Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **520,** 3626–3640 (2023) https://doi.org/10.1093/mnras/stad293
Advance Access publication 2023 January 31

Baryonic post-processing of *N*-body simulations, with application to fast radio bursts

Ian M. Williams [1★] Adnan Khan[2] and Matthew McQuinn[1★]

[1] *University of Washington, Department of Astronomy, 3910 15th Ave NE, Seattle, WA 98195, USA*
[2] *University of Washington, Department of Physics, 3910 15th Ave NE, Seattle, WA 98195, USA*

Accepted 2023 January 23. Received 2023 January 20; in original form 2022 July 20

ABSTRACT

Where the cosmic baryons lie in and around galactic dark matter haloes is only weakly constrained. We develop a method to quickly paint on models for their distribution. Our approach uses the statistical advantages of *N*-body simulations, while painting on the profile of gas around individual haloes in ways that can be motivated by semi-analytic models or zoom-in hydrodynamic simulations of galaxies. Possible applications of the algorithm include extragalactic dispersion measures to fast radio bursts (FRBs), the Sunyaev–Zeldovich effect, baryonic effects on weak lensing, and cosmic metal enrichment. As an initial application, we use this tool to investigate how the baryonic profile of foreground galactic-mass haloes affects the statistics of the dispersion measure (DM) towards cosmological FRBs. We show that the distribution of DM is sensitive to the distribution of baryons in galactic haloes, with viable gas profile models having significantly different probability distributions for DM to a given redshift. We also investigate the requirements to statistically measure the circumgalactic electron profile for FRB analyses that stack DM with impact parameter to foreground galaxies, quantifying the size of the contaminating 'two-halo' term from correlated systems and the number of FRBs for a high significance detection. Publicly available PYTHON modules implement our `CGMBrush` algorithm.

Key words: gravitational lensing: weak – methods: numerical – software: simulations – cosmic background radiation – fast radio bursts.

1 INTRODUCTION

In a simpler universe in which the baryons could not cool, they would essentially trace the dark matter. However, our universe deviates far from this simple picture, as the baryons can cool and condense to high densities in dark matter haloes. Yet, even adding cooling does not match the real complexity. If there were only cooling, the baryons associated with each dark matter halo would have condensed into stars – the Milky Way galaxy would have been ten times more massive (e.g. Behroozi, Conroy & Wechsler 2010; Tumlinson, Peeples & Werk 2017). Instead, galaxies are host for violent phenomena like supernovae and active galactic nuclei that prevent such massive galaxies from forming. These phenomena inject energy and ultimately redistribute the gas, possibly even to distances well outside their dark matter haloes. In addition to learning the locations of the baryons at low redshifts, understanding this redistribution is likely key to such big questions as why some galaxies have stopped accreting gas and are becoming red and dead (Tumlinson et al. 2017). It is also important for realizing precision cosmology with weak lensing (Chisari et al. 2018; Schneider et al. 2019; Debackere, Schaye & Hoekstra 2020; Osato, Liu & Haiman 2021).

Most of the understanding of the distribution of baryons around haloes is derived from observations of absorption lines against a bright background source, typically a quasar (Bahcall & Spitzer 1969; Thom et al. 2012; McQuinn 2016). This probe has revealed large column densities in atomic hydrogen and in many metal ions to impact parameters that extend 100–200 kpc from their foreground galaxies. However, while photoionization modelling has been used to constrain the amount of 10^4 K gas (e.g. Werk et al. 2014), there is no direct way to use these observations to probe the bulk properties and radial profile of all the gas (most of which is likely at higher temperatures). Our best bet is the absorption from highly ionized oxygen (Tumlinson et al. 2011; Faerman, Sternberg & McKee 2017; Voit 2019), but, even when this absorption does trace the bulk of the gas, metallicity and temperature uncertainties still limit this probe's constraining power. Oxygen absorption constraints on the baryonic profile have met the most success around the Milky Way itself, particularly when combined with observations of oxygen emission (Henley & Shelton 2012; Miller & Bregman 2013, 2015; Voit 2019).

More direct probes of the bulk (virialized) gas in the circumgalactic medium (CGM) are X-ray emission as well as the thermal and kinetic Sunyaev–Zeldovich (SZ) effects (Sunyaev & Zeldovich 1970, 1972), which measure, respectively, the inverse Compton heating and Doppler shifting of CMB photons when scattering off free electrons. When stacking thousands of haloes, the thermal SZ has been used to measure the projected pressure profile of the gas in haloes down to $\sim 10^{13}\,M_\odot$ (Planck Collaboration 2013; Greco et al. 2015; Schaan et al. 2016; Amodeo et al. 2021), and X-ray emission to somewhat lower haloes masses (Bregman et al. 2018; Chadayammuri et al. 2022). The kinetic SZ is a new frontier that has recently been used to

★ E-mail: ianw89@live.com (IMW); mcquinn@uw.edu (MMcQ)

measure the electron profile in the $\sim 10^{13.5}$ $M_\odot$ haloes of massive luminous red galaxies (Amodeo et al. 2021). One limitation for both SZ effects is that the wide beam of CMB instruments makes it difficult to probe structure below the virial radius of galactic haloes, whereas (while higher angular resolution) X-ray emission is most sensitive to the denser gas in the inner halo. Additionally, the kinetic SZ stacking requires redshifts to reconstruct the velocities, and the largest spectroscopic catalogues are of galaxies considerably more massive than the Milky Way.

A new class of sources that provide an avenue for localizing the baryons is fast radio bursts (FRBs), which were rediscovered by Thornton et al. (2013), confirming an earlier report (Lorimer et al. 2007). FRBs are a class of bright extragalactic millisecond radio transients, which likely are sourced by magnetars (Cordes & Chatterjee 2019; Petroff, Hessels & Lorimer 2019). There is a detectable FRB roughly once every minute somewhere on the sky, of which only a small fraction are currently detected. However, many wide-field instruments are now coming online and are expected to increase the number of detections by orders of magnitude (see Petroff, Hessels & Lorimer 2022 for a review). FRB observations directly provide the amount of electrons along each sightline as the signal traverses through the intervening medium. The dispersion measure (DM) of this signal is an observed quantity that measures the delay in the arrival time – caused by the total electron column along the line of sight – as a function of the frequency (Deng & Zhang 2014; McQuinn 2014; Shull & Danforth 2018; Macquart et al. 2020; Simha et al. 2021; Connor & Ravi 2022; Lee et al. 2022).[1] The estimates of McQuinn (2014) suggest that FRBs may excel at constraining the gas profiles around galaxies.

This paper develops a new technique to model the distribution of baryons, and is applicable to all these CGM observables. As an initial application, we focus on how the baryonic profile of foreground haloes affects the statistics of DM towards cosmological FRBs. Our technique starts with an *N*-body simulation, which is the primary numerical method for modelling non-linear structure formation in hundreds of megaparsec regions (Centrella & Melott 1983; Davis et al. 1985). *N*-body simulations do not follow how gas cools and is redistributed by feedback. Hydrodynamic simulations of galaxy formation are the preferred method to capture this feedback physics (e.g. Fielding et al. 2017; Stern et al. 2019; Li & Tonnesen 2020; Lochhaas et al. 2020; Su et al. 2020). Since the energy injection from supernovae and AGN is not resolved, these hydrodynamic simulations adopt varied prescriptions for injecting energy into the gas on resolved scales, around sites of star formation and AGN activity in the simulation (Somerville & Davé 2015; Faucher-Giguère et al. 2016). The relatively small box of these hydrodynamic simulations means that they still struggle to adequately sample the cosmological density field (Zhu, Feng & Zhang 2018), an issue that is significantly mollified in the largest *N*-body simulations. Thus, we opt for a hybrid approach that uses the statistical advantages of large *N*-body simulations, while painting on the distribution of gas around individual haloes in ways that can be motivated by zoom-in hydrodynamic simulations of galaxies (or semi-analytic models). Our method can further use CGM models based on observational measurements or even phenomenological parametrizations to understand what baryonic profiles the data can support.

It is useful to compare our baryon pasting method to other approaches. The 'baryonification' method of Schneider & Teyssier (2015) adds a displacement owing to baryonic pressure and feedback in post-processing to the positions of particles in *N*-body simulations. This method has been shown to reproduce the effect of baryons on the matter power spectrum seen in more complex hydrodynamic simulations and has been used to interpret data (Schneider et al. 2022). A later version, Aricò et al. (2020), has also been favourably compared to hydrodynamic simulations in the context of weak lensing surveys (Lu & Haiman 2021; Lee et al. 2023). The algorithm to build weak lensing convergence maps used in Lu & Haiman (2021) is similar to our approach, as the baryon correction model is applied in 2D density planes. Osato & Nagai (2023) presents two post-processing methods to construct tSZ and kSZ maps from *N*-body outputs. The first method is halo-based, where a spherical profile is assigned to the positions of $\geq 10^{13} M_\odot$ dark matter haloes, ignoring the contribution of matter not associated with such haloes. This approach has the thermal SZ effect in mind, where most of the contribution originates from massive haloes. In their second method, they describe a more expensive particle-based method that includes both halo and field particles. The particle-based method handles halo particles by estimating the dark matter density from an *N*-body simulation and mapping this along with halo mass to the baryonic pressure and density obtained from an intracluster gas model. Particles not associated with haloes contribute to the density, but not the pressure. The key difference between all of these approaches and the `CGMBrush` algorithm presented here is that instead of processing every particle or applying a correction model halo by halo, we operate on a 2D projected density field, bin haloes by mass, and use fast 2D convolutions to apply a gas model to the entire mass bin. This allows faster calculations of full 2D maps with less resources (an ordinary home computer suffices), and our implementation also adds flexibility in easily applying different models.[2]

This paper is organized as follows. We discuss our `CGMBrush` method in Section 2. In Section 3, we investigate the results of an application to FRBs. We summarize our conclusions in Section 4. Appendix A discusses the simulation utilized in our application, Appendix B provides additional justification for our method and Appendix C discusses convergence testing.

2 THE `CGMBrush` METHOD

We present a method that can be applied to any cosmological *N*-body simulation (also generalizable to hydrodynamic simulations). The primary requirement, at least for the FRB application presented here, is that the *N*-body simulation must resolve sub-Milky Way halo masses (down to $\sim 10^{11} M_\odot$) and be in a $\gtrsim$100 comoving Mpc box that is large enough for a cosmologically representative sample of structures. The method realizes where the baryons are located with a significant reduction in the computational cost relative to hydrodynamic simulations of galaxy formation at the expense of some self-consistency.

This method adds to *N*-body simulations one thing they lack – a prescription for the locations of the baryons, especially around galaxies where the baryons' dynamics are distinct from the purely gravitational dynamics of the dark matter owing to cooling and feedback. Algorithmically, we extract the matter associated with

[1] The DM, expressed in the usual pc cm^{-3}, can be converted to column density as $N_e = 1.5 \times 10^{20}$(DM/50) cm^{-2}.

[2] A fast emulator for the matter power spectrum for their flexible seven-parameter baryonic-profile model has been developed in Giri & Schneider (2021) for the Schneider & Teyssier (2015) method, allowing faster parameter space scans for this statistic.

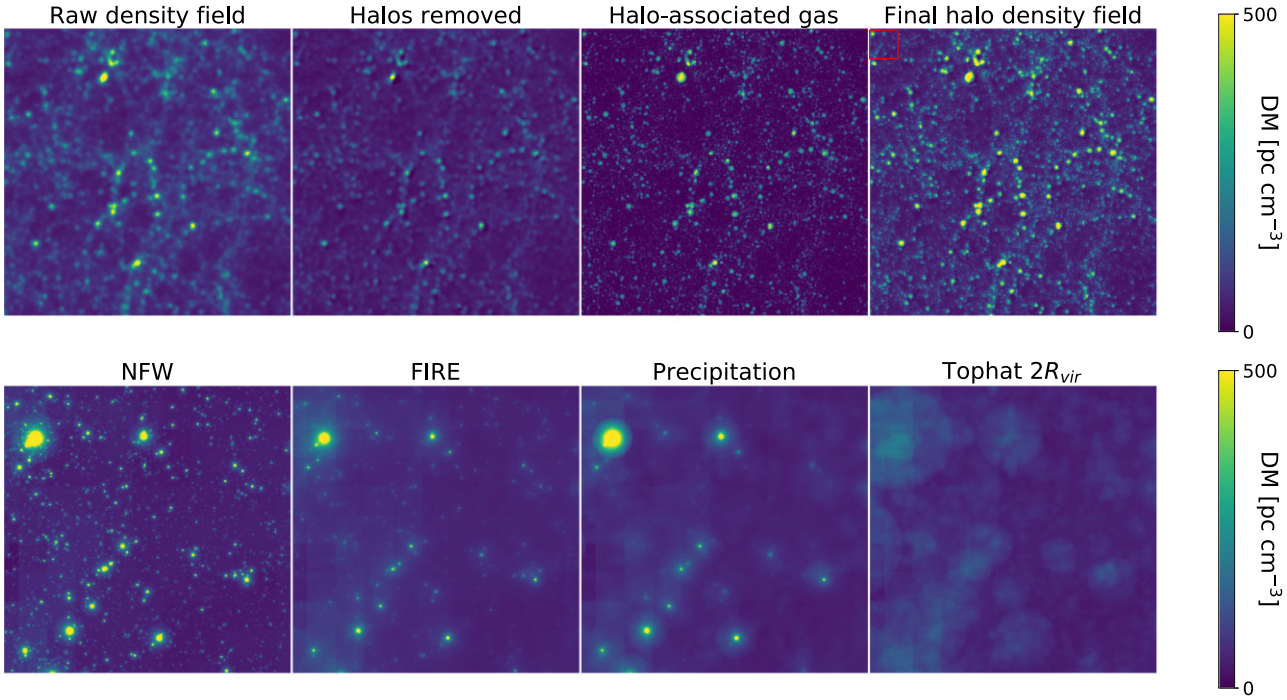

Figure 1. Illustration of the `CGMBrush` method for different baryonic halo profiles applied to a 512^3 grid of the $z = 0$ snapshot of the Bolshoi N-body simulation. Each panel is projected over the simulation boxsize of $250h^{-1}$Mpc. Top panels: Shown is a 140×140 Mpc^2 subregion of the full map of the dispersion measure (DM) – the total electron column density along the line of sight – through different stages of the algorithm. The left-most panel shows the projection of the raw gridded density field from the simulation. From this density field, we subtract the template halo density field, and the result is shown in the second panel. The third panel shows the halo-associated gas density field using the 1 $r_{\rm vir}$ top-hat profile that we wish to add back to the subtracted field, where a fine Eulerian grid of $N_f^2 = 32,768^2$ is used (note that we find such a large N_f is overkill). The right-most panel shows the projected final halo density field after the finely-gridded halo-associated gas density field is added back to the coarse subtracted field. Bottom panels: A zoom-in on the 14×14 Mpc^2 region indicated by the red rectangle in the top-right panel, but shown for several other baryonic halo profiles. From left to right in the bottom panels, we show the final field when re-adding baryons with the NFW, FIRE, Precipitation, and 2 $r_{\rm vir}$ tophat profiles. These models are described in Section 2.2, and we note that the NFW and tophat models are more of a toy character and not meant to be realistic.

dark matter haloes from the density field. We then redistribute the 20 per cent baryon fraction of this matter using physically motivated models for the baryonic profiles around haloes. For the dark matter that lies outside of dark matter haloes, this method assumes that there is a component of the gas that perfectly traces the dark matter with the cosmic ratio. While this is almost certainly true when averaging over megaparsec scales and larger, it may not be on the outskirts of haloes; as discussed in Appendix B, the clustering observables that are of most interest are fortunately not very sensitive to how one models the diffuse gas at halo outskirts.[3]

Large N-body simulations follow typically billions of particles, with the largest surpassing trillions. To reduce the size of outputs, the particles are often projected on to an Eulerian grid with some interpolation kernel (such as 'nearest grid point' or 'cloud-in-cell'; Klypin, Trujillo-Gomez & Primack 2011). Our method has the advantage of working on the combination of these Eulerian grid outputs and a halo field we create from the simulations' catalogue of collapsed haloes. Moreover, the method only requires the 2D projection of these grids, which allows our algorithm to achieve an enormous dynamic range. A fixed grid, while wasteful in many circumstances, is justified for the projected distribution of baryons as many haloes intersect with essentially every sightline.

2.1 Algorithm at a single redshift

We first describe how we generate a baryon field for a single redshift. To start, the algorithm inputs a cosmological density field projected onto an Eulerian grid of N^3 numbers in a box of comoving size L. The algorithm projects the 3D density grids to 2D grids with N^2 cells along a selected axis. An example of this is shown using the $z = 0$ snapshot of the Bolshoi simulation (Klypin et al. 2011) in the top left panel of Fig. 1 (refer to Section A for details on this simulation). The acronym DM on the axes is because the gas column density is identical (aside from redshift factors) to the dispersion measure – an FRB observable considered in detail in later sections.

The next step in our algorithm is to remove the mass associated with dark matter haloes from the Eulerian density field, which requires that we create a template halo density field. We create the template halo density field from the catalogue of collapsed haloes in the simulation, where we require at a minimum their positions

[3]We note that extrapolating a $c = 10$ NFW profile indicates that 40 per cent of the haloes mass lies between one and two virial radii, and detailed calculations show that this percentile is on-average higher for sub M_* haloes and lower for more massive haloes (Wang et al. 2009). This means that even if this mass is also redistributed by feedback, it is a smaller fraction than the virialized mass. The agreement of the observed Ly α forest column density distribution across redshift with simulations that use various feedback models suggests to us that around most haloes gas outside the virial radius is not significantly redistributed (McQuinn 2016), although we note that around Luminous Red Galaxies kinetic SZ observations suggest that such redistribution may even happen out to three virial radii (Amodeo et al. 2021).

and halo masses. First, we divide the haloes into n_M equally spaced logarithmic mass bins between a minimum and maximum halo mass ($M_{\rm min}$ to $M_{\rm max}$) and assume all haloes within a mass bin have the same halo profile. Then we calculate the virial radius of the haloes for each mass bin, assuming the density profiles in a given mass bin specified by M – the average mass in the mass of bin – is described by a single NFW profile (e.g. Cooray & Sheth 2002)

$$\rho(r|M) = \frac{\rho_s}{(r/r_s)[1+(r/r_s)^2]}, \quad (1)$$

parametrized by $r_s(M) = r_{\rm vir}/c$, where $r_{\rm vir}$ and c are given by

$$r_{\rm vir} = \left(\frac{3M}{4\pi[18\pi^2 - 82q(z) - 39q(z)^2]\rho_c}\right)^{1/3}; \quad (2)$$

$$c(m,z) = \frac{9}{1+z}\left(\frac{M}{M_*}\right)^{-0.13}, \quad (3)$$

where $q(z) = \Omega_\Lambda/(\Omega_m(1+z)^3 + \Omega_\Lambda)$, $\rho_c = 3H^2/(8\pi G)$, $M_* = 5 \times 10^{12}{\rm M}_\odot$, and this form for the concentration parameter c is taken from the numerical study of Peacock & Smith (2000). The characteristic density, ρ_s, is set by the mass within the virial radius being M.

We then convolve the halo density profiles with the position of the centres of the haloes (the physical location is provided by the simulation), yielding the halo density field template that we subtract off. Our algorithm does this projection using convolutions performed with fast Fourier transforms, which allows it to be very fast. One worry is that we have assumed a single halo profile, when actual relaxed halo profiles show significant scatter in their triaxiality and concentration (Kravtsov 2013). This problem likely only has a minor effect on our results for reasons detailed in the next paragraph.

The original projection of the N-body particles on to an Eulerian grid is effectively smoothing the density field, and this smoothing must be matched in this subtraction. The simulator has choices in how this smoothing is performed. In our application to the Bolshoi simulation, discussed later, the field is gridded by cloud-in-cell (CIC) interpolation and Gaussian smoothing with a standard deviation of one cell. To create the template halo field, the algorithm puts down the halo profile at the position of each halo and also convolves this profile by the 2D projection of this smoothing. A finer grid is used for this step than the N^2 projected simulation density outputs in order to avoid additional smoothing effects. In the application presented, we choose a 1024^2 grid, slightly larger than our original density field. It then down-grids to the N^2 resolution of the simulation density outputs. This final down-gridded halo density field is then used as a template for subtracting the halo contribution to the density field; the top left-middle panel of Fig. 1 shows the result of the halo field subtracted from the raw density field. Because the algorithm is designed so that the size of the coarse Eulerian grid cell (1 Mpc in our applications) is larger than the virial radius of all but the largest haloes, the method is relatively insensitive to the assumption of a single intrinsic halo profile in each mass bin (as given by equation 1). The cost is that structures outside of haloes (like sheets and filaments) that are smaller than the Eulerian grid cell are not resolved. Appendix B argues that for our envisioned applications these structures are a subdominant source of fluctuations, and this is further quantified by our convergence test presented in Appendix C for different sizes N of the simulation's density grids.

So far we have described how the method creates a halo field to subtract from the density field. The most important step is to create a field that represents the halo-associated gas, which we aim to add back to the subtracted density field. This step takes a profile for halo gas that can be motivated by simulations or models. We describe some choices in Section 2.2. To resolve the sub-halo structures when adding the haloes' gas back with our desired profile, we increase the resolution of the 2D grid by a factor of η, so that our halo baryon grid and final total baryon density grid has resolution $N_f = \eta N$. The third panel from the right in the top row of Fig. 1 shows the halo-associated gas grid for a final output grid of $N_f^2 = 32,768^2$, assuming a simple 3D top-hat model for the halo-associated baryons that extends $1\,r_{\rm vir}$ with all the baryons enclosed. This large N_f is overkill as all statistical quantities we consider are converged for considerably smaller grids, but it demonstrates how much resolution the 2D nature of our algorithm enables.

Finally, we add the halo density field back to the raw density field from which the haloes were previously removed to create a final grid of size N_f^2 (the top right-hand panel in Fig. 1).[4]

The summary of how the final baryonic halo field is generated for a single redshift is as follows:

(i) Start with a simulations density field on a 3D Eulerian grid, and project it over the 'line of sight' dimension to make the N^3 3D grid an N^2 2D grid (the top left-hand panel in Fig. 1). The algorithm is designed to work where the N^2 grid does not resolve the virial radius of most haloes, as will be the case for essentially any computable 3D Eulerian grid of a large N-body simulation.

(ii) Create an N^2 template halo density field by convolving halo centers with a mean NFW halo profile in each halo mass bin and, then, by further applying the same smoothing/interpolation scheme as used to create the Eulerian density grid. (For our application to the Bolshoi simulation, CIC and Gaussian smoothing.)

(iii) Subtract the template halo density field from the 2D density field to create a halo-subtracted baryon field (the top left-middle panel in Fig. 1).

(iv) Add the baryons associated with haloes back to the subtracted density field. First, choose a baryonic profile at each halo mass (see Section 2.2 for potential models), and, next, convolve the halo centers with these profiles. As halo gas can have structure on tens of kiloparsec scales, do this on a grid N_f with much higher spatial resolution than the halo subtracted baryon field ($N_f \gg N$) (the top middle-right panel in Fig. 1).

(v) Add the halo field back to the halo subtracted field to create the full model for the baryonic field on a N_f^2 grid (the top right-hand panel in Fig. 1).

2.2 Halo profiles

The previous section describes a procedure for modelling the distribution of gas around haloes. This allows one to study the effects of different baryonic profiles. Here, we consider several models for the redistribution of gas in and around dark matter haloes, some motivated by their simplicity, others by their analytical nature, and others still by sophisticated zoom-in simulations of the gas around haloes. Our code `CGMBrush` can also take a user specified profile, and supports combining multiple profiles together in a mass-dependent fashion.

In this paper, we consider the following models:

[4]In the implementation presented in this paper, we do not subtract the ∼ 5 per cent of baryons that form stars from the diffuse baryons, even though it does not contribute to the envision statistics such as regarding FRB DM. This omission is justified by the fact that removing this contribution will have a minor effect on any of our results, which are shaped by the large uncertainty in baryonic halo profiles.

3D top hat. This is an unrealistic but instructive density model for the haloes. It assumes the closure density in baryons is distributed uniformly within a 3D sphere of radius r around each halo centre. Perhaps the most motivated value is $r = r_{\rm vir}(M)$, but we also consider twice this extent with $r = 2r_{\rm vir}(M)$.

NFW. The NFW density profile (equation 1) describes the dark matter in haloes rather than the gas (Navarro, Frenk & White 1996). However, it becomes a better approximation for the gas profile in the most massive galaxy clusters. NFW models overestimate the concentration of the gas in smaller haloes because cooling and feedback tend to redistribute the gas to larger radii (e.g. Maller & Bullock 2004).

FIRE. The FIRE profile is based on cosmological hydrodynamic 'zoom-in' simulations with detailed subgrid physics to account for feedback processes (Hopkins et al. 2018). Hafen et al. (2019) show that the CGM density profiles in FIRE are well described by an r^{-2} scaling. Our method requires us to redistribute *all* the baryons associated with these haloes, whereas Hafen et al. (2019) investigated the profile out to the virial radius, which misses baryons redistributed to further extents. Therefore, we assume an exponential cut-off at $r_{\rm max}$ that is determined so that the profile integrates to the total associated gas mass. Thus, our FIRE density profile is given by:

$$\rho = \rho_0 \left(\frac{r}{r_{\rm vir}}\right)^{-2} \exp\left[-r/r_{\rm max}\right]. \tag{4}$$

Hafen et al. (2019) measured the baryons contained within the virial radius of haloes with 10^{10} M$_\odot$, 10^{11} M$_\odot$, 10^{12} M$_\odot$, and $z = 0.25$ using the FIRE cosmological simulations to be 0.1, 0.2, and 0.3. We use this to determine ρ_0 in equation (4) and then set $r_{\rm max}$ to conserve the total mass. We extrapolate the FIRE profile to higher halo masses using values for the baryon fraction motivated by *X*-ray observations (Davies et al. 2020).[5] Hafen et al. (2019) find a similar fraction of baryons within the virial radius for $z = 0{-}2$, and so we assume no evolution in this fraction.

Precipitation: This semi-analytic model specifies the density by assuming a threshold ratio of the cooling to dynamical time is satisfied (Sharma et al. 2012; Voit et al. 2017). Idealized simulations show that exceeding this threshold results in dramatic cooling and fragmentation, that should source star formation and, hence, stellar feedback, and restore balance (McCourt et al. 2012). These models have found success at explaining various observations of gas in the Milky Way CGM (Voit 2019) as well as in other systems (Voit 2019; Voit et al. 2019). Voit (2019) provides a fitting formula to their precipitation profile in physical units:

$$n_e(r) = \left\{ \left[n_1 \left(\frac{r}{1\ {\rm kpc}}\right)^{-\zeta_1} \right]^{-2} + \left[n_2 \left(\frac{r}{100\ {\rm kpc}}\right)^{-\zeta_2} \right]^{-2} \right\}^{-1/2} \tag{5}$$

The parameters of this model are given for a range of halo masses in Voit (2019).[6] This density scales, respectively, at small and large radii as $n_e \propto r^{-\zeta_1}$ with $\zeta_1 \approx 1.2$ and $n_e \propto r^{-\zeta_2}$ with $\zeta_2 \approx 2.3$. The latter scaling is similar to the FIRE profile, although the fraction of baryons within the virial radius in the Precipitation model scales more strongly with halo mass than in FIRE. The normalizing factors n_1 and n_2 are interpolated from Table 1 in the appendix of Voit (2019), as are ζ_1 and ζ_2. Just like with FIRE, we need to find a way to redistribution all the halo-associated gas. This density profile given by equation (5) is integrated out to $3r_{\rm vir}$. If at some radius the total associated gas mass is exceeded, this becomes the maximum radius of the profile. Otherwise, if by $3r_{\rm vir}$ some of the halo-associated gas mass is still not accounted for, this gas is then distributed in a tophat out to $3r_{\rm vir}$ (this tophat contains 98 per cent of the gas for a 10^{11}M$_\odot$ halo, 78 per cent for a 10^{12}M$_\odot$ halo, and no gas for a 10^{13}M$_\odot$ halo at $z = 0$). The generalized result is that above $5 \times 10^{12}(1+z)^{-3/2}M_\odot$ there is no mass in the tophat. In order to prevent unrealistic densities in our extrapolation to large haloes (as the tables in Voit 2019 only go to 10^{13}M$_\odot$), we linearly extrapolate above this, and at $10^{14.1}$M$_\odot$ we transition to a NFW profile as the model starts to overshoot the NFW at this mass (there are few haloes above this mass in our application to the Bolshoi simulation). This model is evolved in redshift from the $z = 0$ fits in Voit (2019) by keeping the profile fixed with constant virial temperature.[7]

2.3 Multiple redshifts and a light-travel image

Most applications of our algorithm require capturing a light-travel image of the baryon field across a span in redshift. To create such an image with a periodic simulation box captured at discrete times, we follow the standard procedure used in many cosmological analyses of stacking shifted simulation snapshots, as well as interchanging which axis is the line of sight (e.g. Saro et al. 2014). The average redshift extent of our example box of size $250h^{-1}$ Mpc is $\Delta z \approx 0.1$ at $z \lesssim 1$. So to create a map out to higher redshifts than $z = 0.1$, we must stack a number of boxes on top of each other. To do so, we use the density and halo field snapshots closest to the mean redshift desired for the next box in the stack.[8] A given snapshot can be selected multiple times if the cadence of simulation outputs is coarse.

To create a light-travel image across multiple boxes, a randomization procedure is used to diminish artifacts from the same structures aligning in the stack. For each box, the original density field is projected from 3D to 2D randomly along one of the three major axes. Our normal procedure of removing and adding haloes (Section 2.1) is then applied to get the final density field for each box. The outputs are then randomly mirrored along each axis, and translated in the

[5]The extrapolated fraction of baryons in CGM for 10^{13}, 10^{14}, 10^{15} M$_\odot$ are 0.5, 0.8, 1, respectively.

[6]Specifically, we use the Voit (2019) models with $Z = 0.3Z_\odot$ and a cooling time to free fall time threshold of 10.

[7]To extrapolate this model to higher redshifts from $z = 0$ coefficients we use the fact that for an isothermal sphere mass profile, the dark matter density at fixed physical radius scales as $(1+z)M^{2/3}$, which means that fixing the dynamical time requires keeping $(1+z)M^{2/3}$ constant. However, if $(1+z)M^{2/3}$ is fixed, since $T_{\rm vir}$ has exactly the same scaling, this fixes the virial temperature and, hence, the cooling rate. Thus, if we keep the same gas density profile as the $z = 0$ density profiles of Voit et al. (2018), but identify this profile with a halo with mass such that $(1+z)M^{2/3}$ is fixed, this essentially fixes the ratio of cooling to dynamical times, which is what we desire to do to maintain the spirit of the model.

[8]Since this stacking of a handful of boxes is a rather crude Riemann sum, to guarantee that the average over the box matches the expected DM(z) relation for our FRB application, we choose an effective redshift of each box in the stack to rescale its mean density to $\bar{n}_{e,0}(1+z_{\rm eff}^{(n)})^3$ and to assign to the n^{th} snapshot, where $\bar{n}_{e,0}$ is the mean density today, so that the dispersion averages to the mean relation $\overline{DM}(z)$. This requires $z_{\rm eff}^{(n)} = (\overline{\rm DM}(z_{\rm max}^{(n)}) - \overline{\rm DM}(z_{\rm min}^{(n)}))/(L\bar{n}_{e,0}) - 1$, where $z_{\rm min}^{(n)}$ and $z_{\rm max}^{(n)}$ are the redshifts for the edges of the nth box of comoving size L in the stack (calculated by converting the comoving size of the n and n + 1 boxes to their corresponding redshifts). This rescaling results in $z_{\rm eff}$ being just slightly different than the z of the snapshot, but is preferable to the 10 per cent errors from $\overline{DM}(z)$ that result if we do not rescale for the applications presented in this paper. Different weightings should be chosen for other observables.

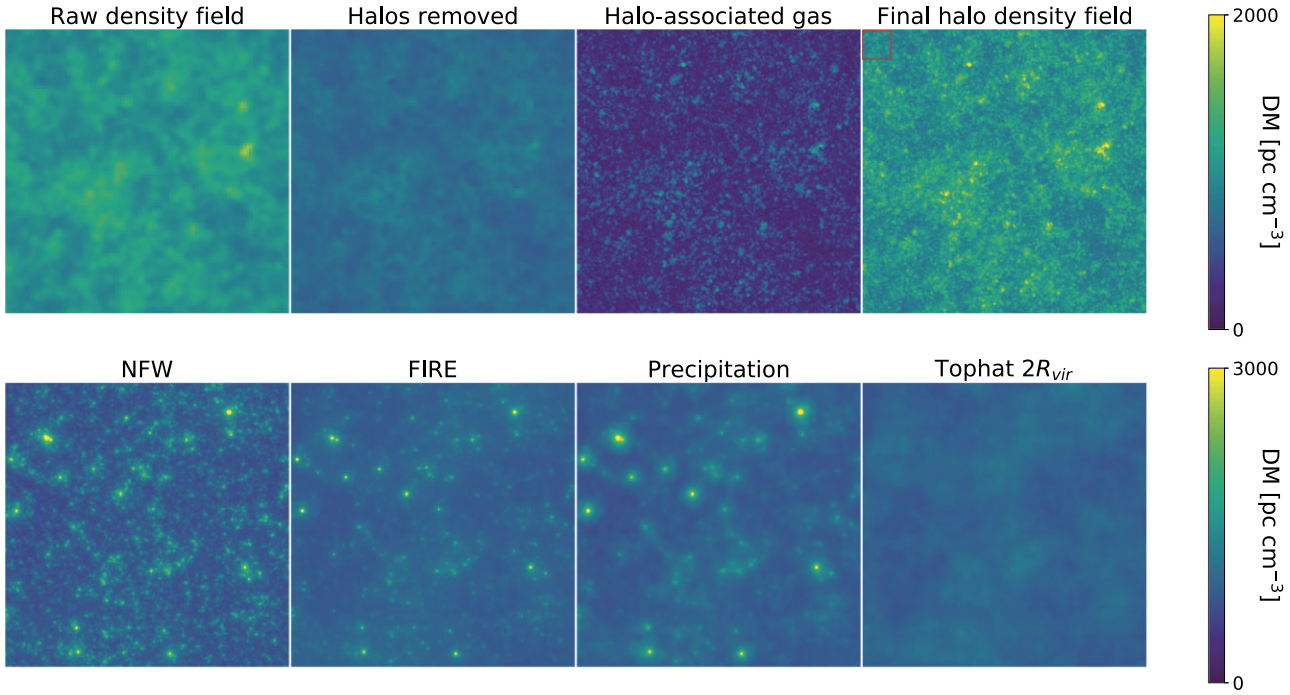

Figure 2. Illustration of the method for different baryonic profiles on a light-rectangle image to a redshift of $z = 1$, where 'light-rectangle' indicates that the projected coordinate follows the light travel delay. Nine boxes are projected over to construct the images. Top panels: Shown is the DM in a 140×140 Mpc^2 subregion through different stages of the algorithm. The left-most panel shows the projection of the raw gridded density field from the simulation. From this density field, we subtract the template halo density field, and the result is shown in the second panel. The third panel shows the halo-associated gas density field using the 1 $r_{\rm vir}$ top-hat profile that we wish to add back to the subtracted field, where a fine Eulerian grid of $N_f^2 = 8194^2$ is used. The right-most panel shows the projected final density field after the finely-gridded halo-associated gas density field is added back to the coarse subtracted field. Bottom panels: A zoom-in on the 14×14 Mpc^2 region indicated by the red rectangle in the top-right panel, but shown for other baryonic profiles. From left to right, we show the final light-rectangle field when re-adding baryons with NFW, FIRE, Precipitation, and 2 $r_{\rm vir}$ tophat profiles. These models are described in Section 2.2, and we note that the NFW and tophat models are of a toy character and not meant to be realistic.

direction transverse to the line of sight by a vector of random direction and length sampled from $[0, \sqrt{2}L]$. The translation is performed in a manner that preserves the simulation's periodicity. This stacking procedure assumes that correlations on scales larger than the box size can be neglected. The large box sizes our algorithm envisions (i.e. $250h^{-1}$ Mpc in the application we present), and the $\lesssim 1$ Mpc halo scale correlations of interest, suggest that this is a good approximation (see also Appendix B, which quantifies the contributions of fluctuations at $k \lesssim 2\pi/L$ to common variance statistics). The net result is that we have simulated the matter fields to high redshifts as shown in Fig. 2.[9] The rightmost image of the final halo density field is a 140 Mpc region, and shows that our method is able to resolve structures and distribute baryons quite effectively without introducing any artifacts from redshift stacking. The zoomed in region of 14 Mpc in the lower panel allows us to see the concentration of baryons for different profiles. As in Fig. 1, we see that the NFW profile is the most concentrated, followed by the FIRE, Precipitation, and then tophat profiles. We discuss the PDFs from these stacked fields in Section 3.1.

[9]Currently our maps are in position space instead of angular space, which is convenient for the applications in this paper. To calculate angular statistics, like the angular power spectrum of the Sunyaev Zeldovich anisotropies or weak lensing, the algorithm needs to be generalized. Angular maps are also interesting for our FRB application to understand DM correlations from more local structures such as the Local Group.

2.4 Computational properties

Our PYTHON `CGMBrush` implementation of the algorithm allows an ordinary home computer to quickly re-paint baryons at high resolution. The implementation is capable of scaling N_f to as a large a value as available memory supports. Our implementation requires ~4 Gb RAM for $N_f = 8, 192$, scaling quadratically with N_f. The runtime for a single box at this resolution is about 7 min on a ~3 GHz core, and scales roughly quadratically with N_f and linearly with the number of boxes. Multiple realizations of the baryon field can be run as separate, parallel operations, provided enough memory is available. In a supercomputing environment, many profiles can be run at very high resolutions in a single day.

3 APPLICATION OF `CGMBRUSH` TO FRBS

As an initial application of our algorithm, we explore how the baryon distribution around structures ranging from dwarf galaxies to rich clusters affect the dispersion measure (DM) of FRBs.

We apply our algorithm to data from the Bolshoi Simulation, described in Appendix A. For all results presented below, `CGMBrush` calculations were made parametrized with $n_M = 60$ logarithmic mass bins between a minimum and maximum halo mass of 1×10^{10}–$8.3 \times 10^{14}\, M_\odot$, resulting in $\Delta M \approx 0.21M$. The upgridded resolution N_f used are always specified with each result; we generally use $N_f = 32, 768$ for $z = 0$ calculations and $N_f = 8, 192$ when computing quantities out to higher redshift (see Appendix C for details on convergence).

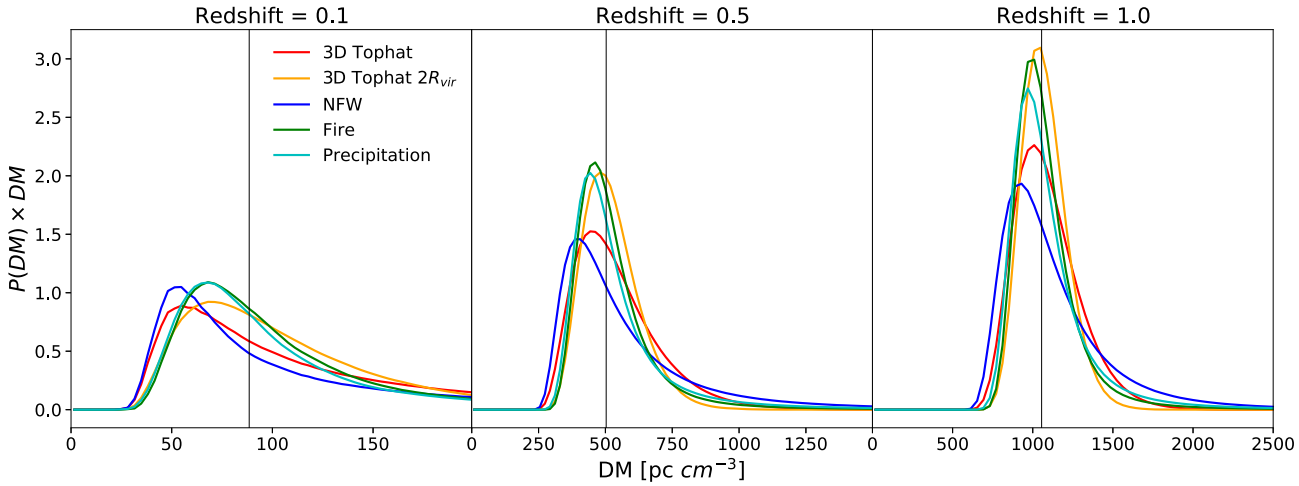

Figure 3. The PDF of DM for sightlines to the specified redshift for different baryonic profiles around haloes. The different curves illustrate the dependence on different profiles including spherical tophat, FIRE, NFW, and Precipitation (the description of these profiles is given in Section 2.2). The vertical line shows the mean DM to the given redshift. More concentrated profiles exhibit stronger skew and greater variance. With increasing redshift, the PDFs become more Gaussian as more structures contribute. These PDFs are calculated with $N_f = 32,768$ resolution, although the Appendix C demonstrates reasonable convergence on this statistic even for $N_f = 1024$.

3.1 The probability distribution of DM(z)

One of our main goals is to analyse the effect of different gas profiles on the probability distribution of DM to a given redshift. The DM is an observed quantity that measures the delay in the arrival time as a function of the frequency caused by the total electron column density along the line of sight. For FRBs at a redshift z_s:

$$\mathrm{DM}(z_s, \hat{\boldsymbol{n}}) = \int_0^{\chi(z_s)} \frac{\rho_e(z, \hat{\boldsymbol{n}})}{(1+z)^2}\, \mathrm{d}\chi, \tag{6}$$

where $\mathrm{d}\chi = c\,\mathrm{d}z/H(z)$ is the differential of the conformal distance, $\rho_e(z, \hat{n})$ is the electron number density at redshift z in direction $\hat{\boldsymbol{n}}$, and z_s is the source redshift. We calculate statistics over the interval $z = 0{-}1$. Since the length of a single box is $250h^{-1}$Mpc, we stack nine different simulation snapshots to create the $z = 1$ PDF.

Fig. 3 shows the PDF of the cosmic distribution of DM to redshifts of 0.1, 0.5, and 1, respectively, for the profiles described above. The more concentrated the profile is (with NFW as the extreme), the greater the skew and variance of the PDF. The PDF becomes more Gaussian as the profiles become more diffuse (like the $2R_{\rm vir}$ tophat) or by going to higher redshift. These trends arise because a random sightline will sample a greater number of diffuse structures and a larger sample at higher z.

Our algorithm enables us to investigate how different halo masses contribute to the PDF of DM. Fig. 4 shows a series of PDFs in which we assume an extremely-puffy $2\,r_{\rm vir}$ spherical tophat profile below the annotated mass threshold and dark-matter-tracing NFW profile above this threshold. This transition has the effect of essentially eliminating the scatter contributed to the PDF from haloes below the threshold mass, whereas the contribution of the more massive haloes is being maximized since we are assuming the NFW profile, which acts as an upper bound on the gas concentration. The shift of the PDFs as a result of increasing the mass threshold shows that all mass scales matter from $10^{13.5}$ $\mathrm{M}_\odot$ down to $10^{10.5}$ $\mathrm{M}_\odot$ haloes (the leftward shift away from the mean vertical line indicates an increase in the scatter/variance of the PDF). The effect of the most massive haloes on the PDF is significant, as observed by the shift in the curves from $10^{12.5}$ $\mathrm{M}_\odot$ to $10^{13.5}$ $\mathrm{M}_\odot$. The most massive haloes contribute to the high-DM tail. While none of the halo mass thresholds have a negligible effect, we observe a gradual convergence as the mass threshold decreases (so that the smaller haloes are transitioning to the concentrated NFW profile). Most of the width of the $10^{13.5}$ $\mathrm{M}_\odot$ case (in which most haloes are evacuated) originates from the large-scale clustering of matter (rather than the Poisson fluctuations from sightlines intersecting haloes) as there are few haloes above this mass. Appendix B shows that matter clustering on 10–100 Mpc always contributes meaningfully to the PDF and can shape it in the more diffuse baryon profile models.

3.2 DM versus impact parameter to galaxies

Ultimately, this baryon science from FRBs will not use the PDF of DM, but instead will stack foreground galaxies with different impact parameters to background FRBs to measure statistically the gas profile around haloes. Unlike the PDF of DM, stacking is unbiased by any DM intrinsic to the host system. Such a stacking measurement was considered in McQuinn (2014), but it has not been investigated using simulations. A concern is the importance of correlating material in the stack, as stacking does not just measure the distribution of gas that is associated with the fraction of mass in baryons associated with each halo of mass M. `CGMBrush` is useful for understanding the size of this effect.

Unlike for P(DM), where we needed to trace through multiple redshifts, for the mean stacked profile of DM in a mass bin, we only need to consider a single box, as material at distances greater than the 250 Mpc/h simulation box – to which our method is applied – is weakly correlated with the halo used in the stack. Material along each FRB sightline that is in the foreground or background will add to the variance of the stack, whose effect we can calculate from the P(DM) reported in the previous section.

We calculate the radial profile of haloes in our final density field for different impact parameters as follows: for each halo within a mass bin, we trim out a grid around the halo from the final density field; then, we stack the trimmed grids of each halo in a mass bin on top of each other; from this stack of haloes, we measure the density profile. Fig. 5 shows the radial profiles of DM for 10^{11} $M_\odot$, 10^{12} $M_\odot$, and 10^{13} $M_\odot$ haloes. The dashed lines show the DM of the baryonic

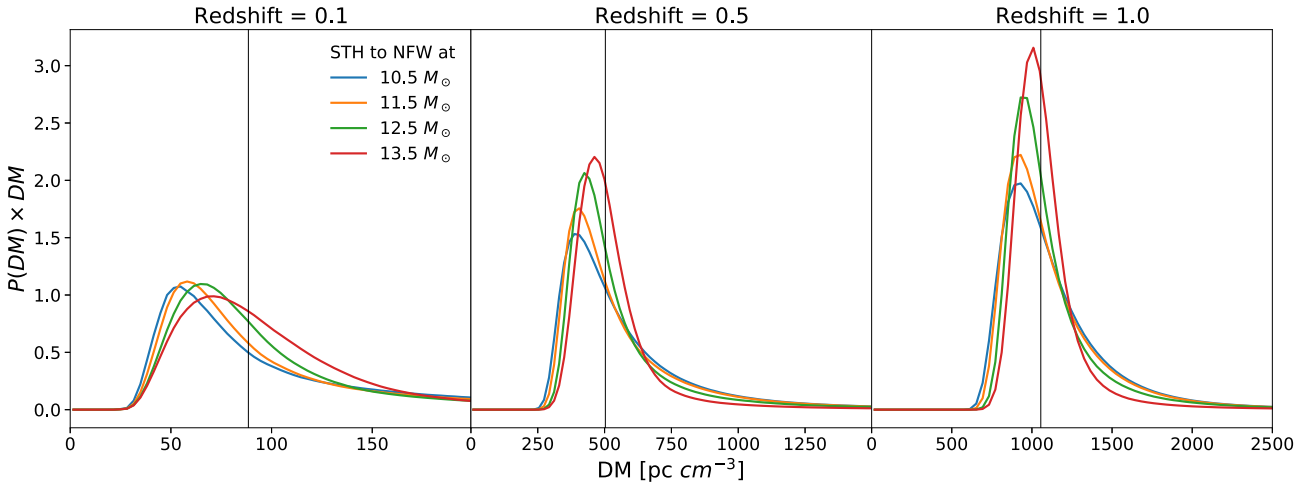

Figure 4. The PDF of DM for sightlines to the specified redshift using a toy model to understand how different haloes contribute. In this model, we transition from a $2r_{\rm vir}$ spherical tophat profile to an NFW profile above the mass thresholds listed in the legend. The former profile we consider as the maximally evacuated case (where haloes do not shape the width of the PDF) and the later NFW as the maximally concentrated. This transition has the effect of essentially eliminating the halo profile contribution to the PDF (the '1-halo' contribution). This toy model illustrates where different mass haloes likely contribute to the PDF and what mass scales need to be captured for convergence. The vertical lines show the mean DM.

profile applied to an isolated halo profile (the 1 halo term), and the solid lines show the DM of the final density field (both 1 and 2 halo terms). The vertical dashed line shows the virial radius of a halo of mass M. To the extent that there is no contribution of correlated material that lies outside of the halo, these two profiles should agree. The differences between the solid and dashed curves illustrates the contribution of all the matter correlated with the halo, the 'two-halo' term. The two-halo term is much larger than the theoretical calculations by McQuinn (2014); and generally the halo model has difficulty capturing the two halo term on such nonlinear scales. In our calculations, the two-halo term is comparable to the one-halo term at all radii in the Milky Way-mass haloes, and sets the trend outside the virial radius. For smaller haloes, the two-halo term is the dominant contribution at all radii, although the one-halo varies more quickly with impact parameter and so can still be distinguished.[10]

We next consider the sensitivity of a survey of $N_{\rm FRB}$ localized FRBs to measuring such a stack. Such a survey's $1\,\sigma$ error bar on DM(b) in a radial bin is

$$\sigma(b) = \frac{\sqrt{\sigma^2_{\rm Cosmic} + \sigma^2_{\rm Host}}}{\sqrt{N_{\rm inter}(b)}}; \quad N_{\rm inter} = N_{\rm FRBs} \times N_{\rm halos} \times P(b) \qquad (7)$$

where $N_{\rm inter}$ is the number of FRB sightlines that fall within $[b - \delta b/2, b + \delta b/2]$ for a halo of mass $\sim M$. In depends on $N_{\rm haloes}$, the average number of haloes intersected within a virial radius by each sightline to a redshift of $z \approx 0.5$ – characteristic of a typical redshift of current FRB samples. As read off from fig. 1 in McQuinn (2014), $N_{\rm haloes}$ equals approximately 4, 2, and 0.6 for haloes with mass above 10^{11}, 10^{12}, and 10^{13} $M_\odot$, respectively. However, in our calculation we take more conservative estimates for haloes intersected by each sightline to a redshift. First, McQuinn (2014) estimates are for intersecting haloes above a mass threshold, whereas we are calculating the error for a mass bin; secondly, small haloes are harder to detect. As such, for 10^{11} $M_\odot$ we reduce 4 to 1, and for 10^{12} $M_\odot$ we reduce 2 to 1. The $P(b)$ factor in $N_{\rm inter}$, gives the probability of the FRB signal passing through the gas for a given impact parameter bin at distance b from the centre of the halo and is given by $P(b) = (\pi(b + \delta b)^2 - \pi b^2)/(\pi r^2_{\rm vir})$.

This error is plotted in Fig. 6 alongside the 1 $r_{\rm vir}$ tophat and FIRE profiles. The dashed black piece-wise line shows an estimate for the sensitivity to the mean DM for a tophat and FIRE profile for a sample of $N_{\rm FRB} = 100$ and $z \approx 0.5$ localized FRBs for haloes of masses 10^{11}, 10^{12}, and 10^{13} $M_\odot$ in the top, middle, and bottom panels, respectively. These estimates use the quoted numbers for $z = 0.5$ bursts as well as a $\sigma_{\rm cosmic} = 150$, the variance of the P(DM) in the FIRE model (the was found to have very similar variance). The significance that each profile can be detected in a radial bin is assessed by the ratio of the profile to the amplitude of the piece-wise line. The intrinsic contribution from the host galaxy is set to zero ($\sigma_{\rm Host} = 0$) in the black dashed piece-wise line. For this case and with 100 FRBs, we predict a stacking analysis will be sensitive to both models from $\sim$0.1 $r_{\rm vir}$ outwards, with this detection constraining the two-halo term outside 1 $r_{\rm vir}$. Adding the $\sigma_{\rm Host} = 300$ pc cm^{-3} as the intrinsic contribution from the host galaxy, which is likely an upper bound on the likely contribution to the error from the hosts (Macquart et al. 2020), the error increases significantly as shown by the dashed blue line, but the two profiles are still constrained down to a quarter or so of the virial radius.[11]

Thus, with as few as 100 FRBs and optical follow-up to find foreground galaxies, we predict that the DM profile can be constrained over an impact parameter of $\gtrsim 0.1 r_{\rm vir}$ for halo masses of 10^{10}–10^{13} $M_\odot$, with the sensitivity in detail depending on the true gas profile. While our method can measure the gas profile outside of the virial radius with such a sample, the two-halo term from correlating gas is generally dominant there, even for our gas profiles that distribute the most gas at $\gtrsim 1\,r_{\rm vir}$ radii.

[10]It may be, especially towards lower mass haloes, that the mapping from galaxy properties to halo mass cannot be approximated as one-to-one. In this case, the interpretation of such a stacking experiment may be more complex.

[11]Down-weighting the high-DM tail in a stack will reduce the effective variance, making our estimates conservative. Since the cosmic contribution to the high-DM tail comes from sightlines through massive systems, such selection can also be done based on the proximity of sightlines to massive structures. Selection can also use scattering, rotation measure, and the location within its host galaxy to excise hosts that likely have larger intrinsic DM.

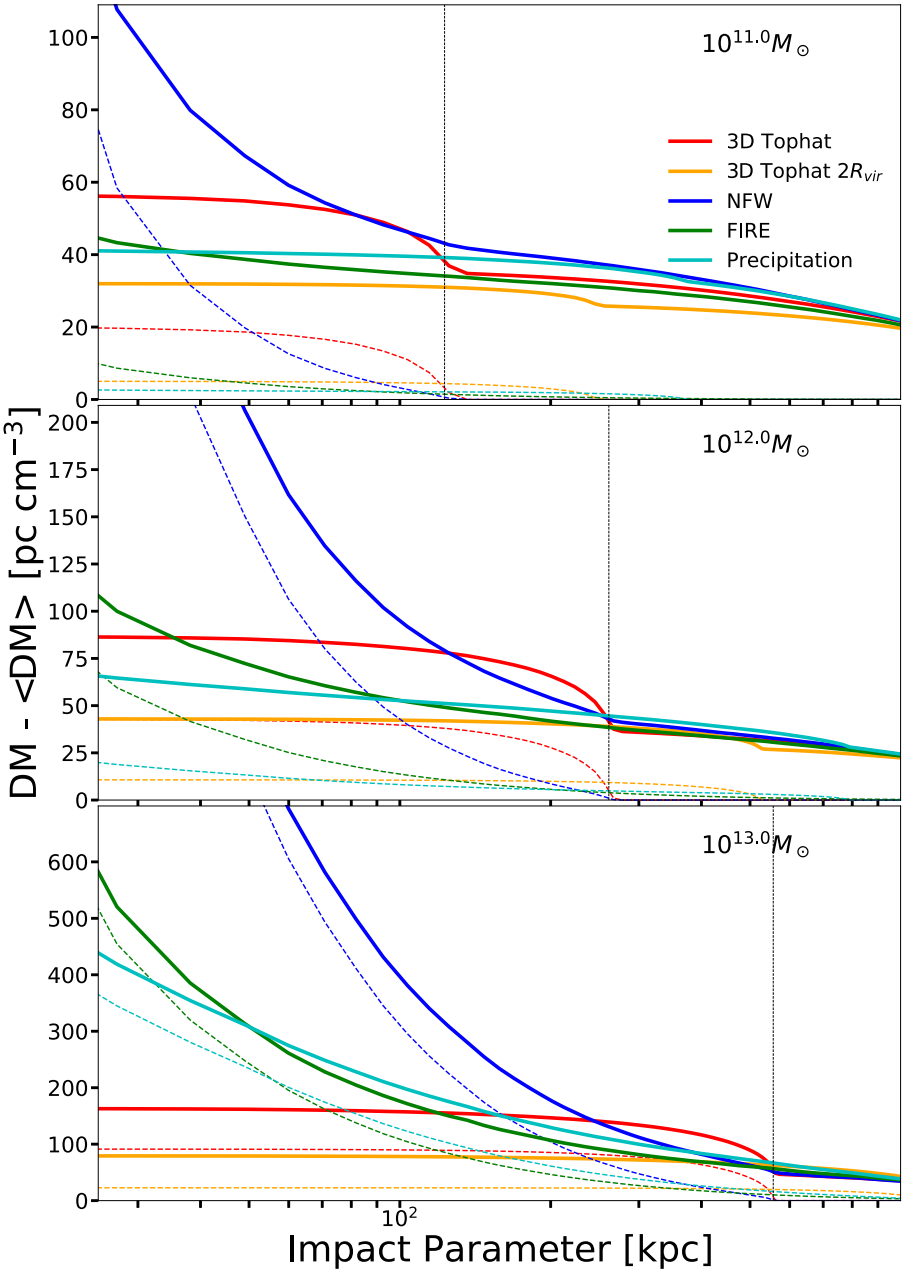

Figure 5. DM versus impact parameter: The angular profile of DM for different halo gas profiles is shown in each panel, which can be measured by stacking FRB DM measurements by their impact parameter to foreground galaxies. The middle panel is for a Milky Way-like halo of $10^{12} M_\odot$, whereas the other panels consider galactic haloes that are an order of magnitude more and less massive. The dashed lines show the DM for the baryonic profile we applied to the mass bin, and the solid lines show the DM extracted from the final density field produced by `CGMBrush`. The difference between the solid and dashed curves owes to the contribution of the gas correlated with haloes that is not contained in the halo gas profile (i.e. the 'two-halo' term). The vertical black line indicates $R_{\rm vir}$. Note that the tophat and NFW models are of a toy character, for illustrative purposes. The calculations are made at $N_f = 32,768$ resolution using the $N = 512$ resolution initial Bolshoi density grid.

4 CONCLUSIONS

Constraining the locations of the cosmic baryons is important for both understanding galaxy formation and achieving precision cosmology from large-scale structure surveys. We have developed a method to redistribute baryons around haloes in *N*-body simulations and presented its application to FRBs. The method is fast, only requiring convolutions in 2D, and it generates Eulerian 2D grids, which are easier to work with than adaptive outputs. A fixed grid, while wasteful for many applications, is more justified for the projected distribution of baryons as many haloes intersect with essentially every sightline. The method enables modelling the baryons in much larger volumes and around smaller galaxies than can be achieved with modern cosmological hydrodynamic simulations, and it can be used to quickly survey potential models.

The distribution of halo-associated baryons that are used can be informed by the results of semi-analytic models, zoom-in hydrodynamic simulations, or previous observations. We implemented several such distributions in this study. In addition to toy analytic

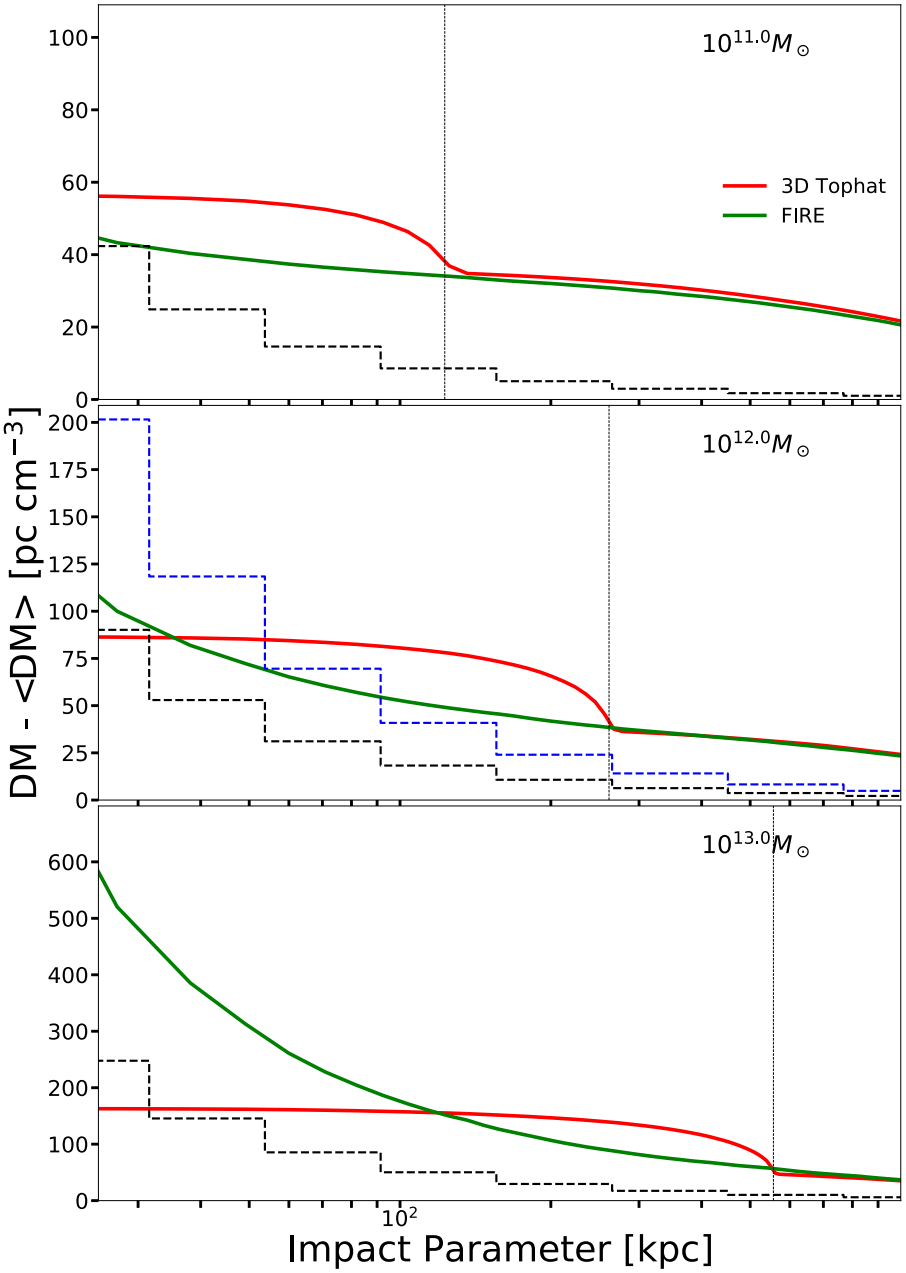

Figure 6. The angular profile of the dispersion measure for the 1 $r_{\rm vir}$ 3D tophat and FIRE gas profile (solid curves) alongside our 1σ error bar estimates for a mock stacking survey of localized FRBs (dashed piece-wise lines). This survey assumes 100 FRB localizations with a typical redshift of $z \approx 0.5$ plus models for the foreground incidence rate of different halo masses (see Section 3.1). We use a variance of $\sigma_{\rm cosmic} = 150$ for both profiles, as they were found to be very similar when computed from our $z = 0.5$ PDFs (Fig. 3.1). The blue dashed piece-wise lines in the middle panel shows how the error increases if the FRB host system contributes scatter in DM of $\sigma_{\rm Host} = 300$ pc cm^{-3}, whereas the other piece-wise lines are in the limit of the host galaxy being a subdominant contribution to the error.

models, we considered models for the baryonic profile motivated by the precipitation-limited models of Voit (2019) and the FIRE galaxy formation simulations Hafen et al. (2019). Our publicly available PYTHON package `CGMBrush` allows the user to create custom models.

By applying the `CGMBrush` method to the statistics of DM towards cosmological FRBs, we have shown that the DM is quite sensitive to the distribution of baryons in galactic haloes, with plausible models producing significantly different probability distributions of DM to a given redshift. We have also investigated the projected profile that could be measured in a analysis that stacks FRBs based on their impact parameter to foreground galaxies of different types. Our investigation included the first assessment of the effect of the contaminating matter from correlating systems, which we find complicates inferring the gas profile towards smaller mass systems. With as few as 100 FRBs that are localized to galaxies plus optical follow-up to find foreground galaxies, we predict that the DM profile can be well constrained over an impact parameter of $0.3-1 r_{\rm vir}$ for halo masses of $10^{11}-10^{13}$ M$_\odot$, with the sensitivity in detail depending on the true gas profile. The gas profile outside of the virial radius would also be measured with such a sample, although the two-halo term from correlating gas is generally dominant for

Milky Way-mass haloes (and for even smaller haloes, the two halo can be larger even at $< 1\,r_{\rm vir}$). Unlike many CGM observables that are most sensitive to group and cluster sized haloes, the limiting factor for measuring the halo gas profile towards even dwarf galaxies is having a deep enough survey to identify these galaxies; if they can be identified, we conclude that their gas profile can be measured statistically with FRBs. With the likely 'exponential' increase of FRB detections that have localizations, with CHIME, HIRAX, CHORD, and DSA forecasting thousands in the next few years (Hallinan et al. 2019; Kocz et al. 2019; Vanderlinde et al. 2019; Amiri et al. 2021; Crichton et al. 2022; Petroff et al. 2022), our estimates suggest these measurements have the potential to finally constrain the total gas profile around galaxies of different types.

Our algorithm is not just relevant to FRB DM science that this paper focuses on. With extensions that allow for magnetic fields and small-scale gas inhomogeneities, it could also be used to predict the rotation and scattering measures to FRBs. Furthermore, our technique is relevant for modeling the thermal and kinetic Sunyaev–Zeldovich effects, which also probe the gas profiles around haloes. Our tool can straightforwardly be extended to generate fast, flexible models for the Sunyaev–Zeldovich angular power spectra. Additionally, our method could be applied to understand the distribution of metals in the universe owing to galactic feedback. Indeed, a related post-processing of tophat enrichment profiles around simulated galaxies was used in Booth et al. (2012) to quantify the extent of enrichment – a technique they devised because their suite of hydrodynamic simulations could not reproduce the scope of metal enrichment. Finally, our models can be used to model the weak lensing power spectrum, where the uncertain gas distribution around haloes is a major systematic for accurately estimating cosmological parameters. Current lensing analyses mainly use the predictions of hydrodynamic simulations to account for this uncertainty (Chisari et al. 2018; Foreman et al. 2020; Osato et al. 2021), whereas `CGMBrush` models have more versatility at the expense of some self-consistency.

ACKNOWLEDGEMENTS

We thank Zachary Hafen for helping with the prescription for the FIRE-inspired profile, G. Mark Voit for help in implementing the precipitation model, and the CosmoSim Database for making the results of simulations publicly available. We acknowledge support from NSF award AST-2007012.

DATA AVAILABILITY

The `CGMBrush` code is available at https://github.com/ianw89/cgm-brush. The Bolshoi simulation data used in this work is available on CosmoSim; see https://www.cosmosim.org/cms/simulations/bolshoi/ and our GitHub instructions for more details on how to query for the data needed to reproduce these results.

REFERENCES

Amiri M. et al., 2021, ApJS, 257, 59
Amodeo S. et al., 2021, Phys. Rev. D, 103, 063514
Aricò G., Angulo R. E., Hernández-Monteagudo C., Contreras S., Zennaro M., Pellejero-Ibañez M., Rosas-Guevara Y., 2020, MNRAS, 495, 4800
Bahcall J. N., Spitzer Lyman J., 1969, ApJ, 156, L63
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Booth C. M., Schaye J., Delgado J. D., Dalla Vecchia C., 2012, MNRAS, 420, 1053
Bregman J. N., Anderson M. E., Miller M. J., Hodges-Kluck E., Dai X., Li J.-T., Li Y., Qu Z., 2018, ApJ, 862, 3
Centrella J., Melott A. L., 1983, Nature, 305, 196
Chadayammuri U., Bogdán Á., Oppenheimer B. D., Kraft R. P., Forman W. R., Jones C., 2022, ApJ, 936, L15
Chisari N. E. et al., 2018, MNRAS, 480, 3962
Connor L., Ravi V., 2022, Nature Astron., 6, 1035
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Cordes J. M., Chatterjee S., 2019, ARA&A, 57, 417
Crichton D. et al., 2022, J. Astron. Telesc. Instr. Syst., 8, 1
Davies J. J., Crain R. A., Oppenheimer B. D., Schaye J., 2020, MNRAS, 491, 4462
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Debackere S. N. B., Schaye J., Hoekstra H., 2020, MNRAS, 492, 2285
Deng W., Zhang B., 2014, ApJ, 783, L35
Faerman Y., Sternberg A., McKee C. F., 2017, ApJ, 835, 52
Faucher-Giguère C.-A., Feldmann R., Quataert E., Kereš D., Hopkins P. F., Murray N., 2016, MNRAS, 461, L32
Fielding D., Quataert E., McCourt M., Thompson T. A., 2017, MNRAS, 466, 3810
Foreman S., Coulton W., Villaescusa-Navarro F., Barreira A., 2020, MNRAS, 498, 2887
Giri S. K., Schneider A., 2021, J. Cosmol. Astropart. Phys., 2021, 046
Greco J. P., Hill J. C., Spergel D. N., Battaglia N., 2015, ApJ, 808, 151
Hafen Z. et al., 2019, MNRAS, 488, 1248
Hallinan G. et al., 2019, BAAS, 51, 255
Henley D. B., Shelton R. L., 2012, ApJS, 202, 14
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Kocz J. et al., 2019, MNRAS, 489, 919
Kravtsov A. V., 2013, ApJ, 764, L31
Lee K.-G., Ata M., Khrykin I. S., Huang Y., Prochaska J. X., Cooke J., Zhang J., Batten A., 2022, ApJ, 928, 9
Lee M. E., Lu T., Haiman Z., Liu J., Osato K., 2023, MNRAS, 519, 573
Li M., Tonnesen S., 2020, ApJ, 898, 148
Lochhaas C., Bryan G. L., Li Y., Li M., Fielding D., 2020, MNRAS, 493, 1461
Lorimer D. R., Bailes M., McLaughlin M. A., Narkevic D. J., Crawford F., 2007, Science, 318, 777
Lu T., Haiman Z., 2021, MNRAS, 506, 3406
Macquart J. P. et al., 2020, Nature, 581, 391
Maller A. H., Bullock J. S., 2004, MNRAS, 355, 694
McCourt M., Sharma P., Quataert E., Parrish I. J., 2012, MNRAS, 419, 3319
McQuinn M., 2014, ApJ, 780, L33
McQuinn M., 2016, ARA&A, 54, 313
Miller M. J., Bregman J. N., 2013, ApJ, 770, 118
Miller M. J., Bregman J. N., 2015, ApJ, 800, 14
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Osato K., Nagai D., 2023, MNRAS, 519, 2069
Osato K., Liu J., Haiman Z., 2021, MNRAS, 502, 5593
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Petroff E., Hessels J. W. T., Lorimer D. R., 2019, A&A Rev., 27, 4
Petroff E., Hessels J. W. T., Lorimer D. R., 2022, A&A Rev., 30, 2
Planck Collaboration, 2013, A&A, 557, A52
Saro A. et al., 2014, MNRAS, 440, 2610
Schaan E. et al., 2016, Phys. Rev. D, 93, 082002
Schneider A., Teyssier R., 2015, J. Cosmol. Astropart. Phys., 2015, 049
Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019, J. Cosmol. Astropart. Phys., 2019, 020
Schneider A., Giri S. K., Amodeo S., Refregier A., 2022, MNRAS, 514, 3802
Sharma P., McCourt M., Quataert E., Parrish I. J., 2012, MNRAS, 420, 3174
Shull J. M., Danforth C. W., 2018, ApJ, 852, L11
Simha S. et al., 2021, ApJ, 921, 134
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Stern J., Fielding D., Faucher-Giguère C.-A., Quataert E., 2019, MNRAS, 488, 2549
Su K.-Y. et al., 2020, MNRAS, 491, 1190
Sunyaev R. A., Zeldovich Y. B., 1970, Ap&SS, 7, 3

Sunyaev R. A., Zeldovich Y. B., 1972, Comments Astrophys. Space Phys., 4, 173
Thom C. et al., 2012, ApJ, 758, L41
Thornton D. et al., 2013, Science, 341, 53
Tumlinson J. et al., 2011, Science, 334, 948
Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389
Vanderlinde K. et al., 2019, Zenodo. Available at:
Voit G. M., 2019, ApJ, 880, 139
Voit G. M., Meece G., Li Y., O'Shea B. W., Bryan G. L., Donahue M., 2017, ApJ, 845, 80
Voit G. M., Ma C. P., Greene J., Goulding A., Pandya V., Donahue M., Sun M., 2018, ApJ, 853, 78
Voit G. M., Donahue M., Zahedy F., Chen H.-W., Werk J., Bryan G. L., O'Shea B. W., 2019, ApJ, 879, L1
Wang H., Mo H. J., Jing Y. P., Guo Y., Van Den Bosch F. C., Yang X., 2009, MNRAS, 394, 398
Werk J. K. et al., 2014, ApJ, 792, 8
Zhu W., Feng L.-L., Zhang F., 2018, ApJ, 865, 147

APPENDIX A: BOLSHOI SIMULATION SPECIFICATIONS

Throughout this work, both as a demonstration of the algorithm and for our FRB application, we use the Bolshoi simulation (Klypin et al. 2011), publicly available at *CosmoSim*. This *N*-body simulation meets our basic requirements: it resolves sub-Milky Way haloes (to masses as small as $\sim 10^{10}$ $M_\odot$), which we require for convergence on our statistics (Appendix C), and its boxsize $L = 250h^{-1}$Mpc is greater than our 100 Mpc requirement, large enough for a representative sample of structures. There are 2048^3 particles with mass 1.35×10^8 $M_\odot$ h^{-1}, and the simulation is run in ΛCDM cosmology (Klypin et al. 2011) with parameters $h = 0.7$, $\Omega_m = 0.27$, $\Omega_b = 0.0469$, $n = 0.95$, and $\sigma_8 = 0.82$. The outputs are available for a large number of snapshots between a redshift of 0 and 17.

CosmoSim provides the Bolshoi simulation density field in two resolutions, 256^3 and 512^3. These 3D grid are calculated from the simulation snapshots using cloud-in-cell interpolation and, then, a Gaussian convolution with a standard deviation of one cell. This smoothing must be matched in subtracting haloes, as described in Section 2. Since the higher resolution of 512^3 is only provided at $z = 0$, we use the 256^3 grid for whenever $z > 0$ is required, such as in Section 3.1, and 512^2 for $z = 0$ calculations, such as the radial profiles in Fig. 1 and Section 3.2. We compare the $z = 0$ results using the two different base grid resolutions in Appendix C, and show that there are relatively minor differences in our final results.

APPENDIX B: THE SCALES THAT CONTRIBUTE TO THE VARIANCE OF THE BARYONIC FIELD

In this appendix, we discuss what structures contribute to the variance of the cosmological baryon field. We show that the variance of the dispersion measure mainly comes from gas in galactic dark matter haloes and, to a lesser extent, gas that traces ≈100 Mpc cosmic structures. Structures that are somewhat less diffuse than haloes are less important to model, which is fortunate because the method detailed in this paper is the least consistent in modeling them.

The left panel in Fig. B1 shows $k^2 P_e/(2\pi)^2$ at $z = 0.5$ for models where different haloes retain their gas with the same NFW halo profile of the dark matter above the specified halo mass and that

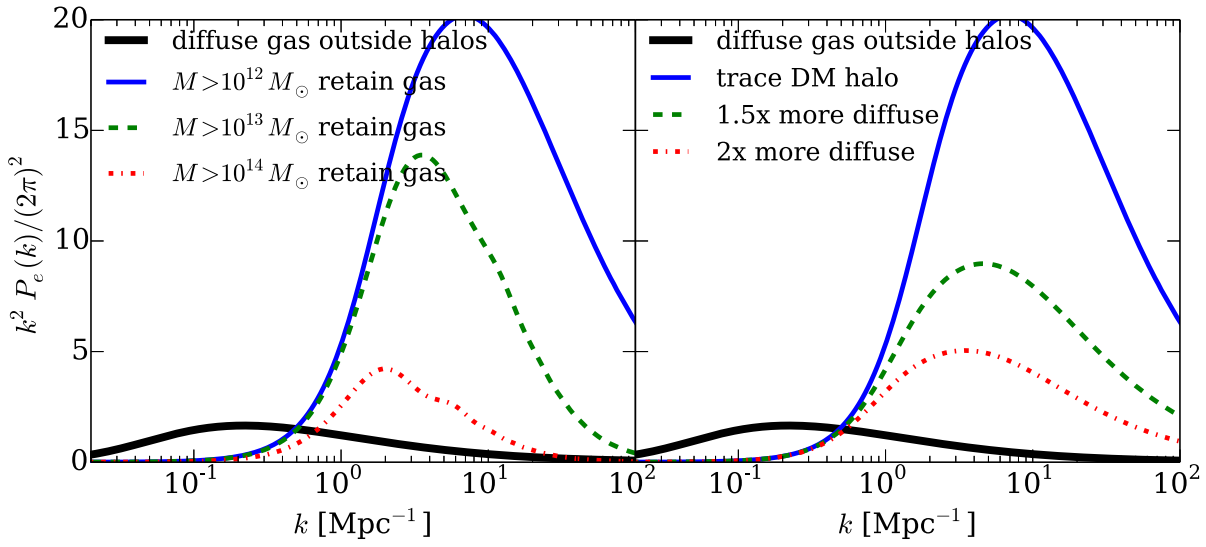

Figure B1. Left-hand panel: Halo model calculation for the $z = 0.5$ contribution of the nonlinear power spectrum of baryons assuming a toy model where they trace the NFW halo profile above the specified halo mass and where haloes are largely evacuated below this mass. The coloured curves show the one-halo term that depends on the profile, and the black curves show the two-halo term (which traces the large-scale matter field). The area under these curves is proportional to the variance of DM (equation B1). Right-hand panel: Considering only the model where $M > 10^{12}$ $M_\odot$ haloes retain their gas, but where the distribution of the gas is assumed to have an NFW profile that extends 1.5 or 2 times more in radius than the dark matter halo. This illustrates how the more diffuse the gas profile, the smaller the contribution of this one-halo term.

haloes are largely evacuated below this mass. These calculations use the halo model (e.g. Cooray & Sheth 2002).[12] Here, k is the wavenumber and P_e the electron overdensity power spectrum with the standard Fourier convention in cosmology such that the area under the curves is proportional to the variance of DM (McQuinn 2014):

$$\sigma^2_{\rm DM} \approx \int_0^z \frac{cdz}{H(z)}(1+z)^2\bar{n}^2_{e,0}\int d\log k\left[\frac{k^2P_e(k,z)}{(2\pi)^2}\right], \tag{B1}$$

where $\bar{n}_{e,0}$ is the cosmic mean electron number density today. The left-hand panel shows that much of the variance potentially owes to the profiles in galactic mass haloes. A lesser contribution to the variance is from large-scale structure shown by the thick black curve. (This thick black curve is the halo model's 'two-halo term'.) Much of the area under this curve is at $2\pi/100\ {\rm Mpc}^{-1} < k < 2\pi/10\ {\rm Mpc}^{-1}$.

Large-scale structure becomes more important as the gas around haloes becomes more diffuse. The right panel in Fig. B1 considers the previous model where haloes with mass satisfying $M > 10^{12}\ {\rm M}_\odot$ retain their gas, but where an illustrative model for the distribution of the gas is assumed where it has an NFW profile that is dilated by a factor of 1, 1.5 and 2 compared to the dark matter. This illustrates how the more diffuse the gas, the smaller contribution to the variance from sightlines intersecting individual haloes, and the relatively larger contribution of the large-scale structure term.

In both panels, the large-scale structure term is calculated assuming the linear theory density power spectrum, as is traditional in the halo model. The linear power spectrum is inaccurate of course on nonlinear wavenumbers of $k \sim 2\pi/(1-10)\ {\rm Mpc}^{-1}$, but fortunately these scales are not a significant contribution to the variance. That 'two-halo correlations' are a somewhat small contribution to the power at $k \sim 2\pi/(1-10)\ {\rm Mpc}^{-1}$ also holds for other baryonic clustering statistics and not just those most relevant for FRBs. Indeed, the unimportance of few Mpc-scale structures is the reason why the halo model (which fails on these physical scales) is successful at predicting the nonlinear power spectrum at all wavenumbers.

On $\sim$1 Mpc scales the baryons that do not accrete on to dark matter haloes may still be affected by feedback, which our `CGMBrush` algorithm is not modeling (since it only considers the gas mass that is associated with dark matter haloes). While this may be a major deficiency of the algorithm, our justification is that regions somewhat outside of haloes are less important for capturing many statistical measures of the cosmic baryon field.

APPENDIX C: RESOLUTION TESTS

We first consider how sensitive our results are to the density grid resolution of N-body simulation. Fig. C1 shows the steps in our calculations for two different Eulerian grids for the N-body outputs. The top panel is for $N = 256$ and the bottom panel for $N = 512$. Visually, we can see that the differences are modest in the final algorithmic output in the rightmost panels. More quantitatively, we test for convergence between the two base grid resolutions by creating a probability distribution function of the gas column density 'DM' before and after our processing for the 1 $r_{\rm vir}$ tophat profile for both base resolutions in the bottom of Fig. C1. At higher base resolutions, the variance is greater and the distribution more skewed – manifesting mostly as a shift by ~ 5 pc cm^{-3} at low DM owing to more voidy sightlines in the high resolution grid. When going to higher redshift (stacking more boxes), the greater statistical sampling reduces the differences between the base resolutions. In Fig. C2, we examine the impact on the radial profiles discussed in Section 3.2. While only the $10^{12}{\rm M}_\odot$ bin is shown in Fig. C2, across 10^{11-13} M$_\odot$ haloes, we find the differences are $\lesssim 20$ per cent, with convergence by 2 Mpc. The differences owe to the two halo term being smoothed out on a scale of a cell, resulting in the 256 falling below the 512. We take these collective results to mean that our algorithm is reasonably converged, being only somewhat sensitive to the base gridding. The calculations in the main text use $N = 256$ for high z calculations and $N = 512$ for $z = 0$ calculations.

We also consider how sensitive our results are to the resolution of our up-sampled fine grid that we create to add haloes back to the grid. In the top panel of Fig. C3, where we have added the haloes back by using a tophat profile (described in more detail in Section 2.2), images of the final `CGMBrush` outputs are shown where as we increase the resolution to $\eta = 32$ in the third image, which corresponds to resolution of $N_f = 8192$, not only do we resolve the internal structure of larger haloes of size $r_{\rm vir} \approx 1$ Mpc significantly, but also of the much smaller haloes that were previously not well resolved as seen the left-most image in the panel with $\eta = 4$. Despite these visual differences in the images, the effect of increasing η on the DM PDF is very small (see the bottom panels in Fig. C3). All calculations in the main text use a resolution of 8192 or higher.

[12]Our calculation uses the same code as McQuinn (2014), except here the Press-Schechter rather than Sheth-Tormen mass function is used. Press-Schechter somewhat overpredicts the abundance of rare massive haloes and underpredicts less rare ones.

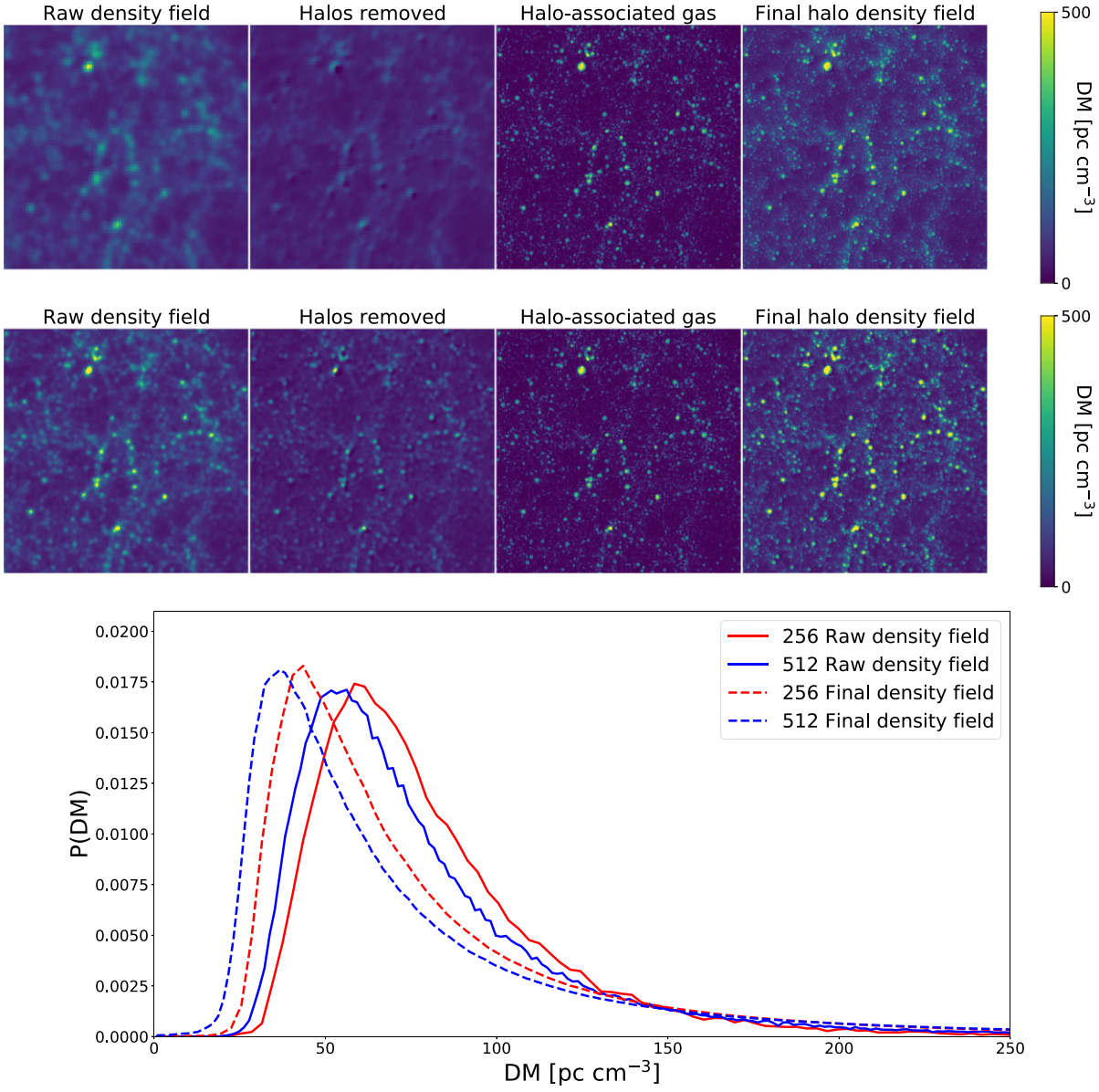

Figure C1. Illustration of method on two Eulerian base grid resolutions: the fiducial $N^2 = 256^2$ grid (top) and the higher resolution $N^2 = 512^2$ grid (middle). The images show the halo subtraction and addition through different stages of our method, applied to the $z = 0$ snapshot of the Bolshoi N-body simulation. Each panel is projected over the simulation boxsize of $250h^{-1}$ Mpc and zooms in on a subregion with transverse comoving size of 140 Mpc. The left-hand panels show the raw gridded density field from the simulation. From this density field, we subtract the halo density field and the result is shown in the second column. The third column shows the halo density field that we add back to the subtracted field, where a fine Eulerian grid of $N_f^2 = 32,768^2$ is used. The last panel shows the net field after the finely gridded haloes are added back to the coarse subtracted field. In this case, we use our simple 1 $r_{\rm vir}$ top-hat model, as the differences are smaller for more concentrated models. The bottom panel is a probability distribution of the gas column density, 'DM', across the $250h^{-1}$ Mpc box. Higher resolution, both in the base grid and from running the algorithm, results in a more skewed distribution and greater variance.

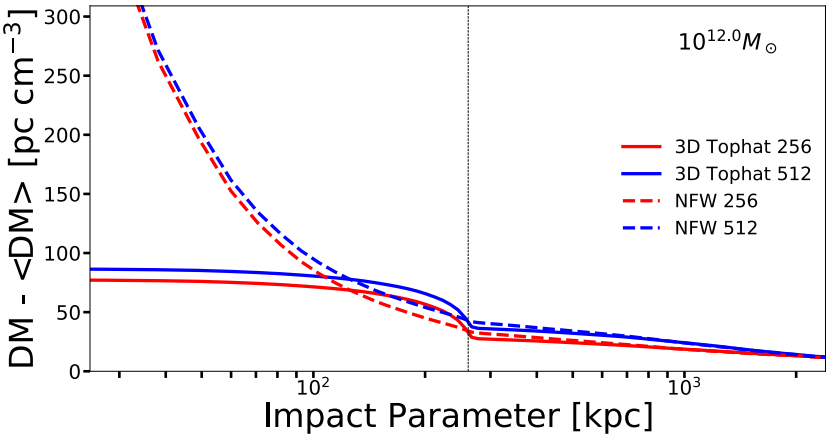

Figure C2. Comparison of the effect of using two different base grid resolutions to grid the simulation density field on the extracted radial profiles. The red lines use the fiducial $N^2 = 256^2$ grid of the Bolshoi simulation, and the blue lines use the higher resolution $N^2 = 512^2$ grid. The boost in DM from using the finer base grid is modest and owes to the two halo term being smoothed out on a scale of a cell. The radial profile calculations use the $N^2 = 512^2$ grid in the main text. The upgridded resolution used for this comparison was $N_f^2 = 32,768^2$.

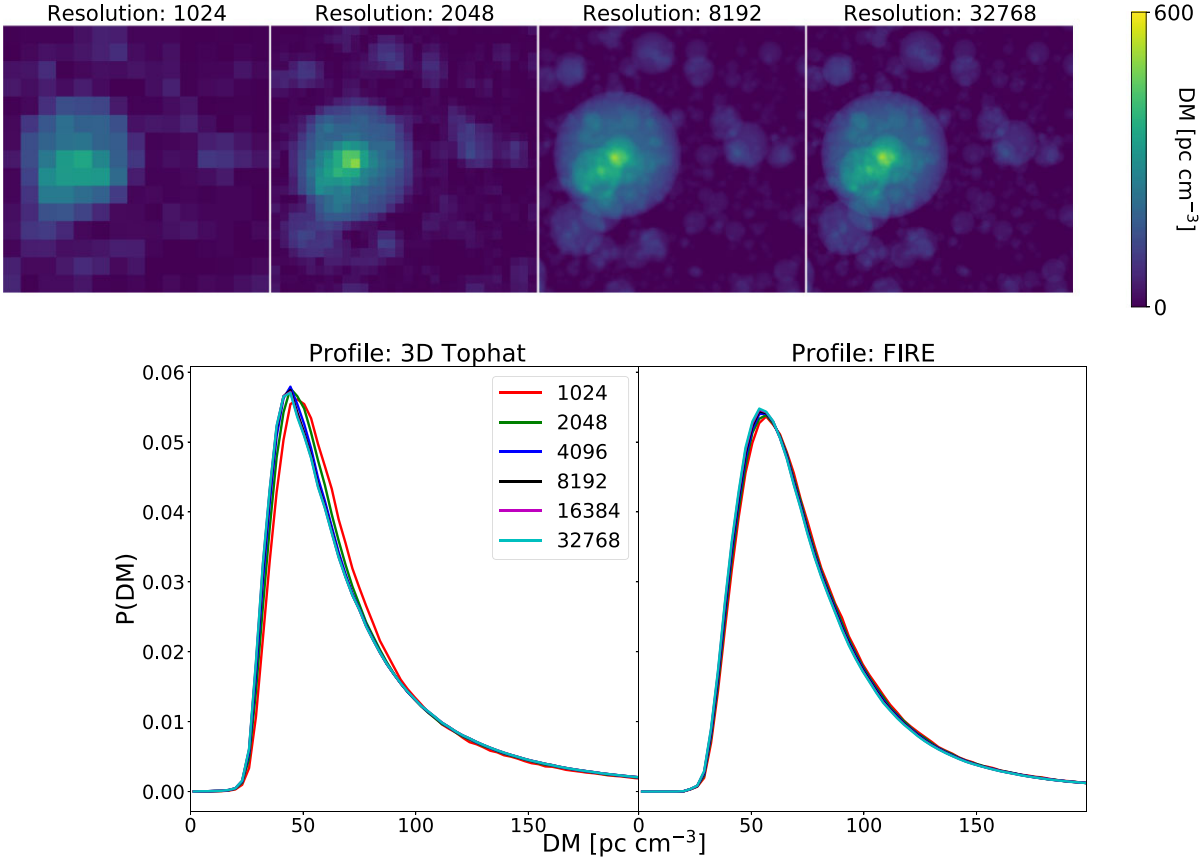

Figure C3. The images at the top show how the final projected density field converges when increasing the resolution of halo addition grid. Each image is computed by projecting the $z = 0$ Bolshoi simulation box over $250h^{-1}$ Mpc and has a transverse comoving size of $5h^{-1}$ Mpc. The bottom panel shows how DM PDFs for the 1 $r_{\rm vir}$ tophat (left) and FIRE (right) profiles tend towards convergence as we increase the resolution. Despite visual differences in the levels of structure resolved in the top panel, the PDF is not strongly dependent on the halo addition grid resolution.

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