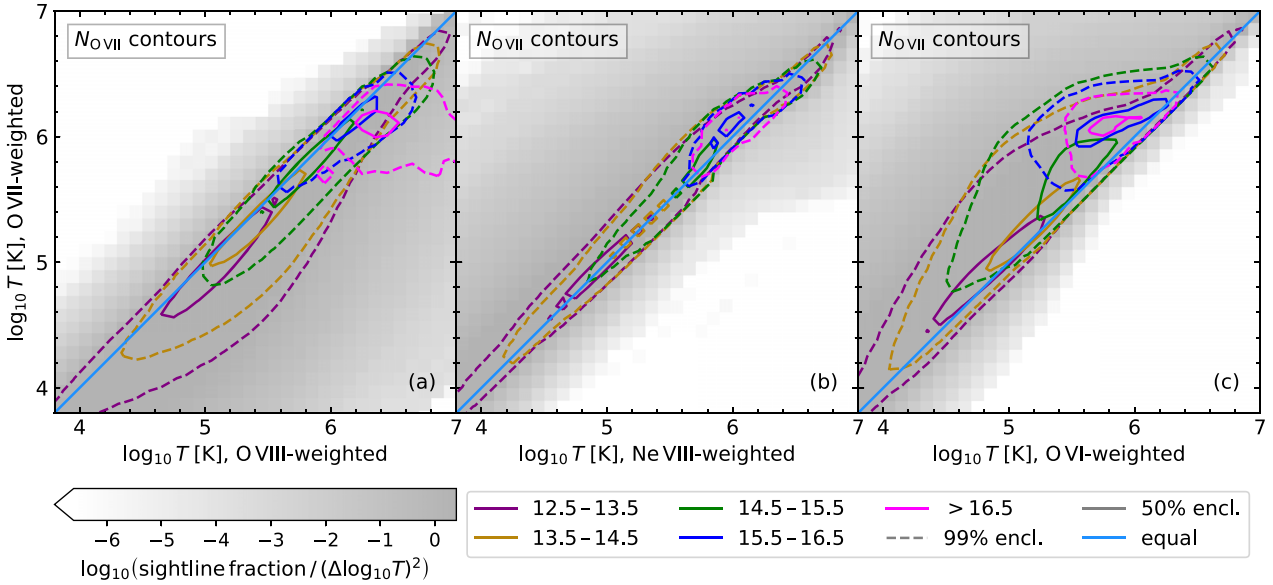

Figure 17. As Fig. 16, but for temperature. Between O VII, and O VI and O VIII, there are systematic temperature differences at all column densities, but especially the larger ones. The O VII and Ne VIII temperatures agree within 0.3 dex for $\gtrsim 98$ per cent of absorbers at all column densities shown.

4 DISCUSSION

4.1 Detection prospects

We examine the prospects for detecting O VII and O VIII line absorption with the planned mission Athena (Lumb et al. 2017; Barret et al. 2018), and the proposed missions Arcus[10] (Brenneman et al. 2016; Smith et al. 2016), and Lynx (The Lynx Team 2018).

---

[10]This was one of three (later two) proposed NASA MIDEX missions in the most recent (2019 February) funding round, but it was not chosen.

Detectability depends on the background source as well as on the absorber. Here, we consider quasars/blazars like those discussed in the mission science cases. The Arcus and Athena science cases include detecting the WHIM and measuring the CDDF for WHIM absorbers (Brenneman et al. 2016; Lumb et al. 2017). For Lynx (The Lynx Team 2018), which would launch later if approved, the focus is on characterizing this gas.

The Arcus XGS (Brenneman et al. 2016; Smith et al. 2016) should be able to find absorbers with EWs ≥ 4 mÅ at 5σ against blazars of a brightness threshold met by at least 40 known examples, at redshifts up to 0.2 with exposure times ≤ 500 ks. There are also plans to

Table 2. Expected number of absorption systems per unit redshift above a given observer-frame EW threshold $dN(> EW_{obs}) / dz$ for O VII and O VIII at different redshifts. Here, we use 6.25 cMpc slice CDDFs and 100 cMpc column density–EW relations (with scatter) at each redshift. The results come from the reference (`Ref-L100N1504`) EAGLE simulation.

EW_{obs} (mÅ)	O VII			O VIII		
	$z = 0.1$	$z = 0.2$	$z = 1.0$	$z = 0.1$	$z = 0.2$	$z = 1.0$
6.8	1.98	2.43	4.33	0.533	0.762	3.23
5.2	2.95	3.53	5.59	0.987	1.33	4.49
4.0	4.06	4.77	6.92	1.60	2.09	5.94
3.0	5.57	6.37	8.61	2.49	3.23	7.74
1.0	14.0	15.1	17.2	9.87	11.8	17.5

measure absorption against gamma-ray bursts. Fig. 6 shows that this threshold is roughly at the knee of the EW distribution. Fig. 5 shows that Arcus would mainly be probing absorption systems that do not lie on the linear part of the curve of growth with this EW regime. Smith et al. (2016) expect to find around 40 O VII absorbers at EW > 4 mÅ, using about 20 blazar background sources probing redshift paths that add up to $\Delta z \approx 8$.

We can make a similar prediction for the number of detections in such a survey. The results are shown in Table 2. At each redshift, we make predictions using EW distributions obtained from 6.25 cMpc slice CDDFs, using the (rest-frame) column density–EW relation for a full sample of 100 cMpc sightlines at each redshift. We obtain the column density–EW relation at $z = 0.2$ and $z = 1$ in the same way as described in Sections 2.3 and 3.3 for $z = 0.1$, except that we select the higher-redshift sightlines by O VII and O VIII column density alone, ignoring O VI. This should not significantly affect our results, since the ion-selected subsamples did not yield substantially different results from the full sample at $z = 0.1$.

In this calculation, changes in the predicted number of absorbers with redshift are due to four effects. First, the CDDFs evolve, as we saw in Fig. 3. The evolution of the CDDFs between redshifts 0.1 and 0.2 is small, but differences with $z = 1.0$ are larger. Secondly, the relation between column density and EW evolves. This evolution is weak, affecting the values in Table 2 by $\lesssim 4$ per cent. Thirdly, these CDDFs concern the absorption system distribution with respect to dX, not dz. Between redshifts 0.0 and 0.2, dX/dz increases from 1 to 1.30. Finally, the instrumental EW threshold applies to the observed EW. This means we can probe lower-EW systems at higher redshift, assuming e.g. quasar fluxes and backgrounds/foregrounds are equal.

For $\Delta z = 8$, we expect to see 32 (38) O VII absorption systems and 13 (17) O VIII absorption systems if redshift ≈ 0.1 (0.2) contributes most to the survey path-length. These values agree roughly with the ≈ 40 O VII absorbers expected by Smith et al. (2016). They expect 10–15 O VIII counterparts to O VII lines along these sightlines, which is compatible with the total number of O VIII absorbers we expect. From the correlations between O VII and O VIII absorption systems (Fig. 15g), and the column density–EW relation we find (Fig. 5), we expect that some of the detectable O VIII absorbers do not have observable O VII counterparts: for absorbers with $N_{O VIII} \geq 10^{15.5}$ cm^{-2}, 31 per cent (3.7 per cent) have O VII counterparts with $N_{O VII} \leq 10^{15.5}$ cm^{-2} ($\leq 10^{15}$).

As shown in Appendix C, we do not see major effects of the large-scale structure on variations in measured CDDFs in the EAGLE reference simulation (`Ref-L100N1504`). We expect the variations in measured EW distributions are Poisson as well. That would imply that the expected survey-to-survey scatter in the number of absorption systems detected is similar to the maximum effect of

varying the background source redshift distribution (as long as $z \lesssim 0.2$ still holds).

If the sensitivity can be increased to a minimum EW of 3 mÅ, we would expect to see about 1.3–1.6 times as many absorption systems at redshifts 0.1–0.2. If, for deep surveys, a sensitivity of 1 mÅ can be achieved, we would expect to find double to triple the number of O VII absorption systems and double to quadruple the number O VIII absorption systems per unit redshift as at 3 mÅ.

As for resolving line widths or line shapes, Fig. 4 shows the expected effects of line broadening and Poisson noise on observations. This figure suggests that constraining the line widths of single absorption components will be difficult, especially since what would appear to be single absorption components often are not. Comparing our approximate line widths from Fig. 5 to the spectral resolution expected for the instrument supports this.

The Arcus XGS resolving power was planned to be $\lambda/\Delta\lambda = 2000$, 2500 below and above 21.6 Å, respectively (Smith et al. 2016). For our broadest O VII lines (≈ 220 km s^{-1}, we have $\lambda/\Delta\lambda \approx 1.4 \times 10^3$ assuming a single line. For O VIII, $\lambda/\Delta\lambda \approx 1.0 \times 10^3$ for the broadest lines. The b parameters for O VIII in Fig. 5 were calculated by fitting the relation between column density and EW by the two doublet lines, each with a width b. Fitting the O VIII curve of growth using a single line instead (using the sum of oscillator strengths and oscillator-strength-weighted average wavelength) only adds about 10 km s^{-1} to the best-fitting b parameter. This would suggest we may be able to constrain some of the line widths with the Arcus XGS, though a good measurement of the line widths may be possible for only the broadest lines. Bear in mind that the b parameters we use are a rough measure, determined by the curve of growth rather than the actual structure of the absorption system, and that unresolved turbulence may cause lines to be broader than predicted from the velocity structure resolved in the simulation alone. Although resolving individual components will be difficult, Arcus will be able to decompose many systems into multiple components.

On Athena, the instrument of interest is the X-IFU (Barret et al. 2018). The science requirements for Athena as a whole, including this instrument, are described by Lumb et al. (2017). Weak line 5σ detections should be possible from EW = 0.18 eV against 'bright sources'. The plan is to measure WHIM absorption against 100 BLLacs and 100 gamma-ray bursts, to study the WHIM at $z < 1$.

The minimum EW translates to 6.8 and 5.2 mÅ (rest-frame) for O VII and O VIII, respectively, for which we show the expected number of absorption systems per unit redshift in Table 2. These limits are for 50 ks observations against a point source with a 0.5 mCrab flux in the 2–10 keV energy band ($F_{2-10 keV} = 1 \times 10^{-11}$ erg cm^{-2}s^{-1} for $\Gamma = 1.8$). These are larger EWs than expected for Arcus, though still above the knee of the EW distribution. Note that the observing time used here is ten times smaller than that of the Arcus specification, so some differences are expected. Rather than attempt to correct for this difference, we explore what Athena would see if the X-IFU is used as described.

The exact number of absorption systems expected from the survey will depend on the redshift distribution of the quasars and gamma-ray bursts used to probe the WHIM. We make predictions for Athena in the same way as for Arcus. Over the redshift range 0.1–1, Table 2 shows large changes in predicted absorber densities in redshift space. The expected number of absorption systems at the Athena sensitivity thresholds is less than expected in the Arcus survey, given the current factor 10 difference in planned exposure times, but if $\Delta z = 50$–100 is achieved, we would expect the Athena survey to

find a larger number of O VII absorption systems. Since the aim is to observe 100 BLLacs and 100 gamma-ray burst afterglows, such a survey size seems reasonable.

For the Lynx XGS (The Lynx Team 2018), a goal is to use O VII and O VIII absorption to characterize the hot CGM of galaxies with halo mass $\gtrsim 10^{12}\,\mathrm{M_\odot}$ and the hot gas in filaments with overdensities $\gtrsim 30$. For the filaments, this means looking for column densities $\gtrsim 10^{15}\,\mathrm{cm^{-2}}$ and rest-frame EWs ~ 1 mÅ. Indeed, The Lynx Team (2018) expects to be able to detect lines at EW $\gtrsim 1$ mÅ outside halo virial radii, against bright AGN ($F_{0.5-2\,\mathrm{keV}} \sim 1 \times 10^{-11}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) with an average exposure time of ≈ 60 ks. Table 2 shows the predicted number of detected O VII and O VIII absorbers in blind surveys with this sensitivity.

In the future we plan to improve on the rough estimates provided in this section by creating many virtual Arcus and X-IFU spectra, processing them through the instrument models, and analysing them as if they were real data.

4.2 Caveats

We have studied the impact of resolution and some technical choices in Appendices A and B. There are some other approximations and uncertainties we will discuss here.

First, we assume the ion fractions are set by the temperature and hydrogen number density of the gas, assuming collisional and photoionization equilibrium with a uniform (but evolving) Haardt & Madau (2001) UV/X-ray background. Cen & Fang (2006) found that there were differences in the EW distributions derived with and without equilibrium assumptions. These differences are comparable to differences between different simulations for O VII, shown in Fig. 6 and smaller than those differences for O VIII. We therefore consider ionization equilibrium a reasonable model simplification. Yoshikawa & Sasaki (2006) also discuss non-equilibrium effects, and plot at what temperatures and densities they expect non-equilibrium effects to be strongest. They found that WHIM detectability was not impacted much by non-equilibrium effects, but that line ratios were impacted more significantly, further complicating measurements of absorber temperatures. Oppenheimer et al. (2016) found using much higher-resolution simulations that non-equilibrium ionization was generally unimportant for O VI.

Another issue relating to ionization and equilibrium is the impact of local ionization sources. For these high-energy ions, we would expect AGNs to be the most important. Oppenheimer & Schaye (2013) pointed out that if, as expected, AGNs vary on time-scales similar to or shorter than the recombination time-scale in the IGM, then a large fraction of high-ionization metal absorbers may be more highly ionized than they would be if they were only ever illuminated by the metagalactic background radiation. This could then even affect gas near galaxies that currently do not show detectable AGN activity. Segers et al. (2017) found that in the inner parts of the halo, a variable AGN could have large effects on O VI column densities, with smaller effects out to $2\,R_{\mathrm{vir}}$. These effects persisted between periods where the AGN was 'on'. They did not study the effect of this on O VII and O VIII, so the precise impact here is unknown. The effect could be important in regions where the temperature is too low for collisional ionization to be effective. Here, AGN radiation could add a population of cooler, high column density systems. However, it might also affect the population of strong absorbers in our current sample, if the X-ray flux is high enough to ionize their hot, dense gas even further.

A further difficulty lies in defining absorbers. We have somewhat avoided this issue when discussing column densities, by limiting

our work to a reasonable proxy for absorption *systems*. However, in observations, observers usually fit line profiles to their spectra and gather information on each line. Although we have synthetic spectra, we have not attempted to identify and characterize individual absorption components in these, since the finding and identification of these components will generally depend on the observing instrument and exposure time, and is time-consuming. From visual inspection of Athena and Arcus mock spectra, it seems that absorption systems would usually look more or less like single lines at the spectral resolution of Arcus, and would nearly always look this way at the Athena spectral resolution. This justifies limiting our discussion to entire absorption systems, though we do plan to revisit this question in future work.

5 CONCLUSIONS

We have used the EAGLE cosmological, hydrodynamical simulations to predict the rate of incidence and physical conditions of intergalactic O VII ($\lambda = 21.6019$ Å, $E = 574$ eV) and O VIII ($\lambda = 18.9671, 18.9725$ Å, $E = 654$ eV) absorbing gas. In the largest simulation, 40 per cent of gas-phase oxygen is in the form of these two ions, making them the key tracers of cosmic metals. We have extracted CDDFs for O VII and O VIII by measuring the number of ions in a grid of long, thin columns within slices through the simulation box. Assuming a uniform metallicity instead of the predicted metallicities results in much steeper CDDFs (Fig. 2) because higher column densities are predicted to correspond to higher metallicities. The most notable feature in the CDDFs, which evolve only mildly between $z = 1$ and $z = 0$, looks like a power-law break, and occurs at column densities $\sim 10^{16}\,\mathrm{cm^{-2}}$ (Fig. 3). This break occurs where absorption systems reach overdensities $\sim 10^2$ (Fig. 7), which indicates it is likely caused by a transition from absorption by sheet and filament gas to absorption by denser halo gas.

AGN feedback substantially modifies the CDDF at high column densities. It impacts both the position of the break in the CDDF and its slope after the break, indicating an impact on haloes. Without AGN feedback, there are slightly fewer absorption systems at column densities of $\sim 10^{15}$–$10^{16}\,\mathrm{cm^{-2}}$ and significantly more at column densities $\gtrsim 10^{16}\,\mathrm{cm^{-2}}$ (Fig. 10). Before the break, this is because AGN feedback increases the metallicity of the gas. Beyond the break, AGN feedback still increases the metallicity, but column densities decrease due to the decrease in the hot gas densities. We used a large set of synthetic absorption spectra to determine the relation between EW and projected column density. The best-fitting b-parameters to the curve of growth are 90 and 158 km s^{-1} for O VII and O VIII, respectively, but the scatter in $b(N)$ is large (Fig. 5). We used the CDDF to predict the EW distribution in the EAGLE reference simulation at redshift 0.1 (Fig. 6), accounting for the scatter by using our sample's EW distribution at fixed column density in the conversion. We predict an O VII EW distribution consistent with the recent observations of Nicastro et al. (2018). By measuring the ion-weighted gas temperatures and densities along the same columns, we investigated the physical properties of our absorption systems. Overall, for 32 per cent (8 per cent) of O VII (O VIII) the ion fractions are within a factor 1.5 of the values corresponding to CIE, while for most of the remaining gas both collisional ionization and photoionization by the metagalactic background radiation are important for the ion fractions. Although O VII and O VIII often trace gas that is (partly) photoionized, O VII and O VIII absorption systems with column densities $\gtrsim 10^{16}\,\mathrm{cm^{-2}}$ mostly trace gas at temperatures close to those where their ion

fractions peak in CIE: $\sim10^6$ and $\sim10^{6.5}$ K, respectively. At column densities $\sim10^{15}$–10^{16} cm^{-2}, much of the gas is at temperatures consistent with CIE, but the ionization fractions are not the same as the purely collisional values (Figs 11 and 12). There are systematic differences between the temperatures traced by O VII and O VIII for the same absorbing structures (Fig. 17). Along with density differences, this may lead to difficulties with using column density ratios to estimate gas temperatures. Instrumental line broadening and unresolved absorption system structure mean that temperatures will also typically not be well-constrained by line widths.

On the other hand, the column densities of O VII and O VIII are correlated with each other, and with those of ions with detectable absorption lines in the far UV (Figs 14 and 15). The presence of neutral hydrogen absorption with column density $\gtrsim10^{15}$ cm^{-2} implies a ~50 per cent chance of finding an O VII or O VIII absorber with column density $\gtrsim10^{15}$ cm^{-2}, but the probability does not increase further for even stronger H I absorption (until ISM densities are reached). High O VI or Ne VIII column densities are good predictors of the presence of detectable O VII and O VIII column densities (Fig. 14). This suggests that pre-selection based on these UV lines is an efficient (though necessarily biased) search strategy for the X-ray lines.

ACKNOWLEDGEMENTS

REFERENCES

Allende Prieto C., Lambert D. L., Asplund M., 2001, ApJ, 556, L63
Bahcall J. N., Peebles P. J. E., 1969, ApJ, 156, L7
Barret D. et al., 2016, in den Herder J.-W. A., Takahashi T., Bautz M., eds, Proc. SPIE Conf. Ser. Vol. 9905, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray. SPIE, Bellingham, p. 99052F
Barret D. et al., 2018, in den Herder J.-W. A., Nikzad S., Nakazawa K., eds, Proc. SPIE Conf. Ser. Vol. 10699, Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray. SPIE, Bellingham, p. 106991G
Bertone S., Schaye J., Dalla Vecchia C., Booth C. M., Theuns T., Wiersma R. P. C., 2010a, MNRAS, 407, 544
Bertone S., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., Wiersma R. P. C., 2010b, MNRAS, 408, 1120
Bonamente M., Nevalainen J., Tilton E., Liivamägi J., Tempel E., Heinämäki P., Fang T., 2016, MNRAS, 457, 4236
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Borgani S. et al., 2004, MNRAS, 348, 1078
Branchini E. et al., 2009, ApJ, 697, 328
Bregman J. N., 2007, ARA&A, 45, 221
Brenneman L. W. et al., 2016, in den Herder J.-W. A., Takahashi T., Bautz M., eds, Proc. SPIE Conf. Ser. Vol. 9905, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray. SPIE, Bellingham, p. 99054P
Burchett J. N. et al., 2019, ApJ, 877, L20
Cen R., 2012, ApJ, 753, 17
Cen R., Fang T., 2006, ApJ, 650, 573
Cen R., Ostriker J. P., 1999, ApJ, 514, 1
Cen R., Ostriker J. P., 2006, ApJ, 650, 560
Collette A., 2013, Python and HDF5. O'Reilly. http://www.h5py.org/
Crain R. A. et al., 2015, MNRAS, 450, 1937
Cyburt R. H., Fields B. D., Olive K. A., Yeh T.-H., 2016, Rev. Mod. Phys., 88, 015004
Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140
Davies J. J., Crain R. A., McCarthy I. G., Oppenheimer B. D., Schaye J., Schaller M., McAlpine S., 2019, MNRAS, 485, 3783
de Graaff A., Cai Y.-C., Heymans C., Peacock J. A., 2019, A&A, 624, A48
Dubois Y. et al., 2014, MNRAS, 444, 1453
Fang T., Bryan G. L., Canizares C. R., 2002, ApJ, 564, 604
Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761
Furlanetto S. R., Phillips L. A., Kamionkowski M., 2005, MNRAS, 359, 295
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, preprint (arXiv:astro-ph/0106018)
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Johnson S. D., Chen H.-W., Mulchaey J. S., 2013, MNRAS, 434, 1765
Jones E. et al., 2001, SciPy: Open source scientific tools for Python. http://www.scipy.org/
Kaastra J. S., Werner N., Herder J. W. A. d., Paerels F. B. S., de Plaa J., Rasmussen A. P., de Vries C. P., 2006, ApJ, 652, 189
Kovács O. E., Bogdán Á., Smith R. a. K., Kraft R. P., Forman W. R., 2019, ApJ, 872, 83
Lide D. R., ed., 2003, CRC Handbook of Chemistry and Physics, 84 edn. CRC Press LLC, Boca Raton
Lumb D., den Herder J.-W., the Athena Science Team, 2017, Issue/Revision 2.01, Athena Science Requirements Document. European Space Agency, European Space Research and Technology Centre, Keplerlaan 1, AZ, Noordwijk, The Netherlands
McAlpine S. et al., 2016, Astron. Comput., 15, 72
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
Meiring J. D., Tripp T. M., Werk J. K., Howk J. C., Jenkins E. B., Prochaska J. X., Lehner N., Sembach K. R., 2013, ApJ, 767, 49
Monaghan J. J., Lattanzio J. C., 1985, A&A, 149, 135
Mroczkowski T. et al., 2019, Space Sci. Rev., 215, 17
Nelson D. et al., 2018, MNRAS, 477, 450
Nicastro F., 2018, preprint (arXiv:1811.03498)
Nicastro F. et al., 2005, ApJ, 629, 700
Nicastro F., Krongold Y., Mathur S., Elvis M., 2017, Astron. Nachr., 338, 281
Nicastro F. et al., 2018, Nature, 558, 406
Oliphant T. E., 2006, Guide to NumPy. Trelgol Publishing, USA
Oppenheimer B. D., Schaye J., 2013, MNRAS, 434, 1063
Oppenheimer B. D. et al., 2016, MNRAS, 460, 2157
Oppenheimer B. D., Segers M., Schaye J., Richings A. J., Crain R. A., 2018, MNRAS, 474, 4740
Oppenheimer B. D. et al., 2019, MNRAS, preprint (arXiv:1904.05904)
Pérez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Perna R., Loeb A., 1998, ApJ, 503, L135
Pierre M., Bryan G., Gastaud R., 2000, A&A, 356, 403
Pillepich A. et al., 2018, MNRAS, 473, 4077

Planck Collaboration I, 2014, A&A, 571, A1

Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427

Rahmati A., Schaye J., Crain R. A., Oppenheimer B. D., Schaller M., Theuns T., 2016, MNRAS, 459, 310

Rohatgi A., 2018, WebPlotDigitizer. https://automeris.io/WebPlotDigitizer

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., 2004, ApJ, 609, 667

Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210

Schaye J. et al., 2010, MNRAS, 402, 1536

Schaye J. et al., 2015, MNRAS, 446, 521

Segers M. C., Oppenheimer B. D., Schaye J., Richings A. J., 2017, MNRAS, 471, 1026

Shull J. M., Smith B. D., Danforth C. W., 2012, ApJ, 759, 23

Smith R. K. et al., 2016, in den Herder J.-W. A., Takahashi T., Bautz M., eds, Proc. SPIE Conf. Ser. Vol. 9905, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray. SPIE, Bellingham, p. 99054M

Springel V., 2005, MNRAS, 364, 1105

Tanimura H. et al., 2019, MNRAS, 483, 223

Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., Wiersma R. P. C., 2011, MNRAS, 413, 190

Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., 2012, MNRAS, 425, 1640

Tepper-García T., Richter P., Schaye J., 2013, MNRAS, 436, 2063

The Lynx Team, 2018, preprint (arXiv:e-print)

Tumlinson J. et al., 2011, Science, 334, 948

Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389

Verner D. A., Verner E. M., Ferland G. J., 1996, At. Data Nucl. Data Tables, 64, 1

Wendland H., 1995, Adv. Comput. Math., 4, 389

Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574

Yoshikawa K., Sasaki S., 2006, PASJ, 58, 641

Yoshikawa K., Yamasaki N. Y., Suto Y., Ohashi T., Mitsuda K., Tawara Y., Furuzawa A., 2003, PASJ, 55, 879

Figure A1. Convergence of the CDDF (*f(N)*) with pixel size. The left-hand panels show O VII, the right-hand panels show O VIII. In all panels, the solid line shows the CDDF as used throughout the paper: the column density is calculated on a grid of $32\,000^2$ pixels, for each of 16 slices through the 100 cMpc the box along the projection axis. These CDDFs are from the reference (Ref-L100N1504) simulation at $z = 0.1$. In the top panels (a, b), this CDDF is compared to that obtained by degrading the resolution of the 3.125^2 ckpc^2 pixel grid to get grids of 6.25^2 (dashed), 12.5^2 (dot–dashed), and 25^2 ckpc^2 (dotted) pixels. The legends indicate the pixel sizes. In the bottom panels (c, d), the column density distributions extracted from two smaller regions (10×10 cMpc^2) at this standard resolution (3.125^2 ckpc^2 per pixel) are compared with the CDDFs from the same region at nine time higher resolution. The CDDFs at different resolution are shown relative to the highest-resolution CDDF for the same region.

APPENDIX A: CDDF CONVERGENCE

In this appendix, we test the convergence of the CDDFs with pixel size, slice thickness, box size, and simulation resolution, and give more details on how we tested the effect of the projection kernel shape. The default projection we will compare other CDDFs to is that obtained from the reference simulation (Ref-L100N1504) at redshift 0.1, with 16 slices of 6.25 cMpc each along the *z*-axis, and $32\,000^2$ pixels in the *x*- and *y*-directions. The assumed gas distribution for one SPH particle in this 'standard' CDDF is given by the C2-kernel (Wendland 1995) used in the simulation. The temperature of star-forming gas is set to 10^4 K for the ion fraction calculation. Any properties of CDDFs that are not otherwise mentioned here, are as described above.

In Fig. A1, we look at the effect of changing the pixel size. Using lower resolution means underestimating the fraction of columns with large column densities, while assessments of lower column densities remain accurate. Our CDDFs appear to be converged at our standard pixel size for column densities $\lesssim 10^{16.5}$ cm^{-2}. At higher resolutions than our standard, we see problems arise at a similar column density.

In Fig. A2, we investigate the effect of using slices of different thickness to calculate the column densities. The clearest effects are at low column densities. Thinner slices result in larger values of *f(N)*, which suggests that at low column densities, absorption often comes from multiple systems along the same line of sight. An optimal slice thickness is difficult to choose; we want to use a size which is small
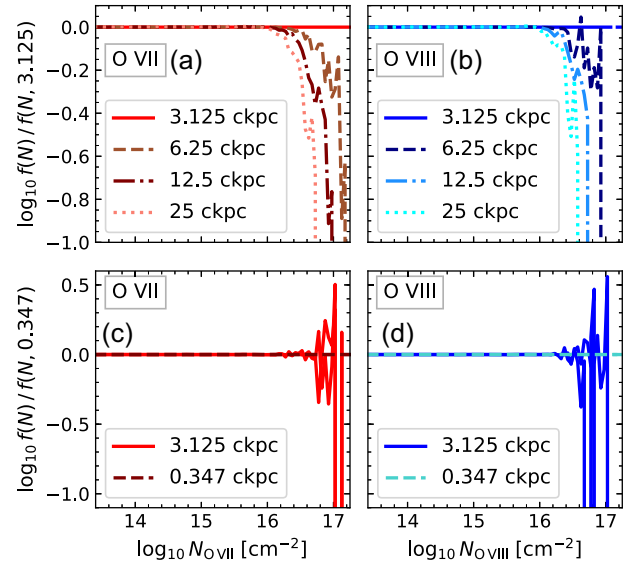
enough that the absorption in one column is (mostly) coming from a single system, but large enough that single systems are (mostly) not split over two columns along the line of sight. We choose to use slices of 6.25 cMpc, which matches UV-ion CDDFs well below the CDDF break (Rahmati et al. 2016), but bear in mind that this choice of, essentially, how to define absorption systems affects our results somewhat. Note, however, that the differences between CDDFs of different slice thickness are small compared to the range spanned by the CDDF. This standard slice thickness corresponds to 404 km s^{-1} in velocity space ($z = 0.1$, rest-frame) when only accounting for the Hubble flow.

Next, we investigate whether the CDDFs are converged in the simulations themselves. Specifically, we investigate the effect of simulation volume and resolution. In Fig. A3, we compare different simulations to each other. Here, differences may also come from differences in initial conditions (cosmic variance). In the top panels, we examine the effect of box size. In smaller boxes, the distributions are noisier because they are less well sampled, but there is also a physical difference: smaller boxes do not sample some of the large-scale modes of the initial density variations, so smaller boxes will contain fewer high-mass, large, hot systems. The 100 cMpc box has no haloes with masses $M_{200c} > 10^{15}$ M$_\odot$, and only seven with $M_{200c} > 10^{14}$ M$_\odot$, so we do not expect complete convergence with box size. However, the figure indicates that, at column densities $\lesssim 10^{16.5}$ cm^{-2}, convergence with box size is sufficient. The $\lesssim 0.2$ dex difference between the CDDFs for the
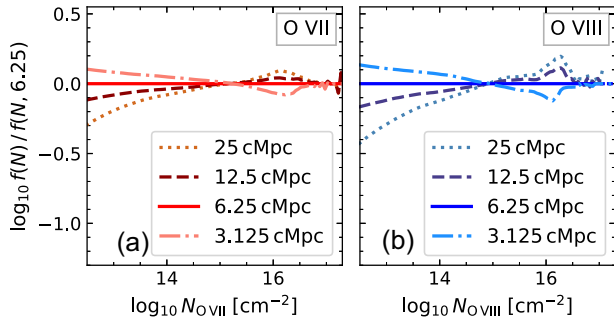
Figure A2. Convergence of the CDDF ($f(N)$) with slice thickness. The left-hand panel shows O VII, the right-hand panel shows O VIII. In both panels, the solid line shows the CDDF as used throughout the paper: the column density is calculated on a grid of $32\,000^2$ pixels, for each of 16 slices the 100 cMpc box along the projection axis. For the other lines, the thickness of the box slices for which the column density is calculated is varied as indicated. These CDDFs are from the reference (`Ref-L100N1504`) simulation at $z = 0.1$. The differences between our default 6.25 cMpc slice CDDF and CDDFs using thicker or thinner slices are mostly $\lesssim 0.2$ dex, and always $\lesssim 0.2$ dex in this column density range when comparing to 3.125 and 12.5 cMpc slices. These differences are small compared to the more than 8 orders of magnitude the CDDFs span in this column density range.
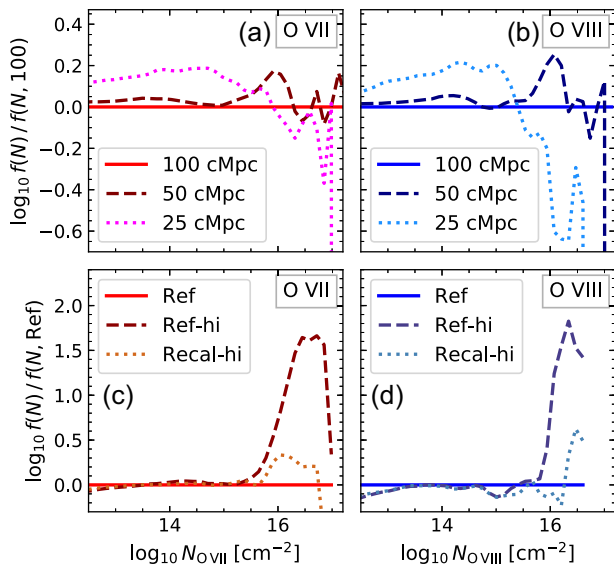

Figure A3. Convergence of the CDDF with simulation box size (top panels) and resolution (bottom panels). The left-hand panels show O VII, the right-hand panels show O VIII. The plots show the CDDF $f(N, z = 0.1)$, as described in equation (1), relative to a reference CDDF. In the top panels, the CDDFs from `Ref-L025N0376` and `Ref-L050N0752` are compared to the standard `Ref-L100N1504` CDDF. In the bottom panels, `Ref-L025N0376` ('Ref') is compared to the `Ref-L025N0752` and `Recal-L025N0752` ('Ref-hi', 'Recal-hi') CDDFs.

largest two boxes is much smaller than the range spanned by the CDDFs in the same column density range, and larger differences occur at column densities where the CDDF is not converged with projection resolution.

In the bottom panels, we investigate the effect of simulation resolution (distinct from the projection resolution we investigated in Fig. A1). We compare three 25 cMpc boxes, since we do not have higher-resolution simulations of larger boxes. The compari-

son to `Ref-L025N0752` tests strong convergence and `Recal-L025N0752` tests weak convergence in the terminology of Schaye et al. (2015). The `Ref-L025N0752` simulation was run using the exact same parameters as `Ref-L025N0376`, except for the resolution, and the addition of initial density perturbations on smaller scales in the higher-resolution box. However, the stellar and AGN feedback in the EAGLE simulation is calibrated to reproduce the redshift 0.1 galaxy stellar mass function, the relation between black hole mass and galaxy mass, and reasonable galaxy sizes. This calibration depends on the simulation resolution, so Schaye et al. (2015) also recalibrated the subgrid model for feedback to produce a similarly calibrated simulation at higher resolution: `Recal-L025N0752`. We consider this the most relevant comparison.

The weak convergence of the CDDFs is particularly good: for column densities $< 10^{16.3}$ cm^{-2}, differences are $\lesssim 0.3$ dex. They get larger at higher column densities (particularly for O VIII), but this is in a regime where the CDDFs in the 25 cMpc box are poorly converged compared to the larger boxes (top panels of Fig. A3), and the behaviour of the CDDFs here may therefore not be representative of what the effect of higher simulation resolution might be in a larger box.

Finally, as we mentioned in Section 2.2, we found that the effect of the choice of SPH kernel is small. Here, we give some more details on the kernels we compared. Our default is the Wendland (1995) C2 kernel, with shape k given by

$$k_{C2}(u) = (1 - u)^4 (1 + 4u), \tag{A1}$$

where u is the distance to the particle position normalized by the smoothing length of the particle. The kernel is zero at $u > 1$. Similarly, the Gadget kernel [used by Springel (2005), from Monaghan & Lattanzio (1985)] is described at $u \leq 1$ by

$$k_{Gadget}(u) = \begin{cases} 2.5465 + 15.2789\,(u - 1)\,u^2 & \text{if } u < 0.5 \\ 5.0930\,(1 - u)^3 & \text{if } 0.5 \leq u \leq 1. \end{cases} \tag{A2}$$

Both kernels are normalized to unity when integrated over the surface of a unit circle to ensure mass conservation in the projection. The C2 kernel is more centrally concentrated. When both kernels are normalized, their absolute difference is largest at the centre of the distribution: $k_{C2}(0) = 2.23$ and $k_{Gadget}(0) = 1.82$. The maximum relative difference is at the edge of the distribution, and diverges as $u \to 1$ due to the different slopes of the kernels near zero. The differences between the CDDFs we obtain using these kernels is $\lesssim 0.05$ dex for 10^{11} cm$^{-2} < N_{\text{O VII, VIII}} < 10^{16.5}$ cm^{-2}.

APPENDIX B: TECHNICAL CHOICES FOR THE EW DISTRIBUTION

Fig. B1 shows the rest-frame EW distribution of O VII and O VIII calculated from the CDDF using different methods. In the first two, we impose a one-to-one relation between column density and EW. We use the linear curve of growth or the curve of growth with the best-fitting b parameter to the relation between log EW and column density obtained for a sample of absorption spectra. The third method, which is our fiducial model, accounts for scatter in the relation between column density and EW by using a matrix generated from the points in the column density–EW relation from Fig. 5. We obtain this matrix by choosing column density and EW bins, then making a histogram of our sample of sightlines in column density and EW. After normalization, we use this matrix to convert our column density histograms into EW histograms.

The differences between the EWs calculated using the different methods are as expected. The matrix conversion is our preferred
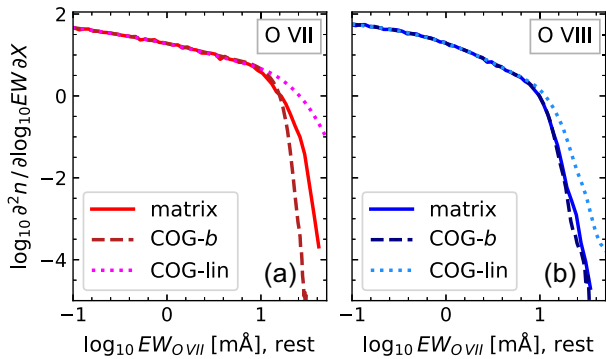
Figure B1. A comparison of methods to obtain the rest-frame EW distribution from the CDDF. The rest-frame EW distributions for O VII, $\lambda = 21.60$ Å (a) and the O VIII, $\lambda = 18.97$ Å doublet (b) in the `Ref-L100N1504` simulation at redshift 0.1 are shown. The lines labelled 'COG-*b*' are for EWs calculated from the CDDF using a single curve of growth with the best-fitting *b* parameters to the relation between log EW and the column density obtained for a set of absorption spectra. The dotted line ('COG-lin') is for EWs calculated using the linear curve of growth for each ion. For the lines labelled 'matrix', the conversion from CDDF to EW distribution was done using a matrix generated from the column density–EW relation in Fig. 5, which accounts for scatter. We use the CDDF for 6.25 cMpc slices.

method, because it models the scatter. Since the linear curve of growth is only valid for unsaturated absorption, it gives the maximum EW for any column density. This method overestimates the number of high EW systems (EW $\gtrsim$ 10 mÅ). The difference between the matrix and best-fitting curve-of-growth methods is whether scatter is included. Including scatter increases the predicted number of absorption systems at the highest EWs because the column density distribution is steep and declining at the highest column densities. This means that scatter at fixed column density attributes larger EWs to more lower column density absorption systems than it attributes smaller EWs to larger column density absorption systems, in absolute terms. This is why the single best-fitting *b*-parameter curve of growth conversion underestimates the number of high EW absorption systems. The discrepancy is larger for O VII than for O VIII, because, as Fig. 5 shows, O VII EWs have a larger scatter at fixed column density and have significant scatter over a larger range of column densities than O VIII EWs. This means that including scatter in the modelling of the EW distribution has a larger effect for O VII.

APPENDIX C: MOCK SURVEY CDDFS

We are interested in the variation in measured CDDFs for different survey sizes. In particular, we investigate whether clustering along the line of sight on scales of 6.25–50 cMpc causes deviations from Poisson statistics, which we find is not the case. We measure survey size by the absorption distance ΔX, as given by equation (2). We measure this at redshift 0.1, and use the Hubble flow across the box to calculate dz. We conduct mock surveys of the 100 cMpc box, of sizes $\Delta X = 0.5, 1.0., 10, 100, 1000$. To create one mock survey, we choose (x, y) positions at random, then take the column densities for all the 6.25 cMpc columns along the z-axis at that position. This mimics searching for absorption systems along longer sightlines than we use to measure the column densities. Note that due to the size (and periodicity) of the box, and the extent over which we measure a single column density, this only probes the effect of large-scale structure correlations for separations between 6.25 and 50 cMpc.

Because we are interested in the statics of these surveys, we conduct 100 such mock surveys for O VII and O VIII for each ΔX. For comparison, we also create 'random' surveys of the same size, where for each random (x, y), we select a single 6.25 cMpc slice at random. We make cumulative distributions for these surveys. Then at each threshold column density, we considered the distribution in absorption system counts among the different mock surveys. We show the 10th (blue), 50th (green, median), and 90th (red) percentiles of this distribution as the dashed lines in Fig. C1.

In Fig. C1, we do not show theoretical distributions, but the random surveys described above. To probe what range in these percentiles is consistent with variation between survey samples, we use ten different samples of 100 random surveys. We plot these in Fig. C1 with colours matching the same percentiles for the mock surveys.

The figure shows that the mock surveys for the different ions and survey sizes are consistent with the random surveys: in other words, large-scale structure on the scales we can probe does not appear to affect the survey statistics for measuring CDDFs. We also checked the relative differences between these percentiles and Poisson and binomial distributions: these were centred on zero, and the mock survey distributions differed from the theoretical values by amounts consistent with the random surveys. Using the CDDF instead of the cumulative distribution gives the same results.

Note that Fig. C1 can also be used to estimate a reasonable range of detected absorption systems above a given column density threshold for different survey sizes.
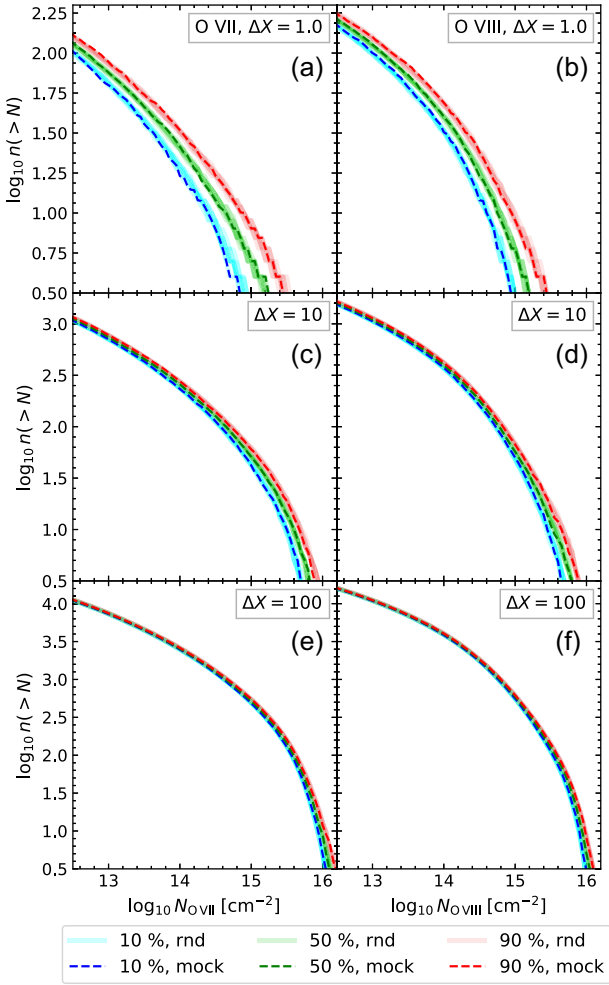
Figure C1. The cumulative column density distributions for mock surveys of sizes $\Delta X = 1.01$ ($\Delta z = 0.878$; a, b), $\Delta X = 10.0$ ($\Delta z = 8.66$; c, d), and $\Delta X = 100$ ($\Delta z = 86.7$; e, f), of O VII (a, c, e) and O VIII (b, d, f) in the reference (Ref-L100N1504) simulation at redshift 0.1. The column densities are calculated for 16 6.25 cMpc slices, each with 3.125^2 ckpc^2 pixels. The dashed lines show mock surveys, where (x, y) positions were selected at random, and then each slice along the z-axis at that position was included, mimicking the measurement of different absorption systems along full 100 cMpc sightlines. This accounts for clustering along the line of sight on scales between 6.25 and 50 cMpc (i.e. half the size of the periodic box). We obtained a sample of 100 such mock surveys. At each column density, show the 10th (blue), 50th (green, median), and 90th (red) percentiles of the distribution of absorber counts from the different mock surveys. To find the effect of large-scale structure on this distribution, we repeated this process for random surveys, show in lighter solid colours, where we select the same number of 6.25 cMpc columns, but choose them randomly. For these random surveys, we show the percentiles for ten different sets of 100 random surveys. The good agreement between the solid and dashed curves indicates that large-scale structure does not seem to affect the statistical variation in CDDFs that surveys may measure on scales of 6.25–50 cMpc.

This paper has been typeset from a TeX/LaTeX file prepared by the author.