Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **495,** 2135–2151 (2020)
Advance Access publication 2020 May 11

doi:10.1093/mnras/staa1313

QSO obscuration at high redshift ($z \gtrsim 7$): predictions from the BLUETIDES simulation

Yueying Ni,[1]★ Tiziana Di Matteo,[1] Roberto Gilli,[2] Rupert A. C. Croft,[1] Yu Feng[3] and Colin Norman[4,5]

[1]*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[2]*INAF Osservatorio Astronomico di Bologna, Via Gobetti 93/3, I-40129 Bologna, Italy*
[3]*Berkeley Center for Cosmological Physics and Department of Physics, University of California, Berkeley, CA 94720, USA*
[4]*Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA*
[5]*Johns Hopkins University – Center for Astrophysical Sciences, 3400 N. Charles Street, Baltimore, MD 21218, USA*

Accepted 2020 May 6. Received 2020 May 5; in original form 2019 December 6

ABSTRACT

High-z AGNs hosted in gas-rich galaxies are expected to grow through significantly obscured accretion phases. This may limit or bias their observability. In this work, we use BLUETIDES, a large volume cosmological simulation of galaxy formation to examine quasar obscuration for the highest redshift ($z \geq 7$) supermassive black holes residing in the centre of galaxies. We find that for the bright quasars, most of the high-column density gas (>90 per cent) resides in the innermost regions of the host galaxy (typically within <10 ckpc), while the gas in the outskirts is a minor contributor to the $N_{\rm H}$. The brightest quasars can have large angular variations in galactic obscuration, over 2 orders of magnitude (ranging from column density $N_{\rm H} \sim 10^{21.5\text{–}24}\,{\rm cm}^{-2}$), where the lines of sight with the lowest obscuration are those formed via strong gas outflows driven by AGN feedback. The obscured fraction P ($N_{\rm H} > 10^{23}\,{\rm cm}^{-2}$) typically ranges from 0.6 to 1.0 for increasing $L_{\rm X}$ (with $L_{\rm X} > 10^{43}\,{\rm erg\,s}^{-1}$), with no clear trend of redshift evolution. Due to the angular variation in $N_{\rm H}$, all relations between $N_{\rm H}$ and $L_{\rm X}$, $M_{\rm BH}$, and galaxy host properties (global M_*, $M_{\rm H_2}$, and star formation rate) show appreciable scatter. The dust optical depth in the UV band $\tau_{\rm UV}$ has tight positive correlation with $N_{\rm H}$. Our dust-extincted UV luminosity function (UVLF) is about 1.5 dex lower than the intrinsic UVLF, implying that more than 99 per cent of the $z \sim 7$ AGNs are heavily dust extincted and therefore would be missed by the UV-band observation.

Key words: galaxies: formation – galaxies: high-redshift – quasars: supermassive black holes.

1 INTRODUCTION

Understanding the origin of the first supermassive black holes (SMBHs) powering the most luminous quasars (QSOs) at $z > 6$ is one of the greatest observational and theoretical challenges. To understand the formation and growth of these SMBHs as well as their co-evolution with their host galaxy, one needs to consider that, while they are extremely massive for this early epoch, their space density is extremely low. The inferred rarity of these high-z QSOs may, however, be partly due to large amounts of star-forming gas in these early host galaxies that may act to obscure an active galactic nucleus (AGN).

The obscuring medium for AGNs is typically composed of dust and/or gas. Dust is the dominant source of obscuration in the UV–IR bands while gas dominates the absorption at X-ray energies (Hickox & Alexander 2018). For moderate column densities, detection of the X-ray emission is an efficient way to reveal the presence of an AGN (that is obscured in the UV). However, when the equivalent neutral hydrogen column exceeds the unit optical depth corresponding to the Thompson cross-section ($N_{\rm H} > 1.5 \times 10^{24}\,{\rm cm}^{-2}$), such 'Compton-thick' AGNs are hard to detect even in X-ray surveys.

The obscuration of AGNs can occur over a range of scales and physical conditions. In the standard unification scheme (e.g. Antonucci 1993; Urry & Padovani 1995), AGNs are surrounded by an optically thick toroidal structure a few parsec from the central BH. This torus then obscures the line of sight depending on the viewing angle. Alternatively, or additionally, obscuration can be produced by gas on the scale of the entire galaxy (>kpc) (Maiolino & Rieke 1995). High-redshift AGNs ($z = 1$–3) residing in star-forming galaxies are often rich in gas (see e.g. Tacconi et al. 2013), and the obscuration is thought to be caused by the high-

★ E-mail: yueyingn@andrew.cmu.edu

density gas fuelling the star formation (SF) as well as the accretion process. Recently, X-ray observations of $z > 2.5$ obscured AGNs (Circosta et al. 2019; D'Amato et al. 2020) indeed confirm this picture, whereby at high redshifts, the host interstellar medium (ISM; at kpc scale) is dense enough to produce the inferred amount of X-ray absorption, explaining also the increased fraction of obscured AGNs at high-z.

The recent observations of the highest redshift, $z > 7$, quasars imply BH masses of order $10^9\,M_\odot$. In order to grow to such high mass in the first few 10^8 yr of the universe, those SMBHs need to undergo continuous near-Eddington or even super-Eddington accretion, during which they are enshrouded by accreting gas with a column density that exceeds even the Compton-thick level (e.g. Pezzulli et al. 2017). Observations of Lyα absorption profiles from two $z > 7$ QSOs (Davies, Hennawi & Eilers 2019) imply that they might have experienced highly obscured growth and could be obscured in more than 82 per cent of their lifetimes (with assumption of similar radiative efficiency as the low-redshift QSOs.)

Different studies have been carried out to assess the fraction of obscured AGNs (including Compton-thick AGNs) as a function of AGN luminosity and redshift (see Hickox & Alexander 2018, for a review). The correlation between obscured fraction and luminosity is still under debate, however. Some studies of lower redshift ($z < 3$) quasars indicate that the obscured AGN fraction decreases towards higher luminosity (e.g. Merloni et al. 2014) or higher Eddington ratio (Ricci et al. 2017), while there are also works that claim that the luminosity dependence is not quite significant (e.g. Mateos et al. 2017). There is also evolution of obscured AGN fraction with redshift: in particular, the obscured fraction of $3 < z < 6$ luminous AGNs is found to increase when going to higher redshift (Vito et al. 2014, 2018).

The studies referenced above of the obscured AGN fraction are mostly based on the AGN population at $z \lesssim 5$. There is rapid ongoing progress in the detection of $z > 6$ and even $z > 7$ QSOs from observations: More than 200 QSOs have been discovered beyond $z = 6$ (see e.g. Fan et al. 2019, and references therein), and a handful found with $z > 7$ (Mortlock et al. 2011; Bañados et al. 2018; Wang et al. 2018; Matsuoka et al. 2019; Yang et al. 2019). Although such studies represent a huge breakthrough, the current detections of these $z > 7$ quasars are mostly obtained from wide-field optical/near-IR surveys, which as a selection method is strongly biased against highly obscured systems (e.g. see discussion in Vito et al. 2019). Studies of AGN populations at lower redshift ($0 < z < 5$) show that the vast majority of the AGN population is obscured (Ueda et al. 2014; Buchner et al. 2015). Recently, the first heavily obscured $z = 6.5$ AGN candidate has been discovered through the *Chandra* X-ray survey (Vito et al. 2019). The hardness of the X-ray photometry leads to an inferred galactic absorption of up to $N_{\rm H} > 2 \times 10^{24}$ and $6 \times 10^{23}\,{\rm cm}^{-2}$ at the 68 and 90 per cent confidence levels, respectively.

Making theoretical predictions for the obscured fraction of the quasar population is essential to reach a complete census of high-z QSOs and further understand the formation and growth of SMBHs in the early Universe. In this paper, we use the BLUETIDES simulation to study the obscuration state of $z > 7$ QSOs in a ΛCDM universe. BLUETIDES is a cosmological hydrodynamic simulation targeted at the study of the first generation of galaxies and QSOs in the high-z universe (Di Matteo et al. 2017; Ni et al. 2018; Tenneti et al. 2018). The large volume and high resolution of BLUETIDES make it ideally suited to study the rare high-z luminous QSOs, along with the detailed structure and physical properties of their surrounding gas. So far, BLUETIDES has been tested against various observations of the high-z universe and has been shown to be in good agreement with all current observational constraints, such as the UV luminosity functions (Feng et al. 2016; Waters et al. 2016a,b; Wilkins et al. 2017), the first galaxies and the most massive quasars (Feng et al. 2015; Di Matteo et al. 2017; Tenneti et al. 2018), the Lyman continuum photon production efficiency (Wilkins et al. 2016, 2017), galaxy stellar mass functions (Wilkins et al. 2018), angular clustering amplitude (Bhowmick et al. 2018), BH–galaxy scaling relations (Huang et al. 2018), and gas outflows from the $z = 7.54$ quasar (Ni et al. 2018).

In this work, we focus on the galactic obscuration that is due to ISM gas in the QSO host galaxies. Numerical zoom-in simulations can better resolve the obscuring gas in the nuclear region of SMBH (see e.g. Hopkins et al. 2016; Lupi et al. 2019; Trebitsch, Volonteri & Dubois 2019). In particular, Trebitsch et al. (2019) have found that the gas from the host galaxy contributes to the total obscuration at a level at least comparable to the gas in the nuclear region. Given the significant role played by galactic obscuration for high-z QSOs, the tens of thousands of SMBHs contained in the BLUETIDES simulation at $z > 7$ enable us to study the statistics of galactic obscuration and the emerging BH population at high redshifts. We are particularly interested in how current detections of high-z quasars may be hampered by obscuration.

This paper is organized as follows. In Section 2, we briefly summarize the subgrid models applied in BLUETIDES and describe the model we use to calculate the quasar and galaxy luminosity. We also introduce the way we calculate the gas and dust obscuration through AGN lines of sight. Section 3 shows and discusses our key results. In Section 4, we conclude the paper.

2 METHOD

2.1 BLUETIDES simulation

The BLUETIDES[1] cosmological simulation (Feng et al. 2016) uses the Pressure Entropy Smoothed Particle Hydrodynamics code MP-Gadget to model the evolution of a $400\,{\rm Mpc}\,h^{-1}$ side box with 2×7040^3 particles. The simulation evolved from $z = 99$ and has now reached $z < 7$. The cosmological parameters used are from the 9-yr Wilkinson Microwave Anisotropy Probe (Hinshaw et al. 2013) ($\Omega_0 = 0.2814$, $\Omega_\Lambda = 0.7186$, $\Omega_{\rm b} = 0.0464$, $\sigma_8 = 0.82$, $h = 0.697$, $n_{\rm s} = 0.971$). The mass resolution of BLUETIDES is $M_{\rm DM} = 1.2 \times 10^7\,{\rm M}_\odot\,{\rm h}^{-1}$ and $M_{\rm gas} = 2.4 \times 10^6\,{\rm M}_\odot\,{\rm h}^{-1}$ in initial condition. The mass of the star particle is $M_* = 1/4 M_{\rm gas} = 6 \times 10^5\,{\rm M}_\odot\,{\rm h}^{-1}$. The gravitational softening length is $1.8\,{\rm ckpc}\,h^{-1}$ for both DM and gas particles.

BLUETIDES implements a variety of subgrid models to model galaxy formation and different feedback processes. Here we briefly list some of its basic features, and we refer the reader to the original papers (Feng et al. 2016) for detailed descriptions. In the simulations, gas is allowed to cool through both radiative processes (Katz, Hernquist & Weinberg 1999) and metal cooling. The metal cooling rate is obtained by scaling a solar metallicity template according to the metallicity of gas particles, following the method described in Vogelsberger et al. (2014). SF is based on a multiphase SF model (Springel & Hernquist 2003) with modifications following Vogelsberger et al. (2013). We model the formation of molecular hydrogen and its effects on SF at low metallicity according to the prescription of Krumholz & Gnedin

[1] http://BlueTides-project.org/

(2011). We self-consistently estimate the fraction of molecular hydrogen gas from the baryon column density, which in turn couples the density gradient to the SF rate. Type II supernova wind feedback (the model used in Illustris; Nelson et al. 2015) is included, assuming wind speeds proportional to the local one-dimensional dark matter velocity dispersion. The large volume of BLUETIDES also allows us to include a model of 'patchy reionization' (Battaglia et al. 2013), yielding a mean reionization redshift $z \sim 10$, and incorporating the UV background estimated by Faucher-Giguère et al. (2009).

In our simulation, we model BH growth and AGN feedback in the same way as in the *MassiveBlack I&II* simulations, using the BH subgrid model developed in Springel, Di Matteo & Hernquist (2005) and Di Matteo, Springel & Hernquist (2005) with modifications consistent with *Illustris*. BHs are seeded with an initial seed mass of $M_{\rm seed} = 5 \times 10^5\, h^{-1}\, {\rm M}_\odot$ (commensurate with the resolution of the simulation) in haloes with mass more than $5 \times 10^{10}\, h^{-1}\, {\rm M}_\odot$. The gas accretion rate on to BH is given by Bondi accretion rate

$$\dot{M}_{\rm B} = \frac{4\pi G^2 M_{\rm BH}^2 \rho}{\left(c_{\rm s}^2 + v_{\rm rel}^2\right)^{3/2}}, \tag{1}$$

where $c_{\rm s}$ and ρ are the local sound speed and density of the cold gas, respectively, and $v_{\rm rel}$ is the relative velocity of the BH to the nearby gas. We allow for super-Eddington accretion in the simulation but limit the accretion rate to 2 times the Eddington accretion rate:

$$\dot{M}_{\rm Edd} = \frac{4\pi G M_{\rm BH} m_{\rm p}}{\eta \sigma_{\rm T} c}, \tag{2}$$

where $m_{\rm p}$ is the proton mass, $\sigma_{\rm T}$ the Thompson cross-section, c is the speed of light, and $\eta = 0.1$ is the radiative efficiency of the accretion flow on to the BH. Therefore, the BH accretion rate in BLUETIDES simulation is determined by

$$\dot{M}_{\rm BH} = {\rm Min}(\dot{M}_{\rm B}, 2\dot{M}_{\rm Edd}). \tag{3}$$

The Eddington ratio defined as $\lambda_{\rm Edd} = \dot{M}_{\rm BH}/2\dot{M}_{\rm Edd}$ would usually range from 0 to 1 during the evolution of AGNs.

The SMBH is assumed to radiate with a bolometric luminosity $L_{\rm Bol}$ proportional to the accretion rate $\dot{M}_{\rm BH}$:

$$L_{\rm Bol} = \eta \dot{M}_{\rm BH} c^2 \tag{4}$$

with $\eta = 0.1$ being the mass-to-light conversion efficiency in accretion disc according to Shakura & Sunyaev (1973). 5 per cent of the radiation energy is thermally coupled to the surrounding gas that resides within twice the radius of the SPH smoothing kernel of the BH particle. This scale is typically about 1–3 per cent of the virial radius of the halo. The AGN feedback energy only appears in kinetic form through the action of this thermal energy deposition, and no other coupling (e.g. radiation pressure) is included.

2.2 Luminosity of AGNs and galaxies

To calculate the UV-band luminosity of AGNs, we apply bolometric corrections to convert the $L_{\rm Bol}$ to rest-frame UV-band absolute magnitude $M_{\rm UV}$ following Fontanot, Cristiani & Vanzella (2012)

$$M_{\rm UV} = -2.5 \log_{10} \frac{L_{\rm Bol}}{f_{\rm B} \mu_{\rm B}} + 34.1 + \Delta_{\rm B,UV}, \tag{5}$$

where $f_{\rm B} = 10.2$, $\mu_{\rm B} = 6.7 \times 10^{14}$ Hz, and $\Delta_{\rm B,\,UV} = -0.48$.

The UV luminosity of the host galaxy is obtained by modelling its spectral energy distribution, which is constructed by attaching the SED of a simple stellar population to each star particle based on the age and metallicity. We employ version 2.1 of the Binary Population and Spectral Populations Synthesis model (Eldridge et al. 2017) utilizing a modified Salpeter initial mass function (Salpeter high-mass slope with a break at $<0.5\, {\rm M}_\odot$) and a high-mass cut-off of $100\, {\rm M}_\odot$.

We also convert AGN $L_{\rm Bol}$ to the luminosity in the hard X-ray band [2–10] keV following the bolometric correction $L_{\rm X} = L_{\rm Bol}/k$ from Hopkins, Richards & Hernquist (2007), with

$$k(L_{\rm Bol}) = 10.83 \left(\frac{L_{\rm Bol}}{10^{10}\, {\rm L}_\odot}\right)^{0.28} + 6.08 \left(\frac{L_{\rm Bol}}{10^{10}\, {\rm L}_\odot}\right)^{-0.020}, \tag{6}$$

where $\rm L_\odot$ refers to the bolometric solar luminosity ${\rm L}_\odot = 3.9 \times 10^{33}\, {\rm erg\, s^{-1}}$.

2.3 Gas obscuration around AGNs

Bright quasars in a high accretion state are typically shrouded by large fraction of high-density gas, and the luminosities of quasars are largely reduced due to scatter and absorption by gas between them and an observer. The gas density field around AGNs (and the corresponding obscuration) can be a complex environment with large spatial and time variations due to accretion and AGN feedback process. In this work, we explore in detail the angular variations of the gaseous environment of simulated AGN and study the resulting obscuration on a statistical basis.

For each AGN, we use HEALPY[2] to cast 972 evenly distributed lines of sight starting from the position of each AGN and calculate the hydrogen column density $N_{\rm H}$, UV optical depth $\tau_{\rm UV}$, and density-averaged radial velocity $v_{\rm r}$ for each line of sight. More specifically, first we determine the gas density field using the SPH formalism

$$\rho(\mathbf{r}_i) = \sum_j m_j W_{ij} = \sum_j m_j W(|\mathbf{r}_i - \mathbf{r}_j|, h_j), \tag{7}$$

where $\sum_j$ is a sum over all the neighbouring gas particles within the smoothing length h, and W is the quintic kernel used in BLUETIDES. We then calculate $N_{\rm H}$ by integrating the hydrogen number density $n_{\rm H}$ along each line of sight:

$$N_{\rm H} = \int_{\rm ray} X \rho(l)/m_{\rm p} {\rm d}l \tag{8}$$

with $m_{\rm p}$ the proton mass and $X = 0.76$ the hydrogen mass fraction.

As is usual in most cosmological simulations, BLUETIDES applies a uniform UV radiation background to all gas particles. It therefore does not include a self-consistent calculation of the neutral hydrogen fraction for gas in the environment of quasars. In order to account for the effect of the local quasar ionizing radiation, we therefore do the following: First, because only neutral hydrogen should contribute to $N_{\rm H}$, it is a good assumption that the star-forming gas is self-shielded from the photon-dissociating UV background and remains neutral. Therefore, we only include gas particles with a non-zero star formation rate (SFR) when calculating $N_{\rm H}$. Secondly, we also assume that any gas that is not dense enough to be forming stars is fully ionized by the nearby quasar. We note that star-forming gas resides in high-density regions and dominates the contribution of $N_{\rm H}$, and we find as expected that there is negligible difference in our results if instead we compute $N_{\rm H}$ from all gas components.

We also quantify the averaged radial velocity for each line of sight and study its relationship with $N_{\rm H}$. To do this, we determine the velocity field using $\mathbf{v}_i = \sum_j m_j \mathbf{v}_j W_{ij}/\rho_i$ and then calculate the averaged radial velocity $v_{\rm r}$ weighted by density for each line of

[2] https://healpy.readthedocs.io

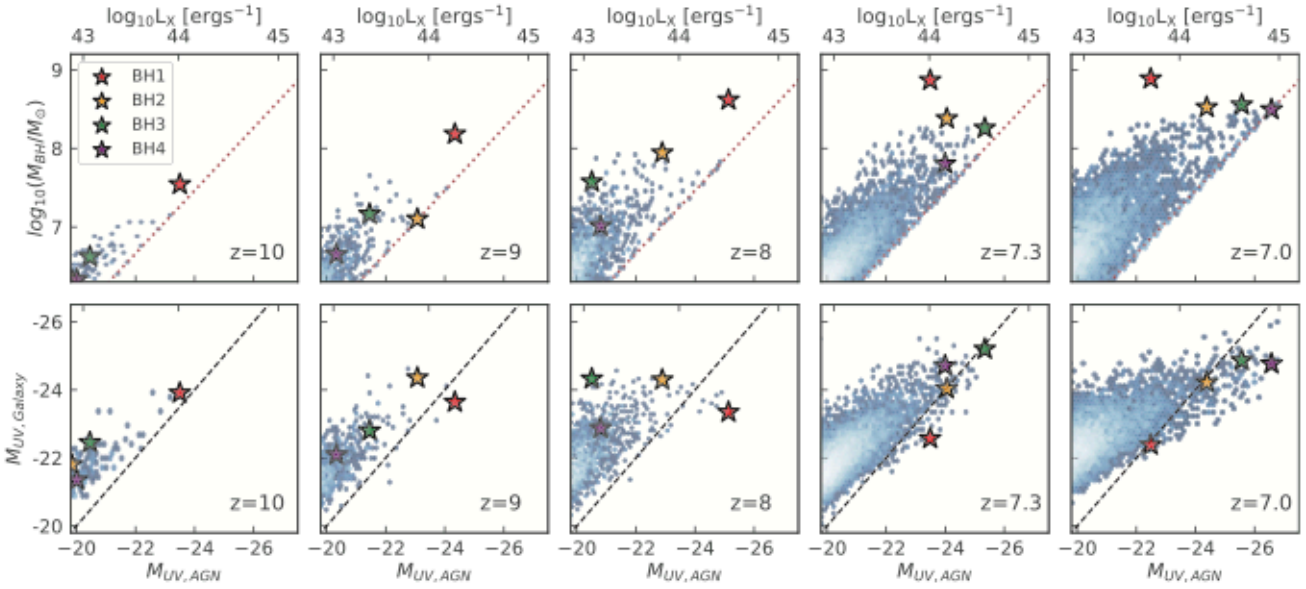


Figure 1. Properties of the SMBH population and their host galaxies in the BLUETIDES simulation measured from $z = 10$ to 7.0 (from left to right). The x-axis shows the QSO luminosity. The upper x-axis shows the corresponding (intrinsic) quasar luminosity in the X-ray band, and the lower x-axis shows the corresponding luminosity in the UV band. The coloured stars mark the four sample QSOs we have selected for illustration in other figures. *Top panel*: 2D histogram of BH mass and BH luminosity. The brown dotted lines are $2 \times L_{\rm Edd}(M_{\rm BH})$ for the corresponding BH mass, which represents the upper bound luminosity for a certain BH mass set in the simulation. *Bottom panel*: 2D histograms of the UV-band (intrinsic) luminosity of AGNs and their host galaxies. The black dashed lines indicate where the AGN and the host galaxy have the same UV-band luminosity.

sight. $v_{\rm r}$ therefore acts as the momentum flux along the specified direction. We use this quantity as a proxy for gas outflow rate in later sections.

To study the AGN obscuration in the UV band brought about by dust attenuation, we employ a scheme same as Wilkins et al. (2017), assuming that the metal density integrated along the line of sight is proportional to dust optical depth:

$$\tau_{\rm UV,AGN} = \kappa \left(\frac{\lambda}{5500\ \text{Å}}\right)^{\gamma} \int_{\rm ray} \rho_{\rm metal}(l)\mathrm{d}l, \tag{9}$$

where $\rho_{\rm metal,\,i} = \sum_j m_j Z_j W_{ij}$ and Z_j is the metallicity (in unit of mass fraction) of gas particle j. Here $\kappa = 10^{4.6}$ and $\gamma = -1.0$ is a free parameter that is calibrated against the observed galaxy luminosity function (see also Marshall et al. 2019). This method is well established in previous studies on luminous galaxy population of BLUETIDES simulation (see e.g. Wilkins et al. 2017, for more detailed descriptions.)

3 RESULTS

3.1 Global BH properties of BLUETIDES

In this work, we consider the host properties of AGNs with a luminosity of $L_{\rm X} > 10^{43}$ erg s^{-1} (corresponding to $L_{\rm bol} > 2 \times 10^{44}$ erg s^{-1}) from $z = 10$ to 7 in the BLUETIDES simulation. For this AGN population, their hosts have at least $M_* > 10^9\,{\rm M}_\odot$ and $M_{\rm gas} > 10^{10}\,{\rm M}_\odot$ within the virial radius R_{200}. The typical host galaxies of $L_{\rm X} > 10^{43}$ erg s^{-1} quasars in the simulation have a physical size of 5–10 ckpc h^{-1} (the half-mass radius of the cold molecular gas). We can resolve the ISM gas content with at least 10^3 gas particles and smoothing length of the order of 1 ckpc h^{-1}.

In Fig. 1, we show some basic properties of the bright AGN population from $z = 10$ to 7. The top panels are the 2D histogram of the BH mass and luminosity. We compute the quasar luminosity in the X-ray band and show it on the upper x-axis and in the rest-frame UV band along the lower x-axis (see Section 2.2). The brown dotted lines in the top panels of Fig. 1 indicate twice the Eddington luminosity, determined from the BH mass (see equation 2). BHs lying on the brown dotted lines are in a state at the upper limit of allowed super-Eddington accretion. The bottom panels of Fig. 1 show the rest-frame UV-band (intrinsic) luminosity of AGNs compared to that of their host galaxies, with the black dashed lines indicating where galaxy and AGN are equally luminous in the UV band. BHs below the black dashed lines are the quasars that have intrinsic UV luminosity that can outshine their host galaxy (without considering the dust attenuation). When evolving to lower redshift, there are more QSOs that can outshine their host galaxy, most of them having high luminosity, $M_{\rm UV} < -22$ (or $L_{\rm X} > 10^{43.5}$ erg s^{-1}).

For illustrative purposes, we choose four sample QSOs to study their host environment, obscuration state, and their time evolution. The objects are marked by coloured stars in Fig. 1. The basic properties of the four sample QSOs at $z = 7$ are listed in Table 1, with the host mass properties (M_*, $M_{\rm gas}$, and $M_{\rm H_2}$) all being calculated within the virial radius of the halo. All the four sample QSOs are massive SMBH with $M_{\rm BH} > 10^8\,{\rm M}_\odot$ and have UV-band luminosities brighter than their host galaxies at $z = 7$. BH1 (in red) is the most massive SMBH in the simulation, having grown to $M_{\rm BH} = 7.7 \times 10^8\,{\rm M}_\odot$ at $z = 7$. We have studied its host galaxy properties and gas outflows in previous publications (Ni et al. 2018; Tenneti et al. 2018).

To allow the reader to briefly examine the four QSO hosts, we plot the gas in their local environments as well as images of their host galaxies in Fig. 2. For each of the panels in Fig. 2, the BH is shown in the centre (position marked by the red cross). The top panels are the gas density field colour coded by temperature (blue to red indicating cold to hot, respectively), the middle panels are stellar density field colour coded by the age of stars (from blue to red

Table 1. The host properties of the four example QSOs at $z = 7$. Here M_*, $M_{\rm gas}$, and $M_{\rm H_2}$ are all calculated within the virial radius R_{200} of the AGN host. Here R_{200} is given in physical units.

	BH1	BH2	BH3	BH4
$L_{\rm Bol,\,BH}$ (erg s^{-1})	1.9×10^{45}	1.1×10^{46}	3.3×10^{46}	8.2×10^{46}
$M_{\rm BH}$ ($M_\odot$)	7.7×10^{8}	3.3×10^{8}	3.6×10^{8}	3.1×10^{8}
M_* ($M_\odot$)	4.2×10^{10}	6.8×10^{10}	8.7×10^{10}	4.8×10^{10}
$M_{\rm gas}$ ($M_\odot$)	3.8×10^{10}	5.1×10^{10}	8.3×10^{10}	6.5×10^{10}
$M_{\rm H_2}$ ($M_\odot$)	6.2×10^{9}	1.7×10^{10}	2.9×10^{10}	2.7×10^{10}
M_{200} ($M_\odot$)	6.3×10^{11}	6.5×10^{11}	9.0×10^{11}	6.2×10^{11}
R_{200} (kpc)	22.2	22.4	25.0	22.1

indicating young to old populations, respectively), and the bottom panels give the density field of the cold molecular gas component in the vicinity of the central BH. In each case, the BH is in the densest part of the gas distribution, surrounded by clumps and filaments of dense gas. The host galaxies vary in their morphology and in the spatial distribution of stellar ages. We study the obscuration caused by their surrounding accreting gas and host galaxy in Section 3.2.

3.2 Gas properties surrounding the QSOs

In this section, we take our four sample QSOs and explore the typical angular variation and time evolution of the column density $N_{\rm H}$, and study its relationship with dust extinction and gas outflow due to AGN feedback.

In Fig. 3(a), we focus on BH1 and show the column density, dust optical depth, and radial velocity of its surrounding gas. The three gas properties are calculated along all lines of sight centred at BH1 and are plotted in Aitoff projection to illustrate the angular distribution. Each row of Fig. 3(a) represents the state at a certain redshift so that we are tracing the time evolution from $z = 10$ to 7. The leftmost column shows the hydrogen column density $N_{\rm H}$, with blue to red indicating low $N_{\rm H}$ to high $N_{\rm H}$, respectively. The middle column gives the corresponding dust optical depth $\tau_{\rm UV}$ calculated based on equation (9), with blue to red representing low to high dust attenuation in this case, respectively. In the third column, we show the averaged radial velocity v_r along the corresponding line of sight. Note that we define the positive values as being in the outward direction and negative values as the inward direction. This means that the yellow to red patches in the third column represent the regions where gas is flowing outward rapidly.

The middle column of dust optical depth $\tau_{\rm UV}$ exhibits a similar angular pattern to the $N_{\rm H}$ map, indicating that directions with larger $N_{\rm H}$ are also more likely to have higher dust extinction. In other words, dust attenuation of AGNs mainly traces the regions with high gas density, and the gas metallicity Z only modulates the variation at a subdominant level.

A comparison between the first and third columns in Fig. 3(a) indicates a clear correlation between the low-column density regions and directions with high outward velocity. This is consistent with our previous study of quasar-driven outflows (Ni et al. 2018), where we found that the outflowing gas tends to channel through low-density regions. The underlying picture is that massive bright quasars were mostly enshrouded by high-density gas, and as the quasars dumped feedback energy into the surrounding gas and drove outflows, this process opened up large regions of low $N_{\rm H}$ where outflows could pass through, opening a window for observations of the central object.

Maps of the other three sample QSOs show a similar pattern relating $N_{\rm H}$, $\tau_{\rm UV}$, and v_r; therefore, we only plot the $N_{\rm H}$ field of the three samples in Fig. 3(b), and put the full three-field maps of BH2, BH3, and BH4 in Fig. B1 of Appendix B as further illustration. We study the relationship of $N_{\rm H}$ with gas outflow and $\tau_{\rm UV}$ on a statistical basis in the next section. To further quantify the $N_{\rm H}$ distribution, we plot in Fig. 5 the $N_{\rm H}$ histograms of the four sample QSOs and mark the contributions from lines of sight with a large outward radial velocity of $v_r > 300$ km s^{-1} in a red colour. Another way to determine the outflow direction is to find in which lines of sight the outflowing gas particles reside. As established in Ni et al. (2018), we find that outflow gas can be defined based on the criterion that a gas particle has a large enough velocity to escape from the potential well of the halo, i.e. peculiar velocity larger than the escape velocity $v_{\rm esc}$:

$$\frac{1}{2} m v_{\rm esc}(r)^2 \geqslant \int_r^{R_{200}} \frac{GM(<r')m}{r'^2} {\rm d}r' + \frac{GM(<R_{200})m}{R_{200}}. \quad (10)$$

The green shaded regions in Fig. 5 mark the lines of sight where the outflow gas constitutes a non-negligible fraction of the overall column density, $N_{\rm out}/N_{\rm H} > 0.01$. Here $N_{\rm out}$ is calculated based on the same formalism as equation (8), but only considering the contribution from outflow gas when calculating equation (7). We see that gas under both criteria occupies the lower $N_{\rm H}$ region in the overall distribution, indicating a correlation between outflow and $N_{\rm H}$.

3.2.1 *Time evolution*

Both Figs 3 and 5 show a rapid evolution of $N_{\rm H}$ with redshift. BH1 is originally heavily obscured at $z = 10$ with $N_{\rm H} \sim 10^{24}$ cm^{-2}, and then with the launching of powerful feedback it clears out its high-density gas environment and ends up with $N_{\rm H} \sim 10^{22}$ cm^{-2} at $z = 7$. On the other hand, BH2, BH3, and BH4 all start with low column density ($N_{\rm H} < 10^{23}$ cm^{-2} at $z = 10$), but get progressively more obscured while also having increasing angular variations with time.

As illustrated by Fig. 5, all the four samples had quite a uniform $N_{\rm H}$ distribution at $z = 10$, with the difference between highest and lowest $N_{\rm H}$ smaller than 1 dex. The angular variation of $N_{\rm H}$ gets larger when going to lower redshift accompanied by the emergence of the outflow. At $z = 7$, the distribution of $N_{\rm H}$ with respect to different lines of sight can span over 2 dex.

More quantitatively, in Fig. 4 we show the time evolution of the four sample QSOs with regard to their BH mass, luminosity, Eddington accretion ratio, and obscuration state. We quantify the level of the obscuration around the QSO using $N_{\rm H,\,ave}$, which is the averaged value of $N_{\rm H}$ over all lines of sight (the fourth panel). To describe the angular variation in $N_{\rm H}$, we calculate the standard deviation of $N_{\rm H}$ based on all lines of sight $\sigma(\log_{10} N_{\rm H})$ (the fifth panel).

Tracing the time evolution, we find that the four sample QSOs have all reached a high accretion state with $L/L_{\rm Edd} \gtrsim 1$ during their evolutionary history (BH1 at $z = 10$, BH2 at $z = 9$, and BH3 and BH4 at $z = 7.5$). At that epoch, they were generally at a higher level of obscuration, with $N_{\rm H,ave} \sim 10^{24}$ cm^{-2}. Fig. 4 also shows that massive BHs are unlikely to stay in a state of close to Eddington accretion, since the strong AGN feedback driven by the high luminosity will self-modulate the surrounding gas density field, launching strong gas outflows that clear out part of the high-density accreting gas (which has large obscuration), and increases the angular variations in the $N_{\rm H}$ field.

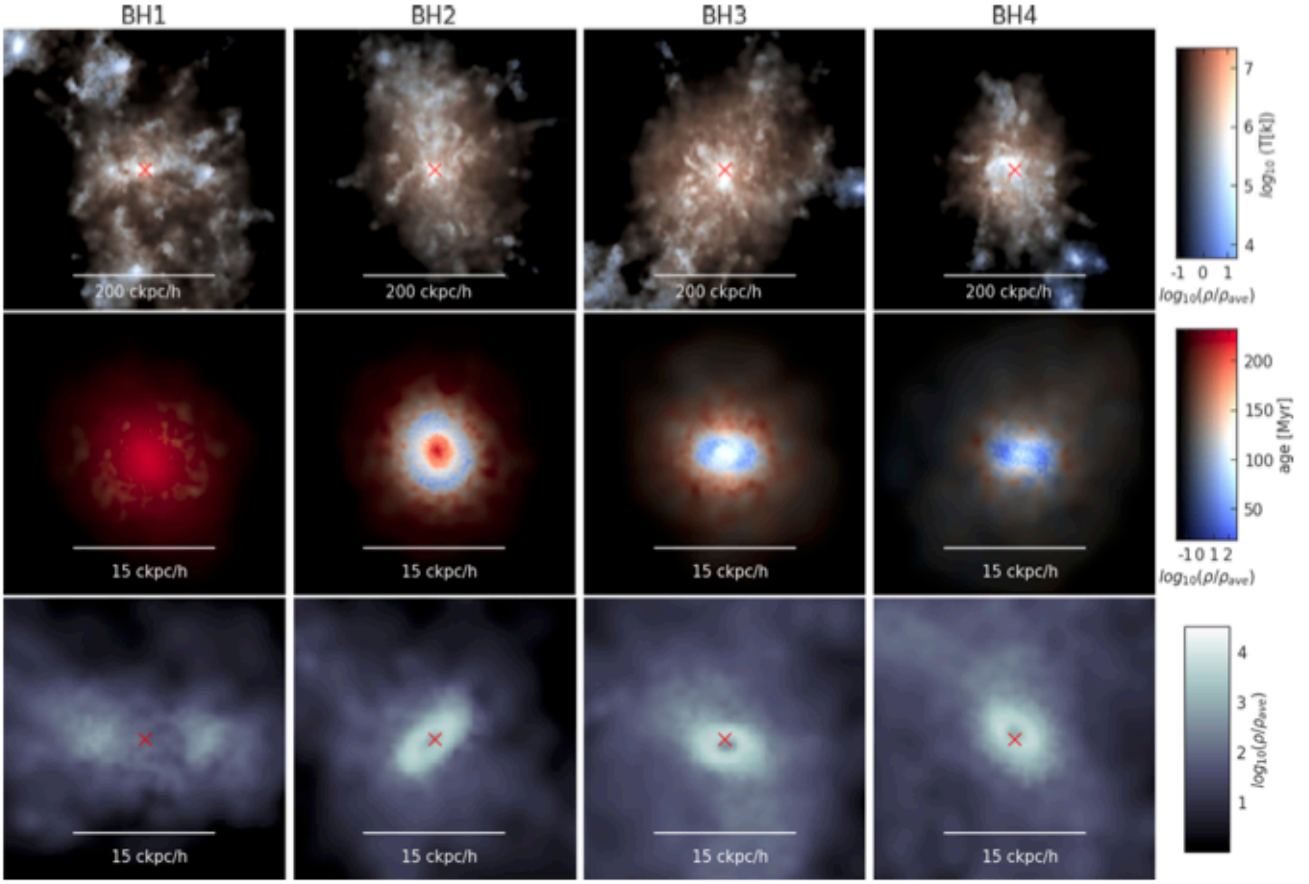


Figure 2. Illustration of the host environment of the four sample QSOs at $z = 7.0$. The top panels give the gas density field colour coded by temperature (blue to red indicating cold to hot, respectively, as shown by the colour bar aside). The boxes are 400 ckpc h^{-1} (comoving unit) per side, with BHs residing in the centre (marked by the red cross in the top panels). The middle panels paint the stellar density field for the host galaxy (in face-on orientation) colour coded by the age of stars (from blue to red indicating young to old populations, respectively). The bottom panels give the density field of the cold molecular gas surrounding the BHs, in contrast with the mean surface density of the host gas $\rho_{\rm ave}$. The middle and bottom boxes are all 30 ckpc h^{-1} per side.

We study the relationship between angular variations of the obscuration and QSO properties using larger, statistical samples in Section 3.3.

3.2.2 *Radial distribution*

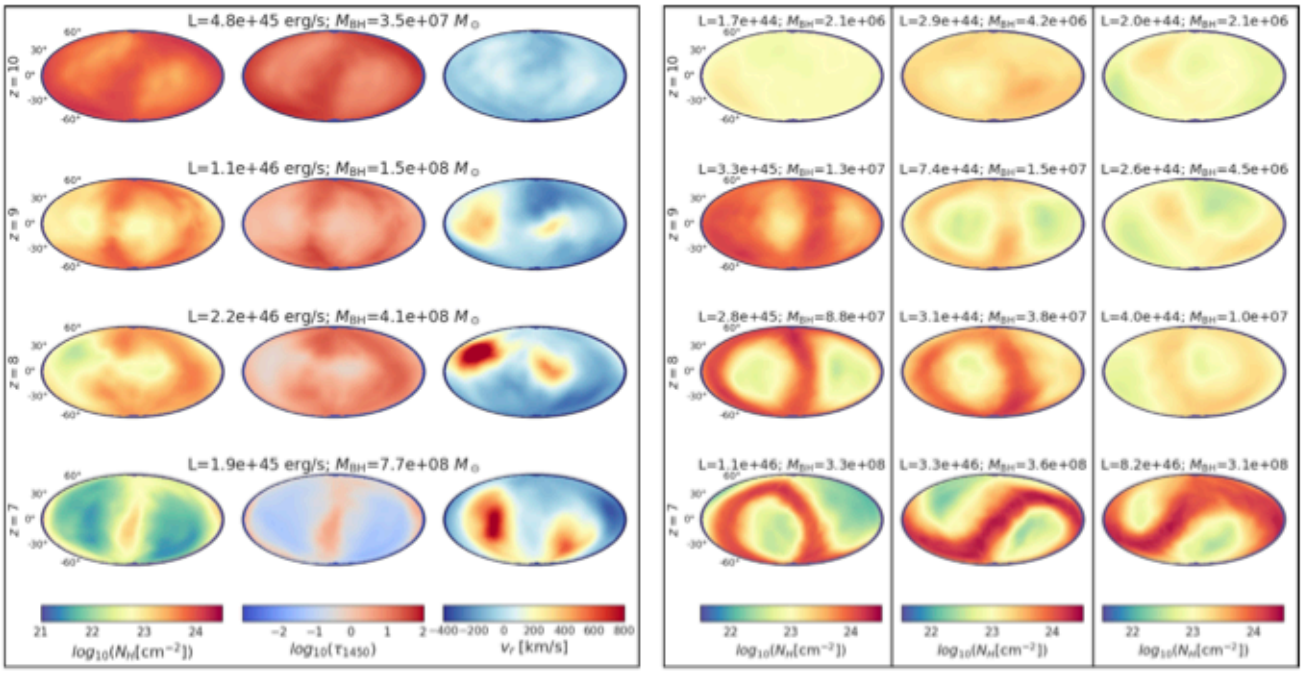
Studying the length-scales that make the highest contribution to galactic obscuration can give us more physical insight. To do this, we use the differential $N_{\rm H}$ profile $dN_{\rm H}/dr$, which shows the radial distribution of $N_{\rm H}$ along each line of sight. In Fig. 6, we select 192 evenly distributed lines of sight (using HEALPY) for the four sample QSOs at $z = 7$ and plot the corresponding differential $N_{\rm H}$ profiles as grey lines. The x-axis indicates the distance (in co-moving coordinates) from the central QSO. Integration of the profile over r gives the column density $N_{\rm H}$ of the corresponding line of sight. For each panel, the blue/green line marks the most/least obscured line of sight, with corresponding $N_{\rm H}$ given in the legends. Comparing the blue and green lines, we can see that the galactic obscuration surrounding a certain QSO can vary by over 2–3 orders of magnitude between different lines of sight, because of spatial variations in the density field.

The differential $N_{\rm H}$ radial profiles show clearly that for sight-lines with large column density, the $N_{\rm H}$ is mostly contributed by high-density gas clumps (e.g. those with $dN_{\rm H}/dr > 10^{22}$ [cm^{-2}/ckpc h^{-1}]) that mostly reside within $r < 10$ ckpc h^{-1} (corresponding to $r < 1.8$ kpc in physical units) of the QSO. More quantitatively, using the mean $N_{\rm H}$ values (black dotted lines), we find that $N_{\rm H}(< 10\,{\rm ckpc}\,h^{-1}) \gtrsim 90$ per cent $N_{\rm H}(< 30\,{\rm ckpc}\,h^{-1})$.

The pink lines in each panel denote the lines of sight with strongly outflowing gas ($v_r > 300$ km s^{-1}) for the sample QSOs. We can see that they mostly channel through the directions without high-density clumps and indeed have low $N_{\rm H}$ values compared to the overall populations.

We would also like to compare the radial $N_{\rm H}$ distribution to the stellar density distribution of the host galaxy. The black dashed lines in each panel of Fig. 6 show the averaged stellar density profile converted to an equivalent hydrogen number density in units of $n_{\rm H}/[{\rm cm}^2 \times {\rm ckpc}\,h^{-1}]$. (To convert to hydrogen number density, we divide the mean stellar density by the hydrogen mass $m_{\rm p}$.) Since the star-forming gas in our model is the source of obscuration, it is straightforward to understand why the radial profile of the stellar density approximately traces the $N_{\rm H}$ profile, indicating that most of the obscuring gas resides in the host galaxy.

We emphasize that the $N_{\rm H}$ radial profile in Fig. 6 clearly shows that, for $z = 7$ QSO, the host ISM is able to produce absorption up to Compton-thick level ($N_{\rm H} > 10^{24}$ cm^{-2}) on scale of a few ckpc. This is due to the ISM at these high redshifts being much denser than that around the AGN at low redshifts ($z \lesssim 2$), where Compton-thick absorption can only be produced by the parsec-scale gas/torus in the nuclear region (Buchner & Bauer 2017). In particular, these



(a) Maps of N_H, dust optical depth τ_{1450} and radial velocity v_r around BH1. (b) Maps of N_H around black holes BH2, BH3, and BH4

Figure 3. *Left figure*: The properties of gas surrounding BH1 (the most massive BH) in an Aitoff projection. Each row represents a specific property at a certain redshift, and the columns its associated evolution from $z = 10$ to 7 (from top to bottom). The leftmost column is the map of hydrogen column density N_H. The middle column is the map of dust optical depth τ_{UV}. The right column gives the averaged radial velocity v_r along each line of sight. The associated BH bolometric luminosity and BH mass at the corresponding redshift are given at the top of each panel. *Right figure*: Maps showing the evolution of N_H with redshift for BH2 (left), BH3 (middle), and BH4 (right). The full N_H, τ_{UV}, and v_r maps of these three sample QSOs are given in Appendix B. Each panel is labelled by the BH bolometric luminosity (in unit of erg s^{-1}) and BH mass (in unit of $M_\odot$) at the corresponding redshift. The maps at $z = 7$ exhibit a bimodal morphology with regard to the angular distribution of the N_H field.

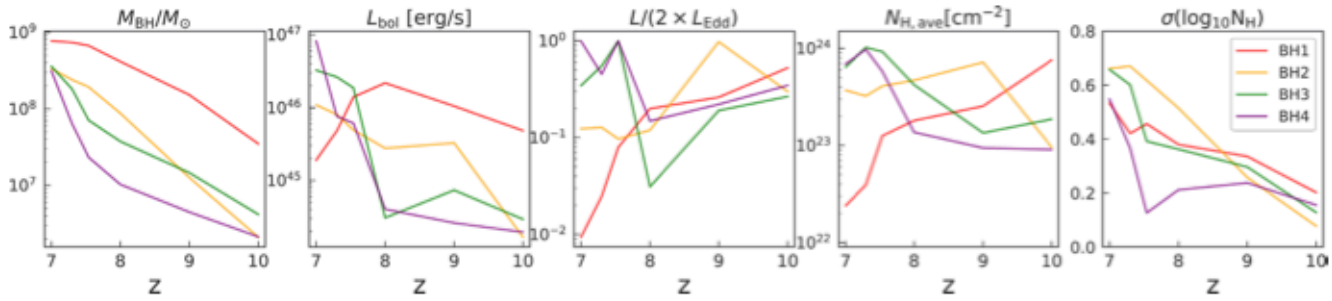


Figure 4. The time evolution of the four samples from $z = 10$ to 7. The first three panels show the following properties of each BH as a function of redshift: BH mass, bolometric luminosity of AGN, and the Eddington accretion ratio. The rightmost two panels quantify the obscuration state of the host environment. The fourth panel shows $N_{H,ave}$, the N_H value averaged over all lines of sight. The rightmost panel displays $\sigma(\log_{10}N_H)$ calculated from all lines of sight, as a simple measure of the angular variation in N_H values.

authors apply observational relation between N_H and stellar mass to the $z < 3$ AGN population, and find that galaxy-scale gas is only responsible for a Compton-thin obscuration, but it does not produce Compton-thick columns.

3.3 Statistics of quasar obscuration

In this section, we study the obscuration state of the bright AGN population on a statistical basis. We mainly focus on the $z = 7$ bright AGN population with $L_X > 10^{43}$ erg s^{-1} (corresponding to $L_{bol} > 10^{44.3}$ erg s^{-1}). In the simulation, we have overall 3504 AGNs with $L_X > 10^{43}$ erg s^{-1} at $z = 7$, and these are widely spread with regard to BH mass and accretion state, residing in all different kinds of hosts. We explore the relation between the N_H distribution and outflows as well as the AGN and galaxy properties.

3.3.1 The N_H distribution around quasars and its relation to outflows

First, we explore the relation between the N_H distribution and the outflow driven by AGN feedback. In Section 3.2 (Figs 3, 5, and 6), we have illustrated that a positive correlation exists between gas outflow and low-N_H regions in the hosts of our four sample QSOs. We now study this relation statistically with larger samples.

In Fig. 7, we plot the N_H versus v_r distribution for all lines of sight around the $L_X > 10^{43}$ erg s^{-1} QSO population at $z = 9$, 8,

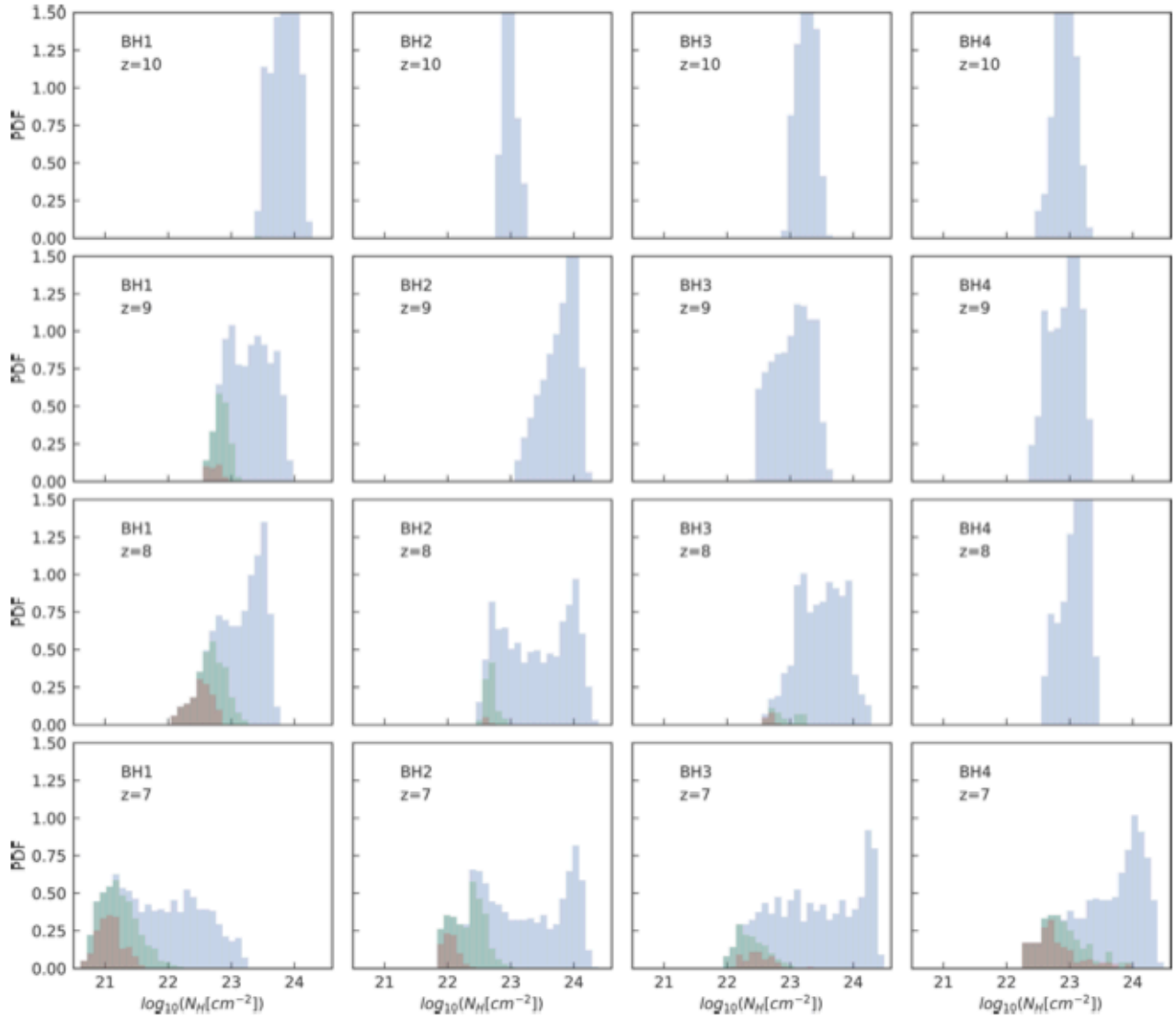


Figure 5. $N_{\rm H}$ histogram of the four sample QSOs from $z = 10$ to 7. The green shade marks out the contribution of lines of sight with $N_{\rm out}/N_{\rm H} > 0.01$ while the brown shade marks out the distribution of lines of sight with $v_{\rm r} > 300\,{\rm km\,s^{-1}}$. All the panels share the same x- and y-axes.

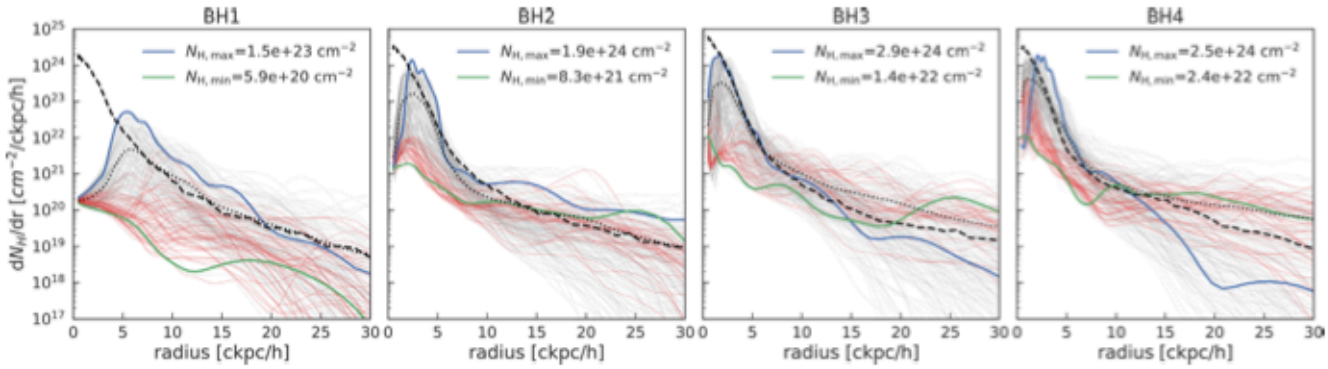


Figure 6. Radial profiles of $N_{\rm H}$ around QSOs. The grey lines are the differential $N_{\rm H}$ profiles for 192 lines of sight around each sample QSO at $z = 7.0$. For each panel, the blue solid line gives the profile of the most obscured line of sight (i.e. that with the largest $N_{\rm H}$ value), and the green solid line corresponds to the least obscured line of sight. The black dotted lines mark out the mean value of all lines of sight at the corresponding radius. The radial profiles for directions with strong outflows ($v_{\rm r} > 300\,{\rm km\,s^{-1}}$) are coloured pink. The dashed black lines in each panel trace the corresponding stellar density profiles, converted into the appropriate units: $n_{\rm H}/[{\rm cm^2} \times {\rm ckpc}\,h^{-1}]$.

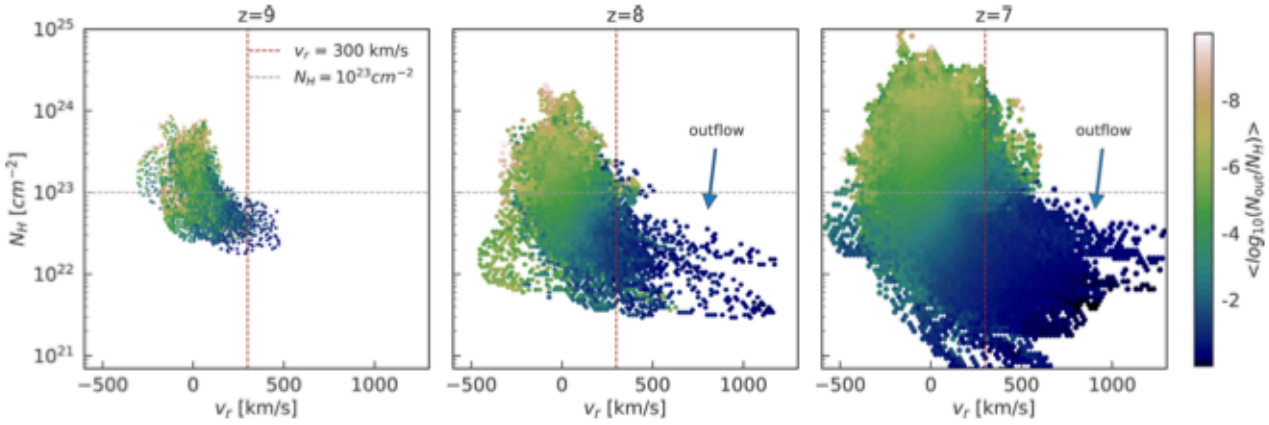


Figure 7. $N_{\rm H}$ versus radial velocity $v_{\rm r}$ for all lines of sight around the QSO population with $L_{\rm X} > 10^{43}$ erg s^{-1}. We show results at redshifts from $z = 9$ to 7. The colour coding indicates the averaged outflow fraction $N_{\rm out}/N_{\rm H}$ in each bin. The red vertical dashed line shows where $v_{\rm r} = 300$ km s^{-1}, and the grey horizontal dashed line is where $N_{\rm H} = 10^{23}$cm^{-2}.

and 7. In particular, we have overall 259, 637, and 3504 such bright AGNs at these three redshifts, and we plot the $N_{\rm H}$ distribution based on 972 lines of sight for each AGN. The colour coding indicates the averaged value of outflow fraction $N_{\rm out}/N_{\rm H}$ in each bin.

We see that the lines of sight with large outward radial velocity indeed correspond to lower $N_{\rm H}$ regions and they generally have large outflow fractions with $N_{\rm out}/N_{\rm H} > 0.01$ (blue colour). At $z = 7$, for lines of sight with $v_{\rm r} > 300$ km s^{-1}, 87 per cent of them have $N_{\rm H} < 10^{23}$ cm^{-2}. The overall $N_{\rm H}$ distribution at high redshift $z = 9$ is narrower because of the lack of powerful gas outflows. When going to lower redshift, the $N_{\rm H}$ distribution is broadened, accompanied by more gas outflows launched by AGN feedback.

We now focus on the latest redshift $z = 7$ and study the relationship between $N_{\rm H}$ and outflow based on the QSO population with $L_{\rm X} > 10^{43}$ erg s^{-1}. In the left-hand panel of Fig. 8, we plot the $N_{\rm H}$ probability distributions obtained by separating the QSO hosts into multiple $M_{\rm out}$ bins, where $M_{\rm out}$ is the mass of the outflow gas calculated based on outflow criteria equation (10). Each bin here contains at least 300 AGNs (with 972 lines of sight used for analysis of each AGN). The overall $N_{\rm H}$ distribution skews towards the low-$N_{\rm H}$ end when more gas is outflowing from the host. For QSO hosts that have $M_{\rm out} > 10^{8.9}$ M$_\odot$ (purple line), the probability of having $N_{\rm H} < 10^{23}$ cm^{-2} can reach 40 per cent.

The Aitoff projection maps in Fig. 3 illustrate how the morphology of the outflow can be bimodal for our four sample QSOs. To study a larger QSO population, we apply a method inspired by angular correlation analysis to quantify the morphologies of outflow and angular $N_{\rm H}$ variations. The results are shown in the right-hand panel of Fig. 8, and are computed as follows. For each AGN, we take pairs of sightlines from the 972 available, and calculate the probability that both members of a pair of sightlines with a specific angular separation θ satisfy particular physical criteria. We calculate the probability by counting the number of sightline pairs in each angular separation bin that satisfy the criteria and divide by the total number of pairs in the bin. Our sample is the AGN population with $L_{\rm X} > 10^{43}$ erg s^{-1} at $z = 7$. The blue line in Fig. 8 shows results for the first criterion, the probability that both sightlines have $N_{\rm H} < 10^{23}$ cm^{-2}. We see that the probability becomes smaller when the separation angle θ increases from 0° to 90°, and becomes larger again when θ keeps growing from 90° to 180°. This convex shape can be produced if the averaged $N_{\rm H}$ distribution is bimodal in angle. This is in fact what we have seen in the example maps in Fig. 3. This morphology points to most of the QSOs having bimodal outflows, with their signature showing up as angular deficits in the obscuration. The red and green lines show the corresponding probability that a pair of lines of sight have $N_{\rm out}/N_{\rm H} > 0.01$ and $v_{\rm r} > 300$ km s^{-1}. They have lower amplitude than for the $N_{\rm H}$ criteria because they are more stringent (as also illustrated in Fig. 5). They, however, yield the same convex feature with the minimum probability at $\theta = 180°$, indicating that on an averaged basis, the outflows in our simulation are indeed consistent with a bimodal structure.

3.3.2 *The relationship of* $N_{\rm H}$ *to AGN properties*

In this section, we explore the $N_{\rm H}$ distribution and its relationship to AGN properties; in particular, we are interested in how it varies with respect to AGN luminosity, BH mass, and Eddington ratio.

In Fig. 9, we split the QSO population with $L_{\rm X} > 10^{43}$ erg s^{-1} at $z = 7$ into different bins of X-ray-band luminosity $L_{\rm X}$ (left-hand panel), $M_{\rm BH}$ (middle panel), and the Eddington accretion ratio $\lambda_{\rm Edd}$ (right-hand panel), and plot the overall probability distribution function (PDF) of $N_{\rm H}$ (including all lines of sight) for the corresponding bin. We aim to show how the overall $N_{\rm H}$ distribution changes with respect to a few basic AGN properties. We note that each bin in Fig. 9 contains at least 300 AGNs (with 972 lines of sight used for each AGN).

The left-hand panel of Fig. 9 shows that $N_{\rm H}$ distribution becomes systematically broader for increasing AGN luminosity. The peak of the distribution changes from $N_{\rm H} = 10^{23.1}$ cm^{-2} in the faintest bin (the blue line) to $N_{\rm H} = 10^{23.6}$ cm^{-2} for the most luminous bin (purple). In addition, the $N_{\rm H}$ distribution develops a longer tail towards low $N_{\rm H}$ with increasing AGN luminosity. The lowest $N_{\rm H}$ lines of sight are created, as discussed in the previous section, as a result of the outflows driven by AGN feedback. For the QSO population with $L_{\rm X} > 10^{44}$ erg s^{-1} (the purple line), the low $N_{\rm H}$ tail with $N_{\rm H} < 10^{23}$ cm^{-2} constitutes 11 per cent of the overall distribution.

The middle panel of Fig. 9 illustrates the relation between $N_{\rm H}$ and BH mass. It shows a similar trend with luminosity: The overall $N_{\rm H}$ distribution becomes broader and also skews towards lower $N_{\rm H}$ end with increasing BH mass. Quantitatively, the peaks of the $N_{\rm H}$ distribution as a function of $M_{\rm BH}$ again move from $10^{23.1}$ to $10^{23.5}$ cm^{-2} when going from the lowest BH mass bin (blue) to the

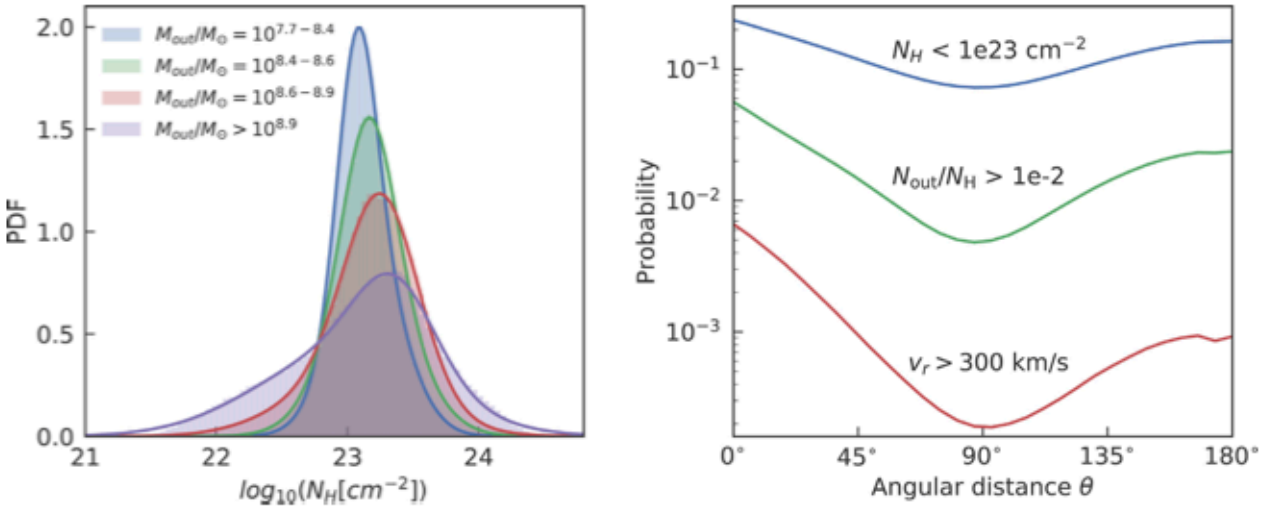


Figure 8. *Left-hand panel*: We split the $L_X > 10^{43}$ erg s^{-1} QSO population at $z = 7$ into different bins of mass of outflowing gas and plot the probability distribution function (PDF) of N_H calculated based on all lines of sight for QSOs in each bin. *Right-hand panel:* The morphology of the gas outflow quantified using angular correlations. The profiles give the probability of both members of a pair of lines of sight satisfying the given criteria as a function of separation angle. The profiles are calculated by averaging over the QSO population with $L_X > 10^{43}$ erg s^{-1} at $z = 7$ (see the text for more detailed description).

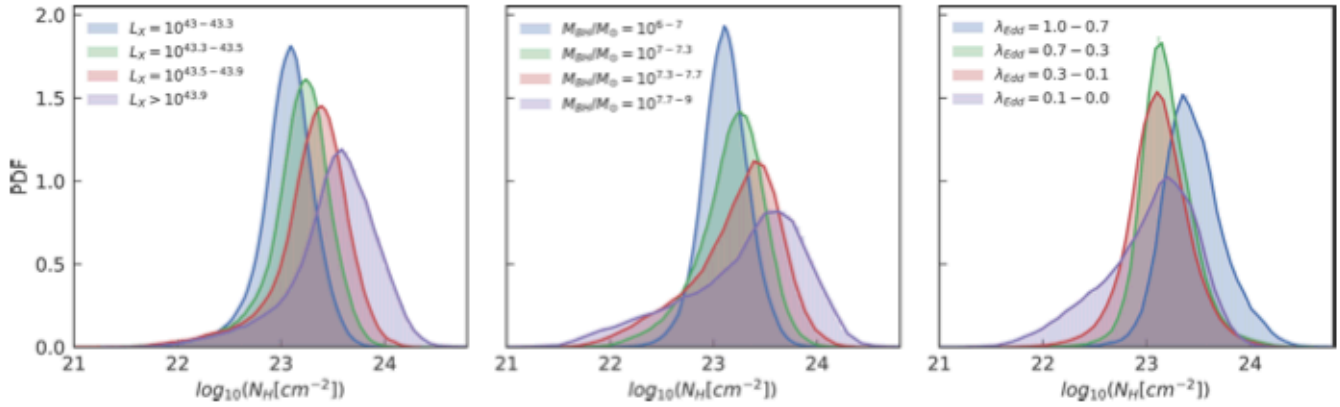


Figure 9. We split the $L_X > 10^{43}$ erg s^{-1} QSO population at $z = 7$ into different bins of luminosity (left), BH mass (middle), and Eddington accretion ratio (right) and study the corresponding probability distribution function (PDF) of N_H for all lines of sight in each bin. The histograms are normalized so that the area under each curve is unity.

highest mass bin (purple). The low-N_H tail (with $N_H < 10^{23}$ cm^{-2}) does, however, take up a larger fraction, 27 per cent, of the highest mass bin. This is again a consequence of the fact that massive AGNs are able to drive more powerful feedback that gives rise to the low-N_H regions in their surroundings. One fact to take into account is that the purple histograms in the left-hand and middle panels both contain about 300 AGNs; however, the histogram in the middle panel is wider and more skewed, indicating that the N_H distribution for the most massive AGN population has larger variations compared to those for the most luminous population.

The right-hand panel of Fig. 9 displays the N_H distribution as a function of Eddington ratio, the ratio of bolometric luminosity to Eddington luminosity (which is only a function of the BH mass via equation 2). There is no clear trend in the N_H distribution in this case. However, comparing the purple with the blue histogram shows that the QSO population with lower Eddington ratio is more likely to be less obscured. For QSOs with $\lambda_{Edd} < 0.1$ (purple line), the probability of finding $N_H < 10^{23}$ cm^{-2} is 40 per cent, while for $\lambda_{Edd} > 0.7$ the probability is only 4 per cent. We note, however, that the low N_H values for AGNs with low Eddington accretion are not necessarily driven by outflows since low λ_{Edd} could also be caused by low luminosity, where the AGN is not surrounded by high-density gas environment.

To further quantify the correlation between N_H and AGN luminosity, we fit a linear relation between logN_H and logL_X. As illustrated in Fig. 9, the scatter in N_H is larger for increasing luminosity. Therefore, we use an up-sampling method to exploit the underlying angular distribution of N_H for each AGN, and compute uncertainties on the regression parameters. We take the N_H value observed from 1 line of sight (out of 972 lines of sight for each AGN) as one realization and carry out a linear regression of N_H versus AGN luminosity L_X for all realizations. In the upper panel of Fig. 10, we plot the results from these fits with the grey lines for all the 972 realizations, and a black dashed line showing the averaged relation. The linear regression gives

$$\log N_H = (0.42 \pm 0.02) \log L_X + (5.1 \pm 0.7). \quad (11)$$

The blue points in the upper panel show N_H versus L_X for a single realization. We quantify the scatter of the linear fitting ϵ as the

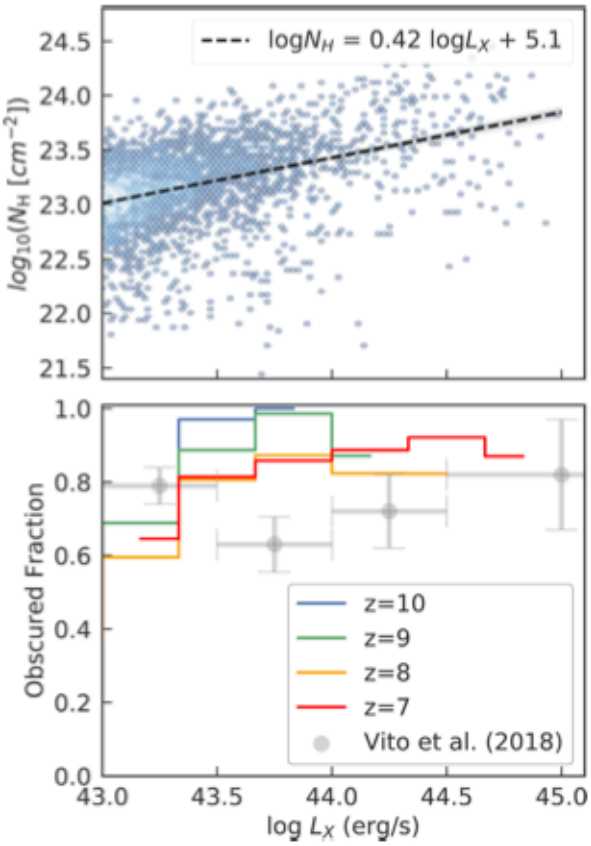


Figure 10. *Upper panel*: $N_{\rm H}$ versus AGN luminosity for the $L_{\rm X} > 10^{43}$ erg s^{-1} AGN population at $z = 7$. The blue points show $N_{\rm H}$ along random lines of sight for AGN versus their intrinsic hard X-ray-band luminosity $L_{\rm X}$. The black dashed line shows the average linear fit for all AGNs, and the grey area overplots the linear fits made from all 972 realizations of $N_{\rm H}$ resulting from different lines of sight. *Bottom panel*: Binned estimates of the obscured AGN fraction (with $N_{\rm H} > 10^{23}$ cm^{-2}) as a function of $L_{\rm X}$, shown from $z = 10$ to 7. The obscured fraction is calculated based on all lines of sight for the AGN populations in the corresponding luminosity bin. The grey data points with error bars are the observational results of Vito et al. (2018), based on the AGN population at redshift $z = 3$–6.

standard deviation of residuals. The $N_{\rm H}$ versus $L_{\rm X}$ relation for a single realization has $\epsilon = 0.29$ dex.

Converting $L_{\rm X}$ back to the unit of $L_{\rm bol}$ (with bolometric correction equation 6), equation (11) can be rewritten as $N_{\rm H} = (1.4 \times 10^{23}\,{\rm cm}^{-2})(L_{\rm bol}/10^{11}\,{\rm L}_{\odot})^{0.32}$. This power-law coefficient between $N_{\rm H}$ and $L_{\rm bol}$ is consistent with the finding of Hopkins et al. (2005) where they examined a series of isolated galaxy merger simulations and give that $N_{\rm H} \sim L_{\rm bol}^{1/3}$. The positive correlation between $N_{\rm H}$ and $L_{\rm bol}$ is essentially driven by the fact that the source of BH obscuration also serves as the fuel for BH accretion. Similar as the interpretation in Hopkins et al. (2005), we expect that $N_{\rm H} \sim M_{\rm c}$, where $M_{\rm c}$ is the gas mass within the core of galaxy that dominates the obscuration, and that $L_{\rm bol} \propto \dot{M}_{\rm BH} \sim M_{\rm c}^3$ for Bondi accretion. We further illustrate this point in Appendix A, where we define $M_{\rm c}$ as the cold molecular gas around central AGN within 5 ckpc h^{-1} (the averaged half-mass radius of the host galaxy), and show that $N_{\rm H} \sim M_{\rm c}^{1.1}$ and $L_{\rm bol} \sim M_{\rm c}^{3.1}$. (See Appendix A for more details.)

To quantify the relation between $N_{\rm H}$ and $M_{\rm BH}$, we also apply a linear regression between $\log N_{\rm H}$ and $\log M_{\rm BH}$. For the QSO population with $L_{\rm X} > 10^{43}$ erg s^{-1} at $z = 7$, we get $\log N_{\rm H} = (0.13 \pm 0.02)\log M_{\rm BH} + (22.3 \pm 0.2)$. However, as discussed for Fig. 9, the variance of the $N_{\rm H}$ distribution for the most massive AGN bin is much larger than that for the most luminous AGN population. The linear fit of $\log N_{\rm H}$ to $\log M_{\rm BH}$ therefore has a larger scatter at the massive $M_{\rm BH}$ end than the low end in each of the realizations.

We now compare our results to current observational constraints on the obscured fraction of high-redshift QSOs. In the lower panel of Fig. 10, we calculate the obscured QSO fraction in BLUETIDES, defined as the probability of a line of sight having $N_{\rm H} > 10^{23}$ cm^{-2}. We do that as a function of the AGN luminosity and plot using coloured solid lines the results obtained from $z = 10$ to 7. The grey data points with error bars are the results from Vito et al. (2018), based on observations of AGNs from $z = 3$ to 6. Note that we apply the same criteria for 'obscured fraction' as that used by Vito et al. (2018) for high-z QSOs. Those authors use $N_{\rm H} = 10^{23}$ cm^{-2} as the minimum column density for (heavily) obscured AGNs at high redshifts. This choice is different from the typical definition of the unobscured column density threshold $N_{\rm H} = 10^{22}$ cm^{-2} because of the limited spectral quality of high-redshift X-ray observations (see discussions in Vito et al. 2018, for more details).

As can be expected from our previous consideration of the left-hand panel of Fig. 9, there is a weak trend whereby the obscured fraction increases when luminosity increases from $L_{\rm X} \sim 10^{43}$ to 10^{44} erg s^{-1}. When going to even higher luminosity ($L_{\rm X} > 10^{44}$ erg s^{-1}), the obscured fraction does not keep growing because of the strong feedback brought by the bright AGNs. Overall, we predict that the obscured fraction for the $L_{\rm X} > 10^{43}$ erg s^{-1} QSO population at $z \geq 7$ in our simulation ranges from 0.7 to 1, which is comparable with (or slightly higher than) the current observational results based on $z = 3$–6 QSOs.

Trebitsch et al. (2019) also carried out similar study of the AGN obscured fraction versus luminosity by tracing the evolution history of a single AGN using zoom-in simulation. They studied a less extreme AGN that only reaches $M_{\rm BH} \sim 10^7\,{\rm M}_{\odot}$ and $L_{\rm X,\,max} \sim 10^{43}$ erg s^{-1}, but with a higher spatial resolution that can resolve the smaller scale obscuration structure. They obtained obscured fraction ranging from 0.6 to 1 when the AGN has a luminosity of $L_{\rm X} \sim 10^{42\text{–}43}$ erg s^{-1}. In contrast, BLUETIDES simulation explores the AGN population at larger mass and higher luminosity that is more comparable with the current high-redshift observations. However, with its large volume, BLUETIDES has a lower spatial resolution than zoom runs like Trebitsch et al. (2019) and therefore cannot resolve the parsec scale around the AGN. The limited resolution may lead to the lower obscured fractions we measure for the smaller objects [compared to Vito et al. (2018)], especially for the AGN with $L_{\rm X} \sim 10^{43}$ erg s^{-1} that tends to have column densities close to the threshold value of $N_{\rm H} = 10^{23}$ cm^{-2} (as shown in Fig. 9).

3.3.3 *Redshift evolution*

From the lower panel of Fig. 10, we can see that there is no clear trend of evolving obscured fraction with redshift. More explicitly, in Fig. 11 we give the redshift evolution of the space density of AGNs with $L_{\rm X} > 10^{43}$ erg s^{-1} and split the total space density at certain redshifts into different $N_{\rm H}$ bins calculated based on all lines of sight for the corresponding AGN population. The blue line represents the $N_{\rm H} < 10^{22}$ cm^{-2} component that corresponds to the typical definition of unobscured QSOs. The green line represents $N_{\rm H}$ values from 10^{22} to 10^{23} cm^{-2}, the red and purple lines represent $N_{\rm H}$ between 10^{23} and 10^{24} cm^{-2}, and the yellow line shows the Compton-thick fraction with $N_{\rm H} > 10^{24}$ cm^{-2}. We see that the $N_{\rm H}$ distribution is mainly dominated by column densities $N_{\rm H} \sim$

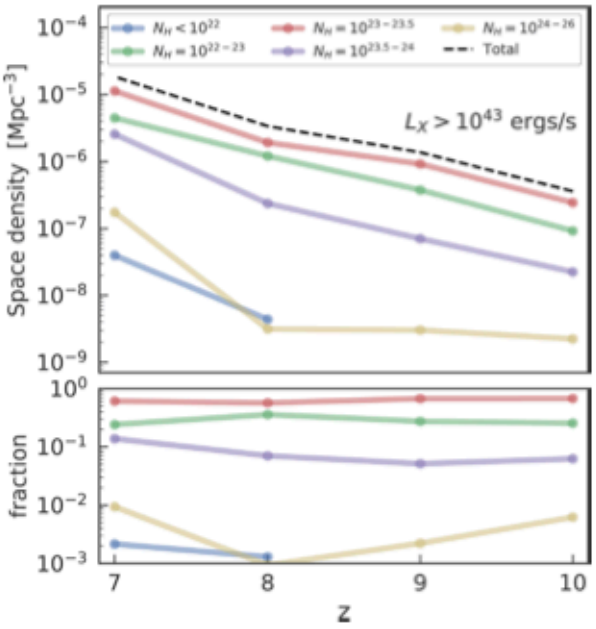


Figure 11. *Upper panel*: The space density of the QSO population with $L_X > 10^{43}$ erg s^{-1} as a function of the redshift. The black dashed line is the total space density. This is split into separate N_H bins (shown by coloured lines) by multiplying the fraction of the corresponding N_H component calculated based on all QSO lines of sight. *Lower panel*: The fraction of different N_H components, obtained by dividing the coloured lines by the black dashed line in the upper panel.

$10^{23-23.5}$ cm^{-2}, and the fraction in each N_H bin does not change much with redshift. The same analysis based on the QSO population above a higher luminosity cut (e.g. $L_X > 10^{43.5}$ erg s^{-1}, not shown in a figure) shows the same evolution of N_H fraction. We therefore conclude that the overall N_H distribution around QSOs above a certain luminosity threshold ($L_X > 10^{43}$ erg s^{-1}) does not show any strong evolution with redshift from $z = 10$ to 7.

3.3.4 *Relationship of* N_H *to host galaxy properties*

In this section, we explore the relationship between the N_H distribution and QSO host properties. In particular, we are interested in the variation of obscuration with stellar mass, molecular gas mass, and the SFR in the host galaxy.

Similar to the style of Fig. 9, in Fig. 12 we split the QSO population with $L_X > 10^{43}$ erg s^{-1} at $z = 7$ into different bins of galaxy stellar mass M_* (left-hand panel), M_{H_2} (middle panel), and SFR (right-hand panel). We give the overall probability density distribution of N_H including all lines of sight for the AGNs in each bin. Here all properties are calculated within the virial radius R_{200} of the QSO host, and each bin contains at least 300 objects (with 972 lines of sight for each AGN).

If we consider the left-hand panel of Fig. 12, we can see that the overall N_H distribution peaks at higher N_H and becomes more broadened as we move to bins representing more massive M_* values. For QSOs with $M_* > 10^{10.3}$ M$_\odot$ (purple contour), the probability of finding $N_H < 10^{23}$ cm^{-2} is 29 per cent. To quantify the relationship between N_H and M_*, we again use the up-sampling method described in the previous section and apply a linear regression analysis to all the realizations. The fit results are given in the upper panel of Fig. 13, with the grey lines being the fits for all 972 realizations, and the black dashed line the averaged value. The linear regression gives

$$\log N_H = (0.24 \pm 0.03) \log M_* + (20.7 \pm 0.3). \quad (12)$$

The blue points show the N_H from one random realization versus M_*, with intrinsic scatter $\epsilon = 0.33$ dex.

Observations at lower redshifts also find a positive correlation between N_H and M_*. For example, Lanzuisi et al. (2017) fit the linear relation between N_H and M_* based on a sample of X-ray-detected AGNs and their far-UV-detected host galaxies in the redshift range $0.1 < z < 4$, giving a slope in the range $\alpha = 0.42$–0.88 for different redshift bins. The purple dashed line in Fig. 13 shows their linear regression result for the $2 < z < 4$ QSO population. The red squares in Fig. 13 show the result found by Rodighiero et al. (2015) for a sample of $z \sim 2$ AGN hosts in the COSMOS field.

The above two observations probe N_H through AGN lines of sight. Additionally, Buchner, Schulze & Bauer (2017) probe galactic obscuration through N_H values inferred from the X-ray spectra of gamma ray bursts (GRB), and study how N_H varies with respect to GRB host galaxy mass. These authors find that $N_H \propto M_*^{0.38}$ in the redshift range $1 < z < 5$, implying that galactic obscuration scales with the galaxy size. We note here, however, that N_H see along AGN lines of sight is mainly contributed by the innermost regions of galaxies (<10 ckpc h^{-1}) and is subject to large variation due to AGN feedback. N_H therefore does not exhibit a tight power-law relation with the M_*, because of the large angular variations present in the massive systems.

We would also like to quantify how N_H is related to the overall gas mass in a host. We choose the mass of molecular gas mass as a proxy for gas mass because it is a direct observable, and also is the dominant contributor to the obscuration, since in high-density regions the molecular fraction is close to 1. In the middle panel of Fig. 12, we split the overall N_H distribution in multiple M_{H_2} bins of QSO hosts. Similar to the trend for M_*, we find a higher probability peak and larger N_H dispersion for QSOs with larger molecular gas mass. For QSOs with $M_{H_2} > 10^{10.4}$ M$_\odot$ (purple line), the probability of getting $N_H < 10^{23}$ cm^{-2} is 22 per cent. Linear regression of $\log N_H$ and $\log M_{H_2}$ gives

$$\log N_H = (0.47 \pm 0.03) \log M_{H_2} + (18.4 \pm 0.3) \quad (13)$$

with intrinsic scatter $\epsilon = 0.31$ dex.

The lower panel of Fig. 13 shows the mass of molecular gas M_{H_2} in the host versus M_* for the same QSO population at $z = 7$. There is a tight positive correlation between M_* and M_{H_2} that can be fitted by $\log M_{H_2} = 0.53 \log M_* + 4.9$. The coefficient 0.53, the slope of the mean relation between M_* and M_{H_2}, explains why the $\log N_H$ slope with respect to M_{H_2} is steeper than that with M_* ($0.53 \times 0.47 \sim 0.24$). This indicates that the positive correlation between N_H and M_* is driven by the fact that larger galaxies with a higher stellar mass also contain more abundant cold molecular gas in the galaxy centre that contributes to AGN obscuration. We now compare our molecular gas data with observations of high-redshift QSO hosts. The green data points with error bars in Fig. 13 are the observations of three QSO host galaxies residing at redshift $6.6 < z < 6.9$ from ALMA observations (Venemans et al. 2017). The brown data point is the observation of a $z = 6.42$ QSO host by Stefan et al. (2015). In the observations, the galaxy (stellar) mass was inferred by subtracting the mass of molecular gas from the dynamical mass. We see that the molecular masses predicted by our simulation are broadly at the level seen in current observations of $z \sim 7$ AGN hosts; this provides some additional validation to the galactic QSO obscuration inferred here.

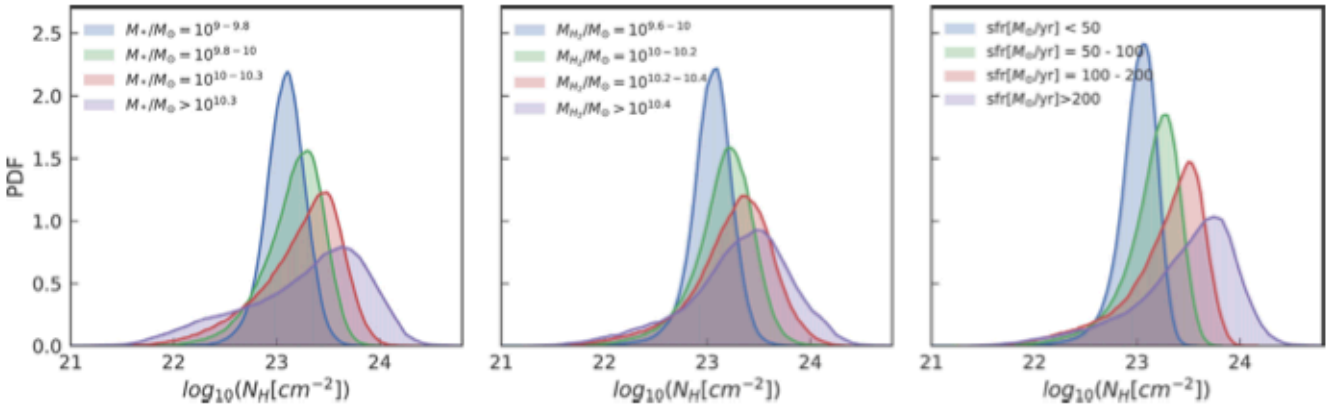


Figure 12. As in Fig. 9, here we plot the overall $N_{\rm H}$ histogram by splitting the QSO population with respect to their host properties: stellar mass (left), molecular mass (middle), and star formation rate (right). All the quantities are calculated within the virial radius R_{200} of the QSO host. We include all QSOs with $L_{\rm X} > 10^{43}$ erg s^{-1} at $z = 7.0$.

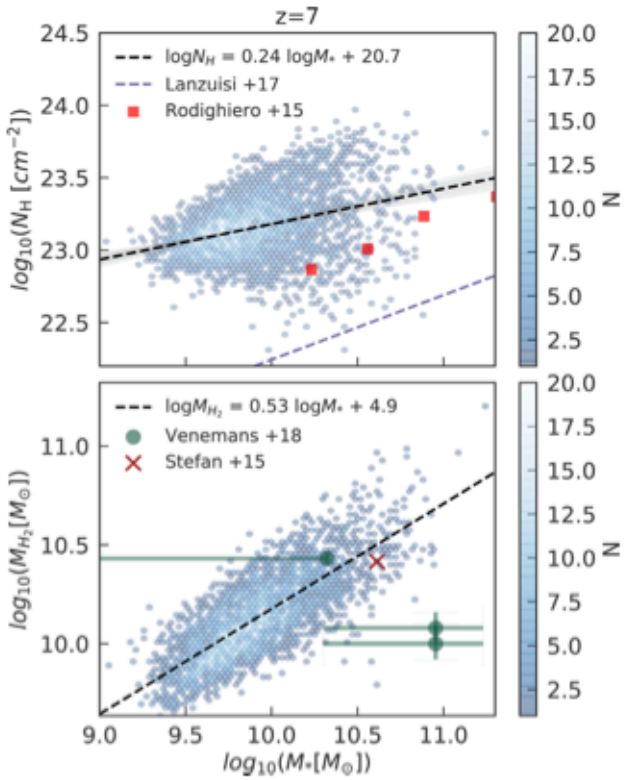


Figure 13. The gas properties of AGN versus the stellar mass of the host galaxy, for the QSO population with $L_{\rm X} > 10^{43}$ erg s^{-1} at $z = 7.0$. *Upper panel*: The blue points give $N_{\rm H}$ for AGN versus the stellar mass of the host galaxy along a random line of sight. The grey areas show the linear fits for 972 realizations of $N_{\rm H}$ versus M_* along different lines of sight, and the black dashed line gives the averaged fitting value. The red squares show results from the observations of Rodighiero et al. (2015) at $z \sim 2$, and the purple dashed line shows the relation found by Lanzuisi et al. (2017) at $2 < z < 4$. *Lower panel*: The molecular mass in the host halo of AGN versus the stellar mass of the host galaxy. The green points with error bars denote the observations of three quasar host galaxies at $6.6 < z < 6.9$ from Venemans et al. (2017). The brown cross represents the measurement from Stefan et al. (2015) for the $z = 6.42$ host of QSO J1148+5251.

We also investigate the relation between $N_{\rm H}$ and the total SFR of the host galaxy. Here the SFR is calculated within the virial radius, R_{200} of the halo. Since the SFR correlates strongly with the local density field and the amount of molecular gas, it is not surprising to find a similar trend as that for $N_{\rm H}$ with $M_{\rm H_2}$. As shown in the third panel of Fig. 12, we split the overall $N_{\rm H}$ distribution with respect to host SFR. It turns out that the $N_{\rm H}$ distribution systematically moves to higher values when we look at hosts with a larger SFR. At the same time, low-$N_{\rm H}$ tail remains, due to the outflows driven by the massive systems. Quantitatively, for QSOs with SFR $> 200\,{\rm M}_\odot\,{\rm yr}^{-1}$ (purple line), the probability of finding $N_{\rm H} < 10^{23}\,{\rm cm}^{-2}$ is 16 per cent. Linear regression between $\log N_{\rm H}$ and the SFR gives

$$\log N_{\rm H} = (0.48 \pm 0.02)\log({\rm SFR}) + (22.31 \pm 0.04), \quad (14)$$

where the SFR is in unit of ${\rm M}_\odot\,{\rm yr}^{-1}$. The intrinsic scatter of the above fitting gives $\epsilon = 0.29$ dex. Compared to M_* and $M_{\rm H_2}$ in the host, $N_{\rm H}$ has slightly tighter correlation with the SFR, since the latter strongly correlates with the cold molecular gas concentrated in the galaxy centre.

Finally, we note that all these relations have about 0.3 dex intrinsic scatter that is likely induced by $N_{\rm H}$ measured for all lines of sights and is sensitive to the angular variations of gas in the inner regions, further complicated by the effects of AGN feedback and associated outflows.

3.4 Dust extinction and UVLF of QSOs

In this section, we evaluate the dust extinction of the QSO hosts in the UV band and make predictions for the corresponding UV luminosity function of QSOs. We calculate the UV-band dust optical depth $\tau_{\rm UV}$ for each line of sight surrounding the AGN and derive the dust-extincted UVLF taking into account the angular variations of $\tau_{\rm UV}$.

As illustrated in the middle column of Fig. 3(a), the dust optical depth $\tau_{\rm UV}$ has a similar pattern to the $N_{\rm H}$ map, indicating that a high level of dust extinction is more likely to happen for directions with high column density $N_{\rm H}$. In Fig. 14, we give the 2D histogram of $N_{\rm H}$ of each line of sight surrounding the subset of the AGN population with $L_{\rm X} > 10^{43}$ erg s^{-1} at $z = 7$, plotted against their $\tau_{\rm UV}$ values. The calculation of $\tau_{\rm UV}$ is based on equation (9) with the assumption that the dust extinction is proportional to the metal column density along the line of sight. We see that for certain $N_{\rm H}$ values (for example, $N_{\rm H} \sim 10^{23}\,{\rm cm}^{-2}$), the corresponding optical depths $\tau_{\rm UV}$ can sometimes be over 1.5 dex because of variations in the metallicity. However, $N_{\rm H}$ and $\tau_{\rm UV}$ have an overall strongly positive correlation, indicating that the high-density regions are

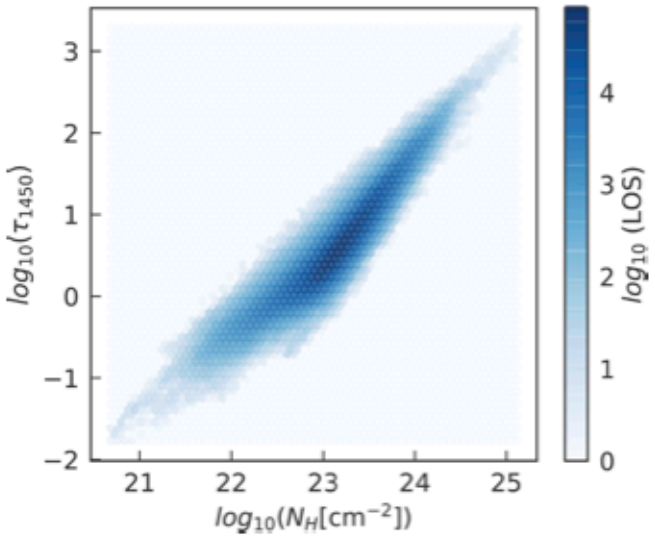


Figure 14. 2D histogram of optical depth in the UV band (τ_{1450}) due to dust grains versus the corresponding $N_{\rm H}$ for the entire AGN population with $L_{\rm X} > 10^{43}$ erg s^{-1} at $z = 7$. The histogram is based on all lines of sight (i.e. 972 lines of sight for every AGN with $L_{\rm X} > 10^{43}$ erg s^{-1} at $z = 7.0$).

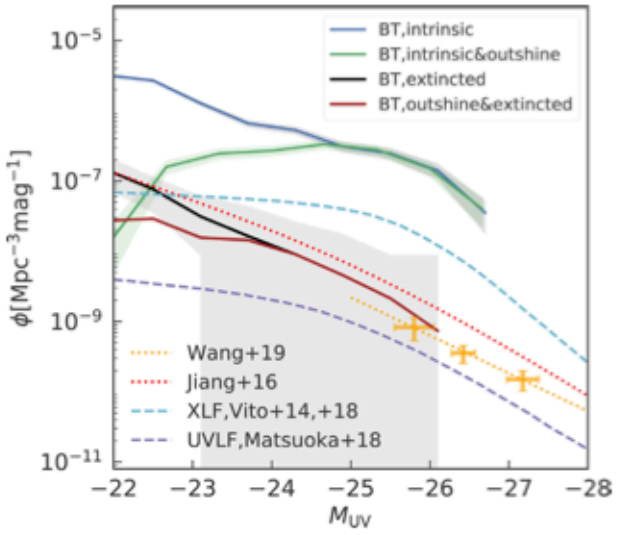


Figure 15. UVLF of QSOs. The blue solid line is the intrinsic UV-band luminosity of QSOs predicted by BLUETIDES at $z = 7.0$. The black solid line is the result of taking dust obscuration into consideration. The grey shaded area gives the range of realizations with respect to different lines of sight (i.e. the result for the UVLF observed along 972 different directions). The green solid line gives the UVLF for QSO population that have intrinsic UV-band luminosity higher than that of their host galaxies, and the red solid line is the corresponding dust-extincted LF. The orange solid symbols with error bars are the measured $z \sim 6.7$ UVLF from Wang et al. (2018). The red dotted line is the $z \sim 6$ fitted UVLF measured by Jiang et al. (2016). The purple dashed line gives the luminosity function of Matsuoka et al. (2018) based on the $5.7 < z < 6.5$ quasar population and extrapolated to $z = 7$. The cyan dashed line shows the QLF by Vito et al. (2018) extrapolated to $z = 7$ (from $z = 4$ AGN population).

more likely to yield higher levels of dust extinction. This in turn implies that the variations in metallicity along the lines of sight is subdominant compared to the density variations.

With knowledge of UV-band extinction for each line of sight, we can predict the dust-extincted UVLF of QSOs taking into account the angular variation of dust extinction. The solid blue line in Fig. 15 shows the intrinsic UVLF at $z = 7$. To predict the dust-extincted UVLF, we collect all the realizations of the dust-extincted UVLF observed from the 972 different lines of sight and take the average value, which we plot with a black solid line in Fig. 15. The grey shaded area gives the estimated error from the 2σ bounds of the 972 collections of dust-extincted UVLFs.

We can see that the dust-extincted UVLF is about 1.5 dex lower than the intrinsic UVLF, implying that more than 99 per cent of the $z \sim 7$ AGNs are heavily dust extincted and might be missed by UV-band observations. Observations of $z > 7$ QSOs also lead to inference of a high level of obscuration in the UV band. Analysis based on the UV spectra of two $z > 7$ QSOs (Davies et al. 2019) indicates that, given the total number of ionizing photons and the accreted black hole mass, the QSO could be obscured over more than 82 per cent of its lifetime (with the assumption of similar radiative efficiency as for low-redshift QSOs). However, since our simulation does not keep enough snapshots to track the time evolution of all QSO hosts, we cannot directly compare with this result by properly tracking the obscuration for specific QSOs on a time averaged basis.

In Fig. 15, we also show some observational data for high-z QSO populations, using dashed and dotted lines. The red dotted line shows the UVLF fitting result of Jiang et al. (2016), which is based on observations of $z \sim 6$ QSOs. The orange data points with error bars are the measured binned UVLF from Wang et al. (2018). The purple dashed line shows the UVLF from Matsuoka et al. (2018) based on observations of $5.7 < z < 6.5$ QSOs. The cyan dashed line is the LF given by Vito et al. (2018), inferred from X-ray observations on $z = 4$ AGN population and extrapolated to $z = 7$. The X-ray LF is converted to UV-band LF via equations (5) and (6). Note that the X-ray-band LF also includes the obscured AGN population while the UVLFs are only for optically selected QSOs.

As indicated by the blue and purple dashed lines, some observations of QSO luminosity functions favour a flattened feature at the faint end. To explore the possible factors that might affect the shape of the UVLF at faint end, we select the QSOs that have an intrinsic UV-band luminosity higher than that of their host galaxy (i.e. the points below the dashed line in Fig. 1) and plot the intrinsic and corresponding dust-extincted UV luminosity functions as green and red solid lines, respectively. This gives us an estimate of how the shape of UVLF changes if we take account of the fact that fainter AGNs might be outshone by their host galaxy and therefore missed by UV observation.

4 SUMMARY

In this work, we study the galactic obscuration surrounding the $z = 7$ QSO population based on the BLUETIDES cosmological hydrodynamic simulation, which in its latest run has just reached these redshifts. We examine the angular variations of the gas obscuring AGNs as well as its correlation with the gas outflows driven by the AGN feedback. With the large size of the BLUETIDES simulation, we are able to study the obscuration state of QSOs on a statistical basis. We explore the relationship between the $N_{\rm H}$ distribution and AGN properties (intrinsic luminosity, BH mass, and the Eddington accretion ratio) and host properties (gas outflow, stellar mass, molecular gas mass, and the SFR), as well as tracing the time evolution of the overall $N_{\rm H}$ distribution. We also evaluate the UV-band dust extinction and make predictions for the UV-band QSO luminosity function.

Our main results are summarized below. Note that the statistics and linear regression fits below are calculated based on the QSO population with $L_X > 10^{43}$ erg s^{-1} at $z = 7$.

(i) For bright AGNs at $z > 7$, galactic obscuration can exhibit large angular variations, spanning over 2 orders of magnitude for different lines of sight. The angular directions with low column density N_H (i.e. less obscured) are clearly correlated with the gas outflows driven by AGN feedback. Strong gas outflows can open up low-column density regions with $N_H \sim 10^{21}$ cm^{-2}. Overall, for lines of sight with radial outflow velocities $v_r > 300$ km s^{-1}, 87 per cent have $N_H < 10^{23}$ cm^{-2} at $z = 7$. The morphology of the outflows typically shows a bimodal structure, as indicated in Fig. 8.

(ii) The host ISM for $z > 7$ AGN is able to produce absorption up to Compton-thick level ($N_H > 10^{24}$ cm^{-2}), which is in contrast with the study of low-z AGNs where it is claimed that Compton-thick obscuration can only be produced by parsec-scale gas in the nuclear region (Buchner & Bauer 2017).

(iii) For lines of sight with $N_H > 10^{22}$ cm^{-2}, the obscuration is mostly contributed by high-density clumps in the inner regions of galaxies, within $r < $ ckpc h^{-1} of the central QSO. More quantitatively, on average we have $N_H(< 10\,\mathrm{ckpc}\,h^{-1}) \gtrsim 90$ per cent $N_H(< 30\,\mathrm{ckpc}\,h^{-1})$.

(iv) The N_H distribution has a positive correlation with QSO luminosity, though when going to higher luminosity, N_H has a larger scatter and skews towards the low-N_H end, driven by the AGN feedback. We fit a linear relationship between $\log N_H$ and $\log L_X$, finding $\log N_H = (0.42 \pm 0.02)\log L_X + (5.1 \pm 0.7)$.

(v) The obscured fraction (defined as the fraction of sightlines to AGNs with $N_H > 10^{23}$ cm^{-2}) typically ranges from 0.7 to 1.0 for $L_X > 10^{43}$ erg s^{-1} AGNs. For AGNs with higher luminosity ($L_X > 10^{44}$ erg s^{-1}), the obscured fraction does not show a strong trend with the luminosity because strong feedback driven by the bright QSOs maintains the fraction of lines of sight with low N_H.

(vi) A similar trend is found to exist between the N_H distribution and BH mass, while the variations of N_H are larger for AGNs in the most massive M_{BH} bins. For BH masses with $M_{BH} > 10^{7.7}$ M$_\odot$, the probability of finding $N_H < 10^{23}$ cm^{-2} is 27 per cent. Linear regression between N_H and M_{BH} gives $\log N_H = (0.13 \pm 0.02)\log M_{BH} + (22.3 \pm 0.2)$.

(vii) The QSO population with large Eddington accretion ratio λ_{Edd} has an overall higher N_H distribution compared to the population with low λ_{Edd}. For QSOs with $\lambda_{Edd} < 0.1$, the probability of finding $N_H < 10^{23}$ cm^{-2} is 40 per cent, while for $\lambda_{Edd} > 0.7$, the probability is only 4 per cent.

(viii) There is no strong redshift evolution of the N_H distribution around QSOs above specific luminosity cuts.

(ix) With regard to host galaxy properties, the trend is for the AGN population in more massive system to have the overall N_H distribution peaking at higher values because larger systems contain more obscuring gas in the galaxy centre, but at the same time becoming more broadened and skewed towards the low-N_H end as a consequence of the AGN-driven feedback. We split N_H by stellar mass, molecular mass, and the SFR in the host, giving linear regression results of $\log N_H = (0.24 \pm 0.03)\log M_* + (20.7 \pm 0.3)$, $\log N_H = (0.47 \pm 0.03)\log M_{H_2} + (18.4 \pm 0.3)$, and $\log N_H = (0.48 \pm 0.02)\log(\mathrm{SFR}) + (22.31 \pm 0.04)$. We note, however, that all of these relations display a large intrinsic scatter typically of the order of 0.3 dex. This is due to the large angular variations of N_H that are induced by the complex structure of the gas in the inner region of the galaxies and the effects of AGN feedback/outflows.

(x) The dust optical depth τ_{UV} has a tight positive correlation with N_H. The regions with large dust extinction are more likely to have high N_H, and the gas metallicity Z only modulates the variation at a subdominant level.

(xi) Our dust-extincted UVLF is about 1.5 dex lower than the intrinsic UVLF, implying that more than 99 per cent of the $z \sim 7$ AGNs are heavily dust extincted and will be missed by the UV-band observations. We expect that upcoming and future X-ray missions [e.g. Athena (Barcons et al. 2012), Lynx (The Lynx Team 2018), and AXIS (Mushotzky et al. 2019)] will be able to reveal the hidden population (in the UV band) of obscured AGNs in the future. More detailed predictions for the number densities of QSOs expected for these deep X-ray surveys will be reserved to a follow-up study.

Finally, we note that, given our findings that QSO-driven feedback is the crucial factor for generating the low-N_H tail (which allows the observation of the high-z QSOs), it may be possible to use the inferred AGN obscured fraction at high-z to provide some constraints for the strength of the AGN feedback. We expect that the future observational samples of $z > 7$ QSOs from the upcoming X-ray missions could provide a good test to the AGN feedback models.

ACKNOWLEDGEMENTS

We thank Fabio Vito for invaluable discussions and comments on this work. YN thanks the help from Nick Gnedin for inspiring suggestions. The BLUETIDES simulation was run on the BlueWaters facility at the National Center for Supercomputing Applications. TDM acknowledges funding from NSF ACI-1614853, NSF AST-1517593, NSF AST-1616168, and NASA ATP 19-ATP19-0084. TDM and RACC also acknowledge funding from NASA ATP 80NSSC18K101 and NASA ATP NNX17AK56G.

REFERENCES

Antonucci R., 1993, ARA&A, 31, 473
Bañados E. et al., 2018, Nature, 553, 473
Barcons X. et al., 2012, preprint (arXiv:1207.2745)
Battaglia N., Trac H., Cen R., Loeb A., 2013, ApJ, 776, 81
Bhowmick A. K., Di Matteo T., Feng Y., Lanusse F., 2018, MNRAS, 474, 5393
Buchner J., Bauer F. E., 2017, MNRAS, 465, 4348
Buchner J. et al., 2015, ApJ, 802, 89
Buchner J., Schulze S., Bauer F. E., 2017, MNRAS, 464, 4545
Circosta C. et al., 2019, A&A, 623, A172
D'Amato Q. et al., 2020, A&A, 636, A37
Davies F. B., Hennawi J. F., Eilers A.-C., 2019, ApJ, 884, L19
Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
Di Matteo T., Croft R. A. C., Feng Y., Waters D., Wilkins S., 2017, MNRAS, 467, 4243
Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, Publ. Astron. Soc. Aust., 34, e058
Fan X. et al., 2019, BAAS, 51, 121
Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416
Feng Y., Di Matteo T., Croft R., Tenneti A., Bird S., Battaglia N., Wilkins S., 2015, ApJ, 808, L17
Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778
Fontanot F., Cristiani S., Vanzella E., 2012, MNRAS, 425, 1413
Hickox R. C., Alexander D. M., 2018, ARA&A, 56, 625
Hinshaw G. et al., 2013, ApJS, 208, 19

Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2005, ApJ, 632, 81
Hopkins P. F., Richards G. T., Hernquist L., 2007, ApJ, 654, 731
Hopkins P. F., Torrey P., Faucher-Giguère C.-A., Quataert E., Murray N., 2016, MNRAS, 458, 816
Huang K.-W., Di Matteo T., Bhowmick A. K., Feng Y., Ma C.-P., 2018, MNRAS, 478, 5063
Jiang L. et al., 2016, ApJ, 833, 222
Katz N., Hernquist L., Weinberg D. H., 1999, ApJ, 523, 463
Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36
Lanzuisi G. et al., 2017, A&A, 602, A123
Lupi A., Volonteri M., Decarli R., Bovino S., Silk J., Bergeron J., 2019, MNRAS, 488, 4004
Maiolino R., Rieke G. H., 1995, ApJ, 454, 95
Marshall M. A., Ni Y., Matteo T. D., Wyithe J. S. B., Wilkins S., Croft R. A., 2019, preprint (arXiv:1912.03428)
Mateos S. et al., 2017, ApJ, 841, L18
Matsuoka Y. et al., 2018, ApJ, 869, 150
Matsuoka Y. et al., 2019, ApJ, 872, L2
Merloni A. et al., 2014, MNRAS, 437, 3550
Mortlock D. J. et al., 2011, Nature, 474, 616
Mushotzky R. F. et al., 2019, BAAS, 51, 107
Nelson D. et al., 2015, Astron. Comput., 13, 12
Ni Y., Di Matteo T., Feng Y., Croft R. A. C., Tenneti A., 2018, MNRAS, 481, 4877
Pezzulli E., Valiante R., Orofino M. C., Schneider R., Gallerani S., Sbarrato T., 2017, MNRAS, 466, 2131
Ricci C. et al., 2017, Nature, 549, 488
Rodighiero G. et al., 2015, ApJ, 800, L10
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Stefan I. I. et al., 2015, MNRAS, 451, 1713
Tacconi L. J. et al., 2013, ApJ, 768, 74
Tenneti A., Di Matteo T., Croft R., Garcia T., Feng Y., 2018, MNRAS, 474, 597
The Lynx Team, 2018, preprint (arXiv:1809.09642)
Trebitsch M., Volonteri M., Dubois Y., 2019, MNRAS, 487, 819
Ueda Y., Akiyama M., Hasinger G., Miyaji T., Watson M. G., 2014, ApJ, 786, 104
Urry C. M., Padovani P., 1995, PASP, 107, 803
Venemans B. P. et al., 2017, ApJ, 845, 154
Vito F., Gilli R., Vignali C., Comastri A., Brusa M., Cappelluti N., Iwasawa K., 2014, MNRAS, 445, 3557
Vito F. et al., 2018, MNRAS, 473, 2378
Vito F. et al., 2019, A&A, 628, L6
Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Wang F. et al., 2018, ApJ, 884, 30
Waters D., Di Matteo T., Feng Y., Wilkins S. M., Croft R. A. C., 2016a, MNRAS, 463, 3520
Waters D., Wilkins S. M., Di Matteo T., Feng Y., Croft R., Nagai D., 2016b, MNRAS, 461, L51
Wilkins S. M., Feng Y., Di-Matteo T., Croft R., Stanway E. R., Bouwens R. J., Thomas P., 2016, MNRAS, 458, L6
Wilkins S. M., Feng Y., Di Matteo T., Croft R., Lovell C. C., Waters D., 2017, MNRAS, 469, 2517
Wilkins S. M., Feng Y., Di Matteo T., Croft R., Lovell C. C., Thomas P., 2018, MNRAS, 473, 5363
Yang J. et al., 2019, AJ, 157, 236

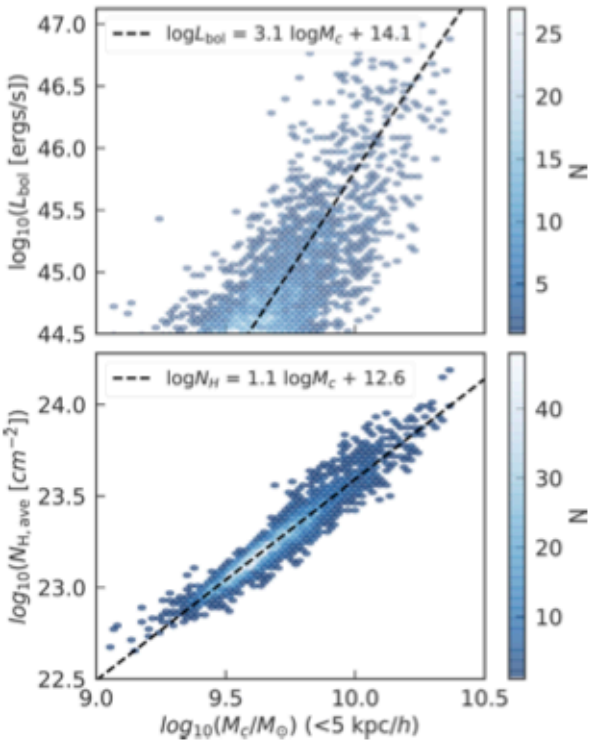


Figure A1. The relation between $L_{\rm bol}$, $N_{\rm H,\,ave}$, and core mass $M_{\rm c}$. Here $M_{\rm c}$ is the mass of cold molecular gas surrounding the central AGN within 5 ckpch^{-1}. $N_{\rm H,\,ave}$ shown in the second panel is the $N_{\rm H}$ value averaged over all lines of sights for each AGN.

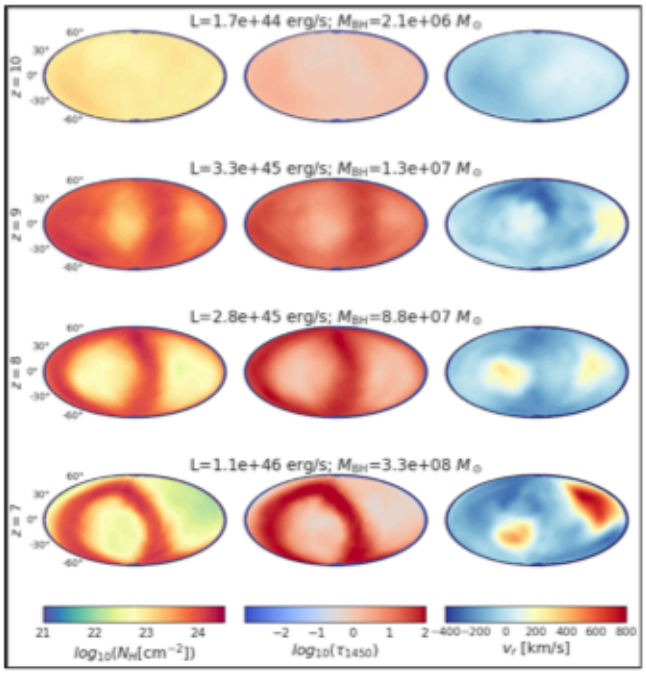
APPENDIX A: RELATION BETWEEN $N_{\rm H}$ AND $M_{\rm C}$

The power-law relation $N_{\rm H} \sim L_{\rm bol}^{1/3}$ discussed in Section 3.3.2 can be interpreted by the scaling relation of $N_{\rm H} \sim M_{\rm c}/R_{\rm c}^2$ and $L_{\rm bol} \sim \rho M_{\rm BH}^2 \sim M_{\rm c}^3$. The latter scaling is from the Bondi formula (cf. equation 1) and the assumption that $M_{\rm BH} \propto M_{\rm c}$. These proposed scaling relations are indeed broadly consistent with the results from our simulations.

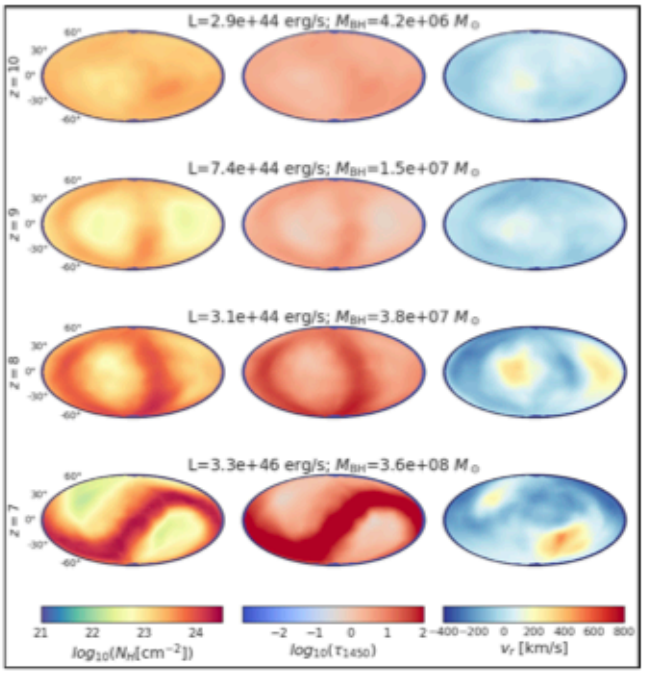
In Fig. A1, we plot the relation between $L_{\rm bol}$, $N_{\rm H,\,ave}$, and core mass $M_{\rm c}$ for the studied AGN population at $z = 7$. Here $M_{\rm c}$ is the amount of cold molecular gas within the vicinity of AGN, with radius $R_{\rm c} = 5$ ckpch^{-1} the averaged half-mass radius of the host galaxies. $N_{\rm H,\,ave}$ is the $N_{\rm H}$ value averaged over all lines of sights for each AGN. The linear fittings show the scaling relation that $L_{\rm bol} \sim M_{\rm c}^3$ and $N_{\rm H} \sim M_{\rm c}$.

APPENDIX B: AITOFF MAP OF SAMPLE QSOS

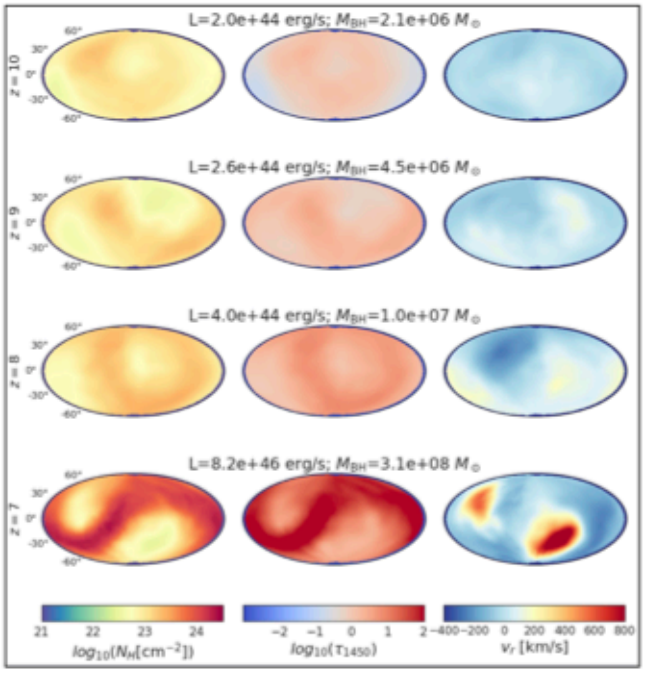
Here we give the full $N_{\rm H}$, $\tau_{\rm UV}$, and v_r maps of three QSO samples BH2, BH3, and BH4 (repeating the format of Fig. 3a) as an extension of Fig. 3, to further illustrate the correlation between the direction with low column density and high outward velocity.



(a) BH2



(b) BH3



(c) BH4

Figure B1. Same as Fig. 3(a); here we plot the map of $N_{\rm H}$, $\tau_{\rm UV}$, and averaged radial velocity $v_{\rm r}$ of the other three sample QSOs. Each row represents the state at a certain redshift, evolving from $z = 10$ to 7 (from top to bottom). The leftmost column is the map of hydrogen column density $N_{\rm H}$. The middle column is the map of dust optical depth $\tau_{\rm UV}$. The right column gives the averaged radial velocity $v_{\rm r}$ in each line of sight. The title gives the BH bolometric luminosity and BH mass at the corresponding redshift.

This paper has been typeset from a TEX/LATEX file prepared by the author.