Characterizing fast radio bursts through statistical cross-correlations

Characterizing fast radio bursts through statistical cross-correlations

Masoud Rafiei-Ravandi,[1] Kendrick M. Smith,[1] and Kiyoshi W. Masui[2,3]
[1]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[2]*MIT Kavli Institute for Astrophysics and Space Research, Cambridge, Massachusetts 02139, USA*
[3]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

(Received 14 January 2020; accepted 26 June 2020; published 16 July 2020)

Understanding the origin of fast radio bursts (FRBs) is a central unsolved problem in astrophysics that is severely hampered by their poorly determined distance scale. Determining the redshift distribution of FRBs appears to require arcsecond angular resolution, in order to associate FRBs with host galaxies. In this paper, we forecast prospects for determining the redshift distribution without host galaxy associations, by cross-correlating FRBs with a galaxy catalog such as the SDSS photometric sample. The forecasts are extremely promising: a survey such as CHIME/FRB that measures catalogs of $\sim 10^3$ FRBs with few-arcminute angular resolution can place strong constraints on the FRB redshift distribution, by measuring the cross-correlation as a function of galaxy redshift z and FRB dispersion measure D. In addition, propagation effects from free electron inhomogeneities modulate the observed FRB number density, either by shifting FRBs between dispersion measure (DM) bins or through DM-dependent selection effects. We show that these propagation effects, coupled with the spatial clustering between galaxies and free electrons, can produce FRB-galaxy correlations that are comparable to the intrinsic clustering signal. Such effects can be disentangled based on their angular and (z, D) dependence, providing an opportunity to study not only FRBs but also the clustering of free electrons.

DOI: 10.1103/PhysRevD.102.023528

I. INTRODUCTION

Fast radio bursts (FRBs) are an astrophysical transient whose origin is not yet understood. Since initial discovery in 2007 [1], interest in FRBs has grown, and explaining the FRB phenomenon is now a central unsolved problem in astrophysics (see [2–4] for recent reviews).

An FRB is a short (usually 1–10 ms), bright ($\sim$1 Jy) radio pulse which is highly dispersed: the arrival time at radiofrequency ν is delayed, by an amount proportional to ν^{-2}. This dispersion relation arises naturally if the pulse propagates through a cold plasma of free electrons. In this case, the delay is proportional to the "dispersion measure" (DM), which is defined as the electron column density along the line of sight:

$$(\text{Delay}) = (\text{DM})\left(\frac{e^2}{2\pi m_e c}\right)\nu^{-2} \tag{1}$$

$$= (4.15\ \text{ms})\left(\frac{\text{DM}}{1\ \text{pc cm}^{-3}}\right)\left(\frac{\nu}{1\ \text{GHz}}\right)^{-2}, \tag{2}$$

where

$$\text{DM} \equiv \int n_e(x)dx. \tag{3}$$

FRBs are a population of dispersed pulses whose observed DM significantly exceeds the maximum Galactic column density DM_{gal} (inferred from a model of the Galaxy [5,6]). On most of the sky, DM_{gal} is $\lesssim 50$ pc cm^{-3}, and FRBs are regularly observed with DM $\gtrsim 1000$. From the outset, the large DM suggested that FRBs were extragalactic, although on its own the large DM could also be explained by a Galactic event with a large local free electron density. As more FRBs were observed, their sky distribution was found to be isotropic (i.e., not correlated with the Galactic plane), conclusively establishing an extragalactic origin.

At the time of this writing, 92 FRB discoveries have been published (according to FRBCAT [7], `frbcat.org`). Ten of these FRBs are "repeaters," meaning that multiple pulses have been observed from the same source [8–11]. Nine of the repeaters were discovered by the CHIME/FRB instrument, and a much larger sample of nonrepeating FRBs from CHIME/FRB is expected soon. (The authors are members of the CHIME/FRB Collaboration, and forecasting the scientific reach of CHIME/FRB was the main motivation for this paper.)

Determining the redshift distribution of FRBs is critical to understanding the FRB phenomenon since a distance scale is required to determine the burst energetics and volumetric rate. In the next few paragraphs, we summarize the current observational status.

FRBs do not have spectral lines, so FRB redshifts cannot be directly determined. When an FRB is observed, an upper bound on its redshift z can be inferred from its DM as follows. We write the total DM of an FRB as the sum of contributions from our galaxy, the intergalactic medium (IGM), and the host galaxy:

$$\mathrm{DM} = \mathrm{DM}_{\mathrm{gal}} + D_i(z) + D_h, \tag{4}$$

where the IGM contribution is related to the FRB redshift as

$$D_i(z) = n_{e,0} \int_0^z dz' \frac{1+z'}{H(z')}, \tag{5}$$

where $n_{e,0}$ is the comoving electron number density and $H(z)$ is the Hubble expansion rate. If we assume that $\mathrm{DM}_{\mathrm{gal}}$ is known precisely and subtracted, then the inequality $D_h \geq 0$ implies an upper bound on z. A $\mathrm{DM} = 1000$ FRB must satisfy $z \lesssim 0.95$, and a $\mathrm{DM} = 3000$ FRB satisfies $z \lesssim 3.08$. However, an alternative hypothesis is that FRBs are at much lower redshifts, and have large host DMs.

Three FRBs have been observed in long-baseline interferometers with sufficient angular resolution to uniquely identify a host galaxy, and thereby determine a redshift [12–16]. The inferred redshifts are $z = 0.19$, 0.32, and 0.66. These observations suggest that most of the DM is IGM-related, but with only three data points it cannot be concluded that this is true for the entire population.

Host galaxy associations are a powerful way to determine FRB redshifts, but require angular resolution around 1 arcsecond or better [17]. Unfortunately, most telescopes capable of finding large numbers of FRBs have angular resolution much worse than this. In particular, for most of the CHIME/FRB sources, the angular resolution is either $\approx 1'$ or $\approx 10'$, depending on whether baseband data are available for the event [11,18,19].

In this paper, we study the following question. Given a catalog of FRBs whose resolution is insufficient for host galaxy associations on a per-object basis, is it possible to associate FRBs and galaxies on a *statistical* basis? To make this question precise, we model the angular cross power spectrum C_ℓ^{fg} between the FRB and galaxy catalogs and forecast its signal-to-noise ratio (SNR). The SNR turns out to be surprisingly large. For example, given a catalog of 1000 FRBs with $1'$ resolution, and the photometric galaxy catalog from SDSS-DR8 [20], we find an SNR of 25–100, depending on the FRB redshift distribution.

As a consequence of this high SNR, the cross-correlation is still detectable if the FRB and galaxy catalogs are binned in various ways. By dividing the galaxy catalog into redshift bins, and separately cross-correlating each bin with the FRB catalog, the FRB redshift distribution can be constrained. By additionally dividing the FRB catalog into DM bins, the FRB redshift distribution of each DM bin can be constrained, pinning down the redshift-DM correspondence.

Other binning schemes are possible. For example, the FRB catalog can be binned in observed flux, so that the galaxy cross-correlation pins down the redshift-flux correspondence, and therefore the intrinsic luminosity distribution of FRBs. Or the galaxy catalog can be binned by the star formation rate before cross-correlating with FRBs, to determine whether FRBs are associated with star formation. This technique can be applied easily to other tracer fields such as supernovae and quasars.

This paper overlaps significantly with work in the galaxy clustering literature on "clustering redshifts" [21–25]. This term refers to the use of clustering statistics to determine the redshift distribution of a source population, by cross-correlating with a galaxy catalog.

However, in the case of FRBs, we find a significant new ingredient: large propagation effects, which arise because galaxies are spatially correlated with free electrons, which in turn can affect the observed density of FRBs and its DM dependence. Propagation effects produce additional contributions to the FRB-galaxy angular correlation, which need to be modeled and disentangled from the cosmological contribution. In particular, if a galaxy catalog and an FRB catalog are correlated, this does not imply that they overlap in redshift. Propagation effects can produce a correlation between low-redshift galaxies and high-redshift FRBs (but not vice versa).

Propagation effects arise from several distinct physical mechanisms: dispersion, scattering, and plasma lensing. In this paper, we will analyze the dispersion case in detail, leaving the other cases to future work. The propagation effects that we will explore have some similarity with magnification bias in galaxy surveys (see, e.g., [26] and references therein).

We also clarify which properties of the FRB population are observable via cross-correlations. It is well known that on large scales ("2-halo dominated" scales), the only observable is $(b_f dn_f/dz)$: the product of FRB redshift distribution dn_f/dz and the large-scale clustering bias $b_f(z)$. We find that there is an analogous observable $(\gamma_f dn_f/dz)$ which determines the FRB-galaxy correlation on smaller ("1-halo dominated") scales. The quantity $\gamma_f(z)$ measures the degree of similarity between the dark matter halos which contain FRBs and galaxies, and is defined and discussed in Sec. IV.

This paper is complementary to previous works that have considered different FRB-related clustering statistics. In [27], the three-dimensional (3D) clustering statistics of the FRB field were studied, using the DM as a radial coordinate. This is analogous to the way photometric galaxy surveys are analyzed in cosmology. Here we generalize to the cross-correlation between the FRB field and a galaxy survey. The FRB-galaxy cross-correlation has higher SNR than the FRB autocorrelation, since the number

of galaxies is much larger than the number of FRBs. Whereas [27] was entirely perturbative, we perform both perturbative calculations and nonlinear simulations using a halo model. In addition, we consider two propagation effects: DM shifting and completeness (to be defined below), whereas [27] considered only the former.

Another idea that has been considered is to cross-correlate a two-dimensional (2D) map of FRB-derived dispersion measures with galaxy catalogs, to probe the distribution of electrons in dark matter halos [28–32]. The cross-correlation of DM vs galaxy density is related to the DM moment of the statistic $C_\ell^{fg}(z, D)$ considered here. Therefore, our statistic contains a superset of the information in the statistic considered in these works.

In [33], a cross-correlation was observed between 2MPZ galaxies at $z \sim 0.01$, and a sample of 23 FRBs from ASKAP operating in "fly-eye" mode with 10′–60′ angular resolution [34,35]. This measurement is seemingly at odds with the three FRB host galaxy redshifts that imply a much more distant population. In the very near future, FRB catalogs will be available with much higher number density and better angular resolution, so it will be possible to measure the cross-correlation at higher SNR and push the measurement to higher redshift. The machinery in this paper will be essential for interpreting a high-SNR cross correlation and separating the clustering signal from propagation effects.

This paper is organized as follows. In Sec. II, we define notation and our modeling assumptions. In Sec. III, we define our primary observable, the FRB-galaxy cross power spectrum C_ℓ^{fg}. We explore and interpret clustering contributions to C_ℓ^{fg} in Sec. IV, and propagation effects in Sec. V. We present signal-to-noise forecasts in Sec. VI, and in Sec. VII we describe a Monte Carlo simulation pipeline that we use to validate our forecasts. We conclude in Sec. VIII.

II. PRELIMINARIES

A. Definitions and notation

Throughout the paper, we use the flat-sky approximation, in which an angular sky location is represented by a two-component vector $\boldsymbol{\theta} = (\theta_x, \theta_y)$, and we assume periodic boundary conditions with no angular mask for simplicity. Angular wave numbers (continuous quantities in a plane) are denoted $\boldsymbol{\ell} = (\ell_x, \ell_y)$, and 3D comoving wave numbers are denoted $\mathbf{k}$. We denote the observed sky area in steradians by Ω.

Let $H(z)$ be the Hubble expansion rate at redshift z, and let $\chi(z)$ be the comoving distance to redshift z:

$$\chi(z) = \int_0^z \frac{dz'}{H(z')}. \tag{6}$$

Let $P_{\rm lin}(k, z)$ denote the linear matter power spectrum at comoving wave number k and redshift z.

We use f and g to denote an FRB or galaxy catalog. Depending on context, the FRB catalog may be binned in DM, or the galaxy catalog may be binned in redshift. For $X \in \{f, g\}$, let n_X^{2d}, $n_X^{3d}(z)$, and dn_X^{2d}/dz denote the 2D number density, 3D number density, and 2D number density per unit redshift. These densities are related to each other by

$$n_X^{3d}(z) = \frac{H(z)}{\chi(z)^2}\frac{dn_X^{2d}}{dz}, \qquad n_X^{2d} = \int dz \frac{dn_X^{2d}}{dz}. \tag{7}$$

B. Halo model

We model FRB and galaxy clustering using the halo model. For a review of the halo model, see [36]. In this section, we give a high-level summary of our halo modeling formalism. For details, see Appendix A.

In the halo model, FRB and galaxy catalogs are simulated by a three-step process. First, we simulate a random realization of the *linear* cosmological density field $\delta_{\rm lin}(\boldsymbol{\theta}, z)$. Since $\delta_{\rm lin}$ is a Gaussian field, its statistics are completely determined by its power spectrum $P_{\rm lin}(k, z)$.

Second, we randomly place dark matter halos, which are modeled as biased Poisson tracers of $\delta_{\rm lin}$. More precisely, the probability of a halo in mass range $(M, M + dM)$ and comoving volume $d^3\mathbf{x}$ near spatial location $\mathbf{x}$ is

$$n_h^{3d}(M, z)(1 + b_h(M, z)\delta_{\rm lin}(x))d^3\mathbf{x}dM, \tag{8}$$

where $n_h^{3d}(M, z)$ is the *halo mass function*, or number density of halos per unit comoving volume per unit halo mass, and $b_h(M, z)$ is the *halo bias*. We use the Sheth-Tormen mass function and bias [Eqs. (A4) and (A6)].

Third, we randomly assign FRBs and galaxies to halos. We always assume that the number counts (N_f, N_g) of FRBs and galaxies are independent from one halo to the next. That is, (N_f, N_g) is a bivariate random variable whose probability distribution (the *halo occupation distribution* or HOD) depends only on halo mass M and redshift z. Once the counts (N_f, N_g) have been simulated, we assign spatial locations to each FRB and galaxy independently, by sampling from the Navarro-Frenk-White (NFW) spatial profile [Eq. (A7)]. We assume that galaxy positions are measured with negligible uncertainty, but FRB positions have statistical errors (θ_x, θ_y) that are Gaussian with full width at half maximum (FWHM) denoted θ_f. Unless stated otherwise, we take the FRB angular resolution to be $\theta_f = 1$ arcminute.

C. Fiducial FRB models

Throughout the paper, we derive analytic results for an arbitrary HOD, but show numerical results for two specific FRB models: the fiducial "low-z" and "high-z" FRB models. Our two fiducial models are intended to bracket the range of possibilities for the FRB redshift distribution currently allowed by observations. The median

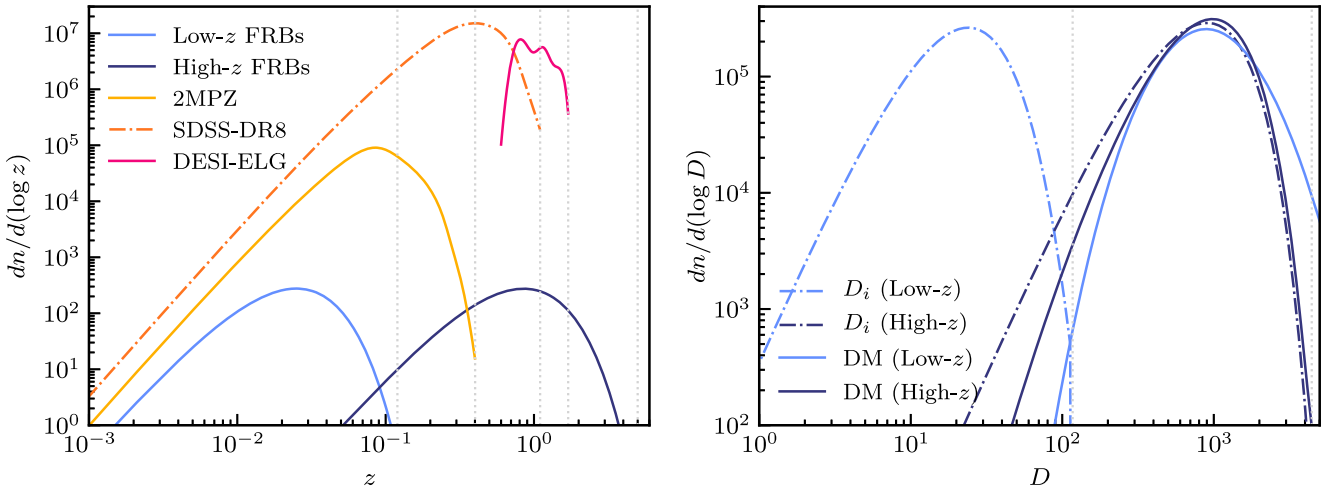

FIG. 1. *Left panel:* FRB redshift distributions in our high-z and low-z fiducial FRB models (see Sec. II), with galaxy redshift distributions shown for comparison. *Right panel:* FRB DM distributions in both fiducial models. We show total extragalactic DM (IGM + host, denoted "DM"), and the IGM contribution $D_i(z)$. The total DM distribution is similar in the two fiducial models, but DMs are usually host dominated in the low-z model and IGM dominated in the high-z model. Vertical dotted lines mark maximum redshift cutoffs.

FRB redshift in the low-z and high-z FRB models is $z \sim 0.022$ and $z \sim 0.76$, respectively. The *host* DM distributions in the two models have been chosen so that the distribution of *total* DMs is similar (Fig. 1). In the high-z FRB model, observed DM is a fairly good indicator of the FRB redshift, whereas in the low-z FRB model, there is not much correlation between DM and redshift. The high-z FRB model was motivated by the FRB host galaxy associations at redshifts 0.19, 0.32, 0.66 reported in [12–16], and the low-z FRB model was motivated by the ASKAP-2MPZ cross-correlation at very low redshift reported in [33].

In both FRB models, we define the FRB HOD so that FRBs have a small nonzero probability to occur in halos above threshold mass $M_f = 10^9\ h^{-1}\ M_\odot$. We have chosen M_f to be small, roughly the minimum halo mass needed to host a dwarf galaxy, since one FRB (the original repeater) is known to be in a dwarf. If M_f is increased (keeping the total number of observed FRBs $N_{\rm FRB}$ fixed), then the FRB-galaxy cross-correlations SNR also increases. Therefore, our choice of small M_f makes our forecasts a bit conservative.

D. Galaxy surveys

We consider three galaxy surveys throughout the paper. First, the SDSS-DR8 optical photometric survey over redshift range $0 \le z \le 1.1$, with redshift distribution taken from [37]. Second, the 2MPZ all-sky infrared photometric survey [38], where almost all (≈98%) of the 2MPZ galaxies have photometric redshifts < 0.2. Finally, the upcoming DESI-ELG spectroscopic survey, whose redshift distribution is forecasted in [39] and covers the range $0.6 \le z \le 1.7$. For photometric surveys, we neglect photometric redshift uncertainties, since these will be small compared to the FRB redshift uncertainty arising from scatter in the FRB host DM.

The galaxy HOD is constructed so that halos above threshold mass $M_g(z)$ contain $(M/M_g(z))$ galaxies on average. The redshift-dependent threshold halo mass $M_g(z)$ is chosen to match the redshift distribution of the galaxy survey ("abundance matching"). Numerical values of $M_g(z)$ are shown in Fig. 2.

For more details of the FRB and galaxy models, including precise specification of the FRB redshift and

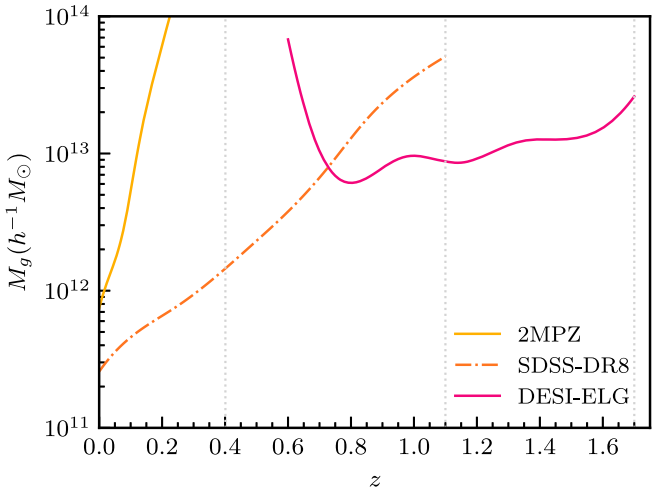

FIG. 2. Threshold halo mass $M_g(z)$ for hosting a galaxy in the 2MPZ, SDSS-DR8, and DESI-ELG galaxy surveys, determined by abundance matching to the redshift distribution dn_g/dz as described in Sec. II and Appendix A 2. Vertical dotted lines mark maximum redshift cutoffs.

host DM distributions in the two fiducial models, see Appendixes A 2 and A 3.

III. THE POWER SPECTRUM C_ℓ^{fg}

A. Definition

Our primary statistic for FRB-galaxy cross-correlations is the angular power spectrum C_ℓ^{fg}, which measures the level of correlation as a function of angular wave number ℓ.

We review the definition of the angular power spectrum. The input data are a catalog of FRB sky locations $\boldsymbol{\theta}_1^f, \ldots, \boldsymbol{\theta}_{N_f}^f$, and a catalog of galaxy sky locations $\boldsymbol{\theta}_1^g, \ldots, \boldsymbol{\theta}_{N_g}^g$. We then define the 2D FRB field $\delta_f(\boldsymbol{\theta})$ as a sum of delta functions,

$$\delta_f(\boldsymbol{\theta}) = \frac{1}{n_f^{2d}} \sum_{i=1}^{N_f} \delta^2(\boldsymbol{\theta} - \boldsymbol{\theta}_i^{(f)}), \tag{9}$$

and similarly for the galaxy field $\delta_g(\boldsymbol{\theta})$.

In Fourier space, the FRB field $\delta_f(\boldsymbol{\ell})$ is a sum of complex exponentials,

$$\delta_f(\boldsymbol{\ell}) = \frac{1}{n_f^{2d}} \sum_{i=1}^{N_f} \exp(-i\boldsymbol{\ell} \cdot \boldsymbol{\theta}_i^{(f)}), \tag{10}$$

and likewise for δ_g. The two-point correlation function of the fields δ_f, δ_g is simplest in harmonic space, where it takes the form

$$\langle \delta_f(\boldsymbol{\ell})^* \delta_g(\boldsymbol{\ell}') \rangle = C_\ell^{fg} (2\pi)^2 \delta^2(\boldsymbol{\ell} - \boldsymbol{\ell}'), \tag{11}$$

where the delta function on the right-hand side (RHS) is a consequence of translation invariance. This equation defines the power spectrum C_ℓ^{fg}.

The power spectrum C_ℓ^{fg} is one representation for the two-point correlation function between δ_f and δ_g. Other representations, such as the two-point correlation function as a function of angular separation, contain the same information as C_ℓ^{fg}. The power spectrum has the advantage that when it is estimated from data, statistical correlations between different ℓ values are small (in contrast with the correlation function, where correlations between different angular separations can be large). For this reason, we choose to use the angular power spectrum throughout the paper.

If the galaxy catalog has been divided into redshift bins, then for each redshift bin j we can define a galaxy field $\delta_{g_j}(\boldsymbol{\theta})$ and a power spectrum $C_\ell^{fg_j}$ by cross-correlating with the (unbinned) FRB catalog.

Similarly, we can bin the FRBs by dispersion measure. Throughout the paper, we assume that the galactic contribution $\mathrm{DM_{gal}}$ can be accurately modeled and subtracted from the observed DM prior to binning. For each FRB DM bin i and galaxy redshift bin j, we can compute an angular power spectrum $C_\ell^{f_i g_j}$. In the limit of narrow redshift and DM bins, the angular power spectrum becomes a function $C_\ell^{fg}(z, D)$ of three variables: angular wave number ℓ, galaxy redshift z, and FRB dispersion measure D.

B. Two-halo and one-halo power spectra

In the halo model, the power spectrum C_ℓ^{fg} can be calculated exactly. Here we summarize the main features of the calculation; details are in Appendix A.

The power spectrum is the sum of *2-halo* and *1-halo* terms,

$$C_\ell^{fg} = C_\ell^{fg(2h)} + C_\ell^{fg(1h)}, \tag{12}$$

which correspond to correlations between FRBs and galaxies in different halos, or in the same halo. Some example 2-halo and 1-halo power spectra are shown in Fig. 3.

The 2-halo term $C_\ell^{fg(2h)}$ is sourced by large-scale cosmological correlations and is responsible for the large bump at low ℓ. For an arbitrary redshift z, the bump is at $\ell \sim k_{\mathrm{eq}} \chi(z)$, where $k_{\mathrm{eq}} \sim 0.02\, h\,\mathrm{Mpc}^{-1}$ is the scale of matter-radiation equality. The 2-halo term arises because FRBs and galaxies trace the same underlying large-scale cosmological density fluctuations. A complete expression for $C_\ell^{fg(2h)}$ is given by Eq. (A39) in Appendix A. Here we give a simplified expression which applies on angular scales larger than the angular size of a halo ($\ell \lesssim \chi / R_{\mathrm{halo}}$, where R_{halo} is the comoving radius of a typical halo):

$$C_\ell^{fg(2h)} \to \frac{1}{n_f^{2d} n_g^{2d}} \int dz \frac{H(z)}{\chi(z)^2} \left(b_f(z) \frac{dn_f^{2d}}{dz} \right) \times \left(b_g(z) \frac{dn_g^{2d}}{dz} \right) P_{\mathrm{lin}}\left(\frac{\ell}{\chi(z)}, z \right). \tag{13}$$

Here, $b_f(z)$ and $b_g(z)$ are bias parameters that measure the coupling of FRBs and galaxies to the cosmological density field on large scales. The FRB bias b_f is defined by the statement that the FRB and matter overdensities are related by $\delta_f \approx b_f \delta_m$ on large scales, and likewise for b_g. An explicit formula for b_f, b_g is given in Eq. (A40), and numerical values are shown in Fig. 4. The 2-halo term mainly depends on the redshift overlap between the FRB and galaxy catalogs, via the factors $(b_f dn_f^{2d}/dz) \times (b_g dn_g^{2d}/dz)$ in Eq. (13).

As a technical comment, power spectra have been computed using the Limber approximation [40–42] throughout the paper. We comment on the accuracy of the Limber approximation in Appendix B.

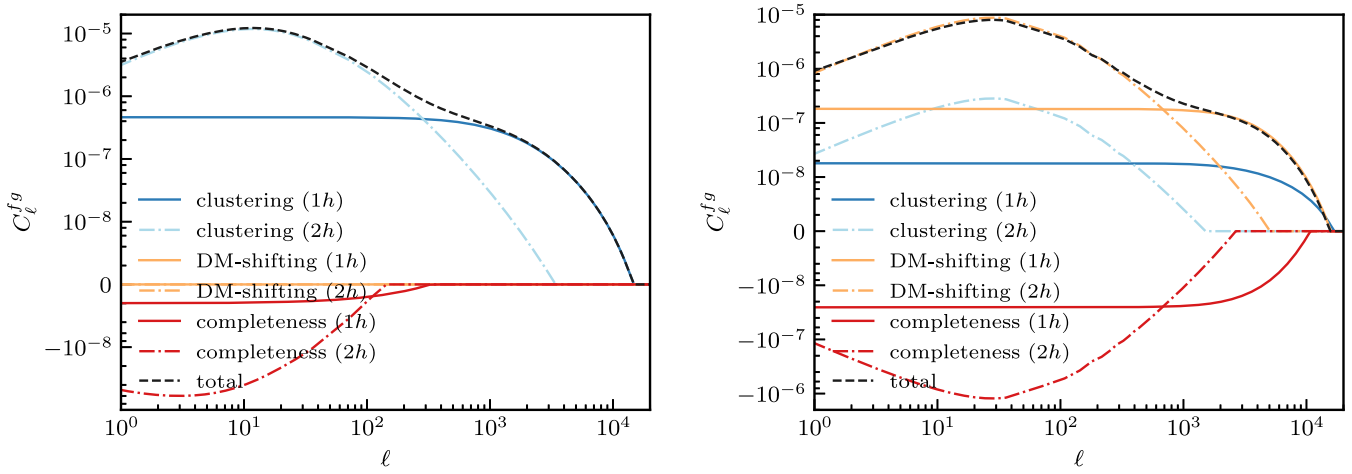

FIG. 3. Angular cross power spectrum C_ℓ^{fg} as a function of ℓ for the fiducial high-z FRB model (see Sec. II) and SDSS-DR8 galaxies. The total observed power spectrum is the sum of clustering and propagation contributions, and each contribution may be split into 1-halo and 2-halo terms, which we show separately here. Disentangling these terms is a challenge and one of the main themes of this paper. The clustering terms are described in Sec. III B, and the "DM-shifting" and "completeness" terms are propagation effects which will be described in Sec. V. *Left panel:* Unbinned FRB and galaxy fields. *Right panel:* FRB dispersion measure bin $1400 < D < 1500$ and galaxy redshift bin $0.63 < z < 0.74$.

The 1-halo term $C_\ell^{fg(1h)}$ arises because FRBs and galaxies occupy the same dark matter halos. A complete expression for $C_\ell^{fg(1h)}$ is given by Eq. (A39) in Appendix A. Here we give a simplified expression that applies on angular scales $\ell \lesssim \chi/R_{\rm halo}$ larger than the angular size of a halo:

$$C_\ell^{fg(1h)} \to \frac{1}{n_f^{2d} n_g^{2d}} \int dz dM \frac{\chi(z)^2}{H(z)} n_h^{3d}(M,z) \langle N_f N_g \rangle_{M,z}, \tag{14}$$

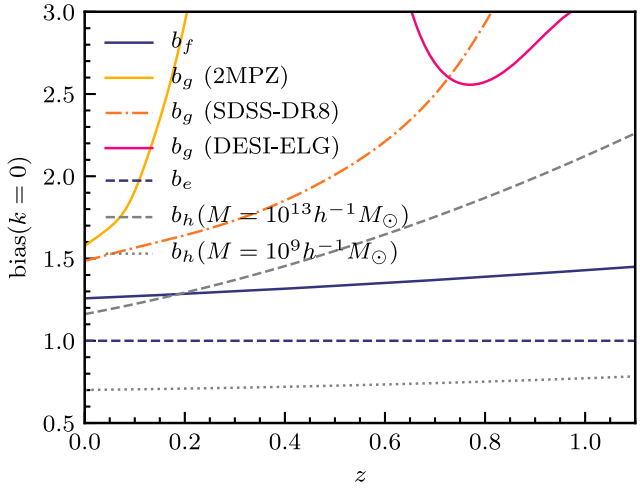

FIG. 4. Large-scale bias parameters. The FRB bias $b_f(z)$ assumes minimum halo mass $M_f = 10^9\ h^{-1}\,M_\odot$. The galaxy bias $b_g(z)$ for the 2MPZ, SDSS-DR8, and DESI-ELG surveys assumes the minimum halo mass $M_g(z)$ from Fig. 2. We take electron bias $b_e = 1$ throughout. We also show the halo bias $b_h(z)$ for two choices of halo mass.

where $\langle\cdot\rangle_{M,z}$ denotes the average over the HOD in a halo of mass M at redshift z.

The 1-halo term is harder to interpret than the 2-halo term, since it depends on the details of the HOD. As an artificial example, suppose that the FRB and galaxy catalogs do overlap in redshift, but the FRB and galaxy HODs do not overlap in halo mass. Then the 1-halo term will be zero. This example is artificial, since halos of sufficiently large mass will contain galaxies of all types, and presumably FRBs as well. However, it illustrates that interpreting the 1-halo term is not straightforward. We will return to this issue shortly.

The 1-halo term $C_\ell^{fg(1h)}$ arises whenever FRBs and survey galaxies occupy the same halos. If FRBs actually inhabit the survey galaxies themselves, there will be an additional "Poisson" term $C_\ell^{fg(p)}$ that dominates on the smallest scales (high ℓ). We have neglected the Poisson term in our forecasts, since we are assuming that the FRB survey has insufficient resolution to associate FRBs and galaxies on a per-object basis, but this does make our forecasts slightly conservative. For more discussion of the Poisson term, see Eq. (A41) in Appendix A.

IV. THE OBSERVABLES $b(dn/dz)$ AND $\gamma(dn/dz)$

In the limit of narrow galaxy redshift and FRB DM bins, the angular power spectrum $C_\ell^{fg}(z, D)$ is a function of three variables: angular wave number ℓ, FRB dispersion measure D, and galaxy redshift z. One may wonder whether the information in C_ℓ^{fg} can be "compressed" into a function of fewer variables.

In this section, we will take a step in this direction, by showing how the ℓ dependence can be absorbed into two

observables, corresponding to the power spectrum amplitude in the 2-halo and 1-halo regimes. These observables, denoted $b(dn/dz)$ and $\gamma(dn/dz)$ for reasons to be explained shortly, will be functions of z and D.

The basic idea is simple. For a narrow galaxy redshift bin $(z, z+\Delta z)$, the 2-halo and 1-halo power spectra in Eqs. (13) and (14) have the following forms for $\ell \lesssim \chi/R_{\rm halo}$:

$$C_\ell^{fg(2h)} \to (\text{Constant}) P_{\rm lin}\left(\frac{\ell}{\chi(z)}, z\right),$$
$$C_\ell^{fg(1h)} \to (\text{Constant}). \tag{15}$$

For $\ell \gtrsim \chi/R_{\rm halo}$, the power spectra acquire additional ℓ dependence which gives information about halo profiles, but we will assume that this profile information is of secondary interest. Thus, the information in the ℓ dependence of the power spectrum can be compressed into two numbers: the coefficients in Eq. (15). Given a measurement of the total power spectrum C_ℓ^{fg}, we can fit for both coefficients jointly, without much covariance between them.

A. The 2-halo observable $b_f(dn/dz)$

Starting with the 2-halo power spectrum, we take Eq. (13) in the limit of a narrow redshift bin $(z, z+\Delta z)$, obtaining

$$C_\ell^{fg(2h)} \to \frac{1}{n_f^{2d}} \frac{H(z)}{\chi(z)^2} \left(b_f(z) \frac{dn_f^{2d}}{dz}\right) b_g(z) P_{\rm lin}\left(\frac{\ell}{\chi(z)}, z\right). \tag{16}$$

All factors on the RHS are known in advance except $b_f(z) dn_f^{2d}/dz$, including the factor $P_{\rm lin}(\ell/\chi(z), z)$ which determines the ℓ dependence. In particular, the galaxy bias $b_g(z)$ can be measured in several ways, for example, by cross-correlating the redshift-binned galaxy catalog with cosmic microwave background (CMB) lensing. Therefore, we can interpret the 2-halo power spectrum amplitude as a measurement of the quantity $b_f(dn_f^{2d}/dz)$.

The observable quantity $b_f(dn_f^{2d}/dz)$ is not as intuitive as the FRB redshift distribution (dn_f^{2d}/dz), but in practice the two are not very different. For example, in our fiducial model with threshold halo mass $M_f = 10^9\ h^{-1}\, M_\odot$, the FRB bias satisfies $1.2 \le b_f \le 1.5$ for $z \le 1$ (see Fig. 4).

This interpretation of the 2-halo amplitude as a measurement of $b(dn/dz)$ is fairly standard and has been explored elsewhere [21–25]. The 1-halo amplitude is less straightforward to interpret and does not seem to have a standard interpretation in the literature. In the rest of this section, we will define an analogous observable $\gamma(dn/dz)$ for the 1-halo amplitude. The definition is not specific to FRBs and may be interesting in the context of other tracer populations.

B. The 1-halo observable $\gamma_f(dn/dz)$

We define the following 3D densities:

$$n_{gg}^{3d}(z) = \int dM\, n_h^{3d}(M,z) \langle N_g^2 \rangle_{M,z}, \tag{17}$$

$$n_{fg}^{3d}(z) = \int dM\, n_h^{3d}(M,z) \langle N_f N_g \rangle_{M,z}, \tag{18}$$

where $\langle\cdot\rangle_{M,z}$ is the expectation value over the HOD for a halo of mass M at redshift z. These can be interpreted as comoving densities of pair counts (g, g') or (f, g) in the same halo. Next we define

$$\gamma_f(z) = \frac{n_g^{3d}(z)}{n_f^{3d}(z)} \frac{n_{fg}^{3d}(z)}{n_{gg}^{3d}(z)}. \tag{19}$$

We will see shortly that the 1-halo amplitude can be interpreted as a measurement of $\gamma_f(dn_f^{2d}/dz)$.

We would like to give an intuitive interpretation of $\gamma_f(z)$. First, note that γ_f is invariant under rescaling the overall abundance of FRBs and galaxies. For example, if we wait until the FRB experiment has detected twice as many FRBs, then densities rescale as $n_{fg}^{3d} \to 2n_{fg}^{3d}$ and $n_f^{3d} \to 2n_f^{3d}$, leaving γ_f unchanged.

Second, note that if the galaxy and FRB HODs were identical (aside from overall abundance), then $\gamma_f(z) = 1$. If the FRB HOD were then modified so that FRBs are in more massive halos (relative to the galaxies), then n_{fg}^{3d} would increase, and $\gamma_f(z)$ will be > 1. Conversely, if the typical FRB inhabits a halo that is less massive than a typical galaxy, then $\gamma_f(z)$ will be < 1.

In Fig. 5, we show $\gamma_f(z)$ for our fiducial HOD [Eqs. (A15) and (A20)] as a function of (M_f, M_g) the threshold halo masses for FRBs and galaxies. Consistent with the previous paragraph, if M_f and M_g are of the same order of magnitude, then γ_f is of order unity. In the regimes $M_f \ll M_g$ and $M_f \gg M_g$, the quantity γ_f will be $\lesssim 1$ and $\gtrsim 1$, respectively.

Now we show how the 1-halo amplitude can be interpreted as a measurement of $\gamma_f(z)(dn_f^{2d}/dz)$. We take Eq. (14) and specialize to a narrow redshift bin $(z, z+\Delta z)$, obtaining

$$C_\ell^{fg(1h)} \to \frac{1}{n_f^{2d}} \frac{n_{fg}^{3d}(z)}{n_g^{3d}(z)}. \tag{20}$$

Similarly, the 1-halo amplitude of the galaxy auto power spectrum is

$$C_\ell^{gg(1h)} \to \frac{1}{n_g^{2d}} \frac{n_{gg}^{3d}(z)}{n_g^{3d}(z)} \tag{21}$$

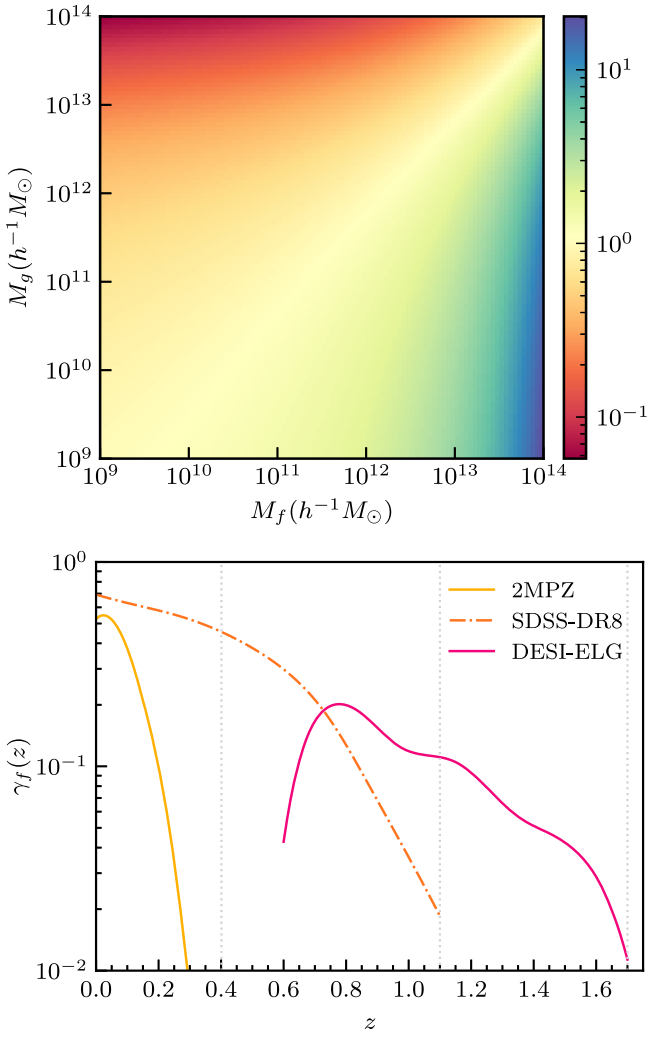

FIG. 5. *Top panel:* Quantity $\gamma_f(z)$ defined in Eq. (19), as a function of threshold FRB halo mass M_f and threshold galaxy mass M_g, for Poisson HODs at redshift $z = 0.5$. If M_f and M_g are comparable, then γ_f is of order 1. *Bottom panel:* Quantity $\gamma_f(z)$ as a function of redshift, assuming FRB threshold halo mass $M_f = 10^9\ h^{-1}\,M_\odot$, and galaxy threshold halo mass $M_g(z)$ from Fig. 2. At high redshifts, γ_f can be $\ll 1$ in our models, since galaxies are rare and our abundance-matching prescription gives a large value of M_g. Vertical dotted lines mark maximum redshift cutoffs.

by specializing Eq. (A39) for $C_\ell^{gg(1h)}$ in Appendix A to low ℓ and a narrow redshift bin. Now we write $C_\ell^{fg(1h)}$ in the following form:

$$C_\ell^{fg(1h)} \to \frac{n_g^{2d}}{n_f^{2d}}\gamma_f(z)\frac{n_f^{3d}(z)}{n_g^{3d}(z)}C_\ell^{gg(1h)} = \frac{\Delta z}{n_f^{2d}}\left(\gamma_f(z)\frac{dn_f^{2d}}{dz}\right)C_\ell^{gg(1h)}, \quad (22)$$

where the second line follows from the first by using Eq. (7). All factors on the RHS are known in advance except $\gamma_f(z)dn_f^{2d}/dz$, including the factor $C_\ell^{gg(1h)}$ which can be measured from the galaxy auto power spectrum. Therefore, the 1-halo amplitude can be interpreted as a measurement of the quantity $\gamma_f(z)dn_f^{2d}/dz$.

Summarizing, we have defined power spectrum observables $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$. By measuring the power spectrum C_ℓ^{fg} as a function of (ℓ, z), both observables may be constrained as functions of z. This extracts all information in C_ℓ^{fg}, except for suppression at high ℓ which contains information about halo profiles. The FRB catalog may be further binned in DM to measure the observables $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$ as functions of (D, z). In the top rows of Figs. 6 and 7, we show the observables as functions of (D, z) in our fiducial model.

V. PROPAGATION EFFECTS

So far, we have considered contributions to C_ℓ^{fg} which arise because 3D positions of FRBs and galaxies are spatially correlated. However, propagation effects also contribute to C_ℓ^{fg}. Galaxies at redshift z_g will spatially correlate with free electrons, which can modulate the observed abundance of FRBs at redshifts $z_f > z_g$, via dispersion, scattering, or lensing. This generates new contributions to C_ℓ^{fg}, which we will study systematically in this section.

Throughout this section, f denotes an FRB catalog, which may be constructed by selecting on FRB properties. For example, f could be a subcatalog of a larger catalog, obtained by selecting a DM bin or a fluence bin.

A. Generalities

Let $\delta_e(\boldsymbol{\theta}, z)$ be the 3D electron overdensity along the past light cone. We will expand propagation effects to first order in δ_e.

Let $\delta_f(\boldsymbol{\theta})$ be the 2D FRB overdensity produced by propagation effects, given a realization of δ_e. We write δ_f as a line-of-sight integral:

$$\delta_f(\boldsymbol{\theta}) = \int dz W_f(z)\delta_e(\boldsymbol{\theta}, z), \quad (23)$$

where this equation defines the "window function" $W_f(z)$. We will show how to calculate $W_f(z)$ shortly.

Given the window function $W_f(z)$, the contribution to C_ℓ^{fg} due to propagation effects may be calculated from Eq. (23). In the Limber approximation, the result is

$$C_\ell^{fg} = \frac{1}{n_g^{2d}}\int dz W_f(z)n_g^{3d}(z)P_{ge}\left(\frac{\ell}{\chi(z)}, z\right), \quad (24)$$

where $P_{ge}(k, z)$ is the 3D galaxy-electron power spectrum at comoving wave number k. We model P_{ge} using the halo

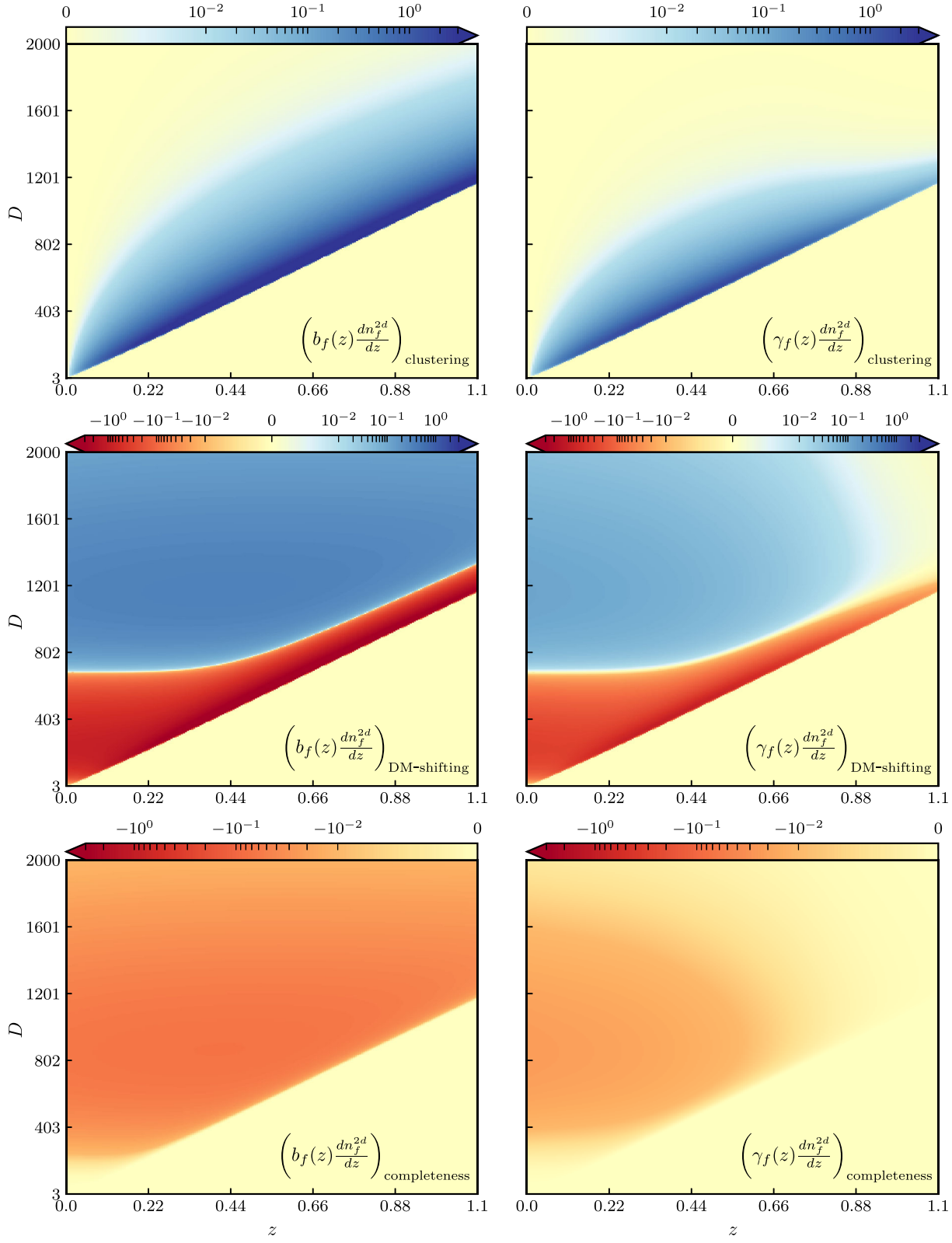

FIG. 6. Visual comparison between clustering and propagation contributions to the clustering power spectrum C_ℓ^{fg}, for our fiducial high-z FRB model and SDSS-DR8. Each row corresponds to one such contribution: clustering (top), DM-shifting propagation effect (middle), and completeness propagation effect (bottom). Since C_ℓ^{fg} is a function of three variables (z, D, ℓ), we compress the ℓ dependence into two clustering observables $b_f dn_f/dz$ (left column) and $\gamma_f dn_f/dz$ (right column), as described in Sec. IV. Qualitatively, it is clear that clustering and propagation effects may be distinguished based on their (z, D) dependence.

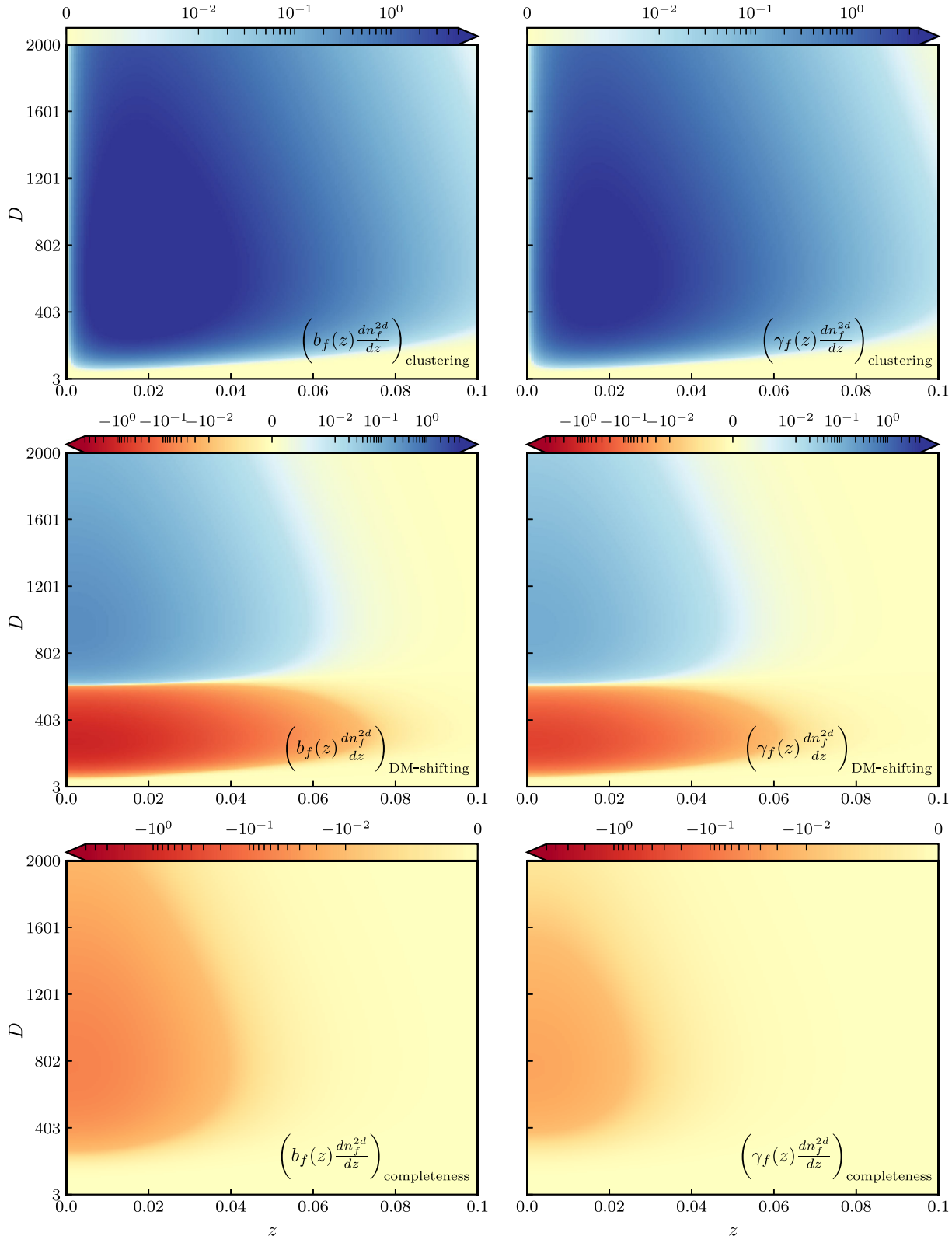

FIG. 7. Same as Fig. 6, but for the fiducial low-z FRB model.

model [Eq. (A42)] in Appendix A. For a narrow galaxy redshift bin $(z, z+\Delta z)$, Eq. (24) becomes

$$C_\ell^{fg} \to \frac{H(z)}{\chi(z)^2} W_f(z) P_{ge}\left(\frac{\ell}{\chi(z)}, z\right). \tag{25}$$

B. Dispersion-induced clustering

In this section we will compute the window function $W_f(z)$ defined by Eq. (23). There will be contributions to $W_f(z)$ from several propagation effects: dispersion, scattering, and lensing. In this paper, we will describe the dispersion case in detail, deferring the other cases to future work.

For an FRB at sky location $\boldsymbol{\theta}$ and redshift z_f, we write the DM as $D = D_i(z_f) + \Delta(\boldsymbol{\theta}, z_f)$, where $\Delta(\boldsymbol{\theta}, z_f)$ is the DM perturbation due to electron anisotropy along the line of sight at redshifts $0 < z < z_f$. Then Δ is given explicitly by

$$\Delta(\boldsymbol{\theta}, z_f) = n_{e,0} \int_0^{z_f} dz \frac{1+z}{H(z)} \delta_e(\boldsymbol{\theta}, z). \tag{26}$$

As usual, let dn_f^{2d}/dz denote the angular number density per unit redshift, so that

$$n_f^{2d} = \int dz \frac{dn_f^{2d}}{dz}. \tag{27}$$

We introduce the notation $(\partial/\partial\Delta)(dn_f^{2d}/dz)$ to denote the derivative of dn_f^{2d}/dz with respect to a foreground DM perturbation $\Delta(z)$ along the line of sight. Then, by differentiating Eq. (27), we can formally write the propagation-induced FRB anisotropy as

$$\delta_f(\boldsymbol{\theta}) = \frac{1}{n_f^{2d}} \int dz_f \Delta(\boldsymbol{\theta}, z_f) \left(\frac{\partial}{\partial\Delta} \frac{dn_f^{2d}}{dz_f}\right).$$

Plugging in Eq. (26) for $\Delta(\boldsymbol{\theta}, z_f)$ and reversing the order of integration, we get

$$\delta_f(\boldsymbol{\theta}) = \frac{n_{e,0}}{n_f^{2d}} \int dz \frac{1+z}{H(z)} \delta_e(\boldsymbol{\theta}, z) \int_z^\infty dz_f \left(\frac{\partial}{\partial\Delta} \frac{dn_f^{2d}}{dz_f}\right). \tag{28}$$

Comparing with the definition of W_f in Eq. (23) we read off the window function:

$$W_f(z) = \frac{n_{e,0}}{n_f^{2d}} \frac{1+z}{H(z)} \int_z^\infty dz' \left(\frac{\partial}{\partial\Delta} \frac{dn_f^{2d}}{dz'}\right). \tag{29}$$

This identity relates the window function W_f to the derivative $(\partial/\partial\Delta)(dn_f^{2d}/dz)$, but it remains to compute the latter quantity. This will depend on the details of how the FRB catalog f is selected.

Generally speaking, the derivative $(\partial/\partial\Delta)(dn_f^{2d}/dz)$ contains two terms. First, there is a term that arises because a DM perturbation changes the probability that an FRB is detected. Increasing DM preserves pulse fluence, but decreases signal to noise.[1] If the FRB catalog is constructed by selecting all objects above a fixed SNR threshold, then this effect gives a negative contribution to $(\partial/\partial\Delta)(dn_f^{2d}/dz)$. We will refer to this contribution as the *completeness* term.

Second, in the case where the FRB catalog is DM binned, there is an additional term in $(\partial/\partial\Delta)(dn_f^{2d}/dz)$ which arises because a DM perturbation can shift observed DMs across a bin boundary. We will refer to this contribution as the *DM-shifting* term.

We give an explicit formula for the DM-shifting term as follows. Suppose that the FRB catalog is constructed by selecting FRBs in DM bin $(D_{\min}, D_{\max})$. Let $(d^2n_f^{2d}/dzdD)$ be the angular number density of FRBs per (redshift, DM), so that

$$\frac{dn_f^{2d}}{dz} = \int_{D_{\min}}^{D_{\max}} dD \frac{d^2n_f^{2d}}{dz\, dD}. \tag{30}$$

Then the DM-shifting term is

$$\left(\frac{\partial}{\partial\Delta}\frac{dn_f^{2d}}{dz}\right)_{\text{DM-shifting}} = \left(\frac{d^2n_f^{2d}}{dz\, dD}\right)_{D_{\min}} - \left(\frac{d^2n_f^{2d}}{dz\, dD}\right)_{D_{\max}}. \tag{31}$$

Next we give an explicit formula for the completeness term. This term is more complicated and depends on both selection and the underlying FRB population. As a toy model for exploring the order of magnitude of this term, we will make the following assumptions:

(1) The FRB catalog is constructed by selecting all objects above threshold signal-to-noise SNR_*.
(2) All FRBs have the same intrinsic pulse width t_i.
(3) In each redshift and DM bin, the FRB luminosity function is Euclidean: the number of FRBs above fluence F_* is proportional to $(F_*^{-3/2})$.[2]
(4) SNR is related to fluence F by

$$\text{SNR} \propto \frac{F}{(t_i^2 + t_s^2 + t_d^2)^{1/4}}, \tag{32}$$

[1]This is true for FRB searches based on incoherent dedispersion, such as the CHIME/FRB real-time search, due to pulse broadening within each frequency channel. If the FRB search were based on coherent dedispersion, then dispersion would not change the SNR. However, a coherent search is computationally infeasible for large blind searches.

[2]The luminosity function is expected to be Euclidean at low z if the FRB catalog is unbinned in redshift. However, within a (z, DM) bin, there is no particular reason why the FRB luminosity function should be Euclidean, so this assumption of our toy model is fairly arbitrary.

where t_s is the instrumental time sample length and t_d is the dispersion delay within a channel, given by

$$t_d = 2\mu(\mathrm{DM})\nu^{-3}(\Delta\nu), \tag{33}$$

where ν is the observing frequency, $(\Delta\nu)$ is the channel bandwidth, and $\mu = 4.15\ \mathrm{ms\,GHz^2}$ is the coefficient in the FRB dispersion relation $(\mathrm{delay}) = \mu(\mathrm{DM})/\nu^2$ in Eq. (2).

Under these assumptions, we can calculate the derivative of $\log d^2n_f/(dzdD)$ with respect to a foreground DM perturbation Δ, as follows:

$$\begin{aligned}\frac{\partial}{\partial\Delta}\left(\log\frac{d^2n_f}{dz\,dD}\right) &= -\frac{3}{2}\frac{\partial\log F_*}{\partial\Delta}\\ &= -\frac{3}{2}\frac{\partial\log(t_i^2+t_s^2+t_d^2)^{1/4}}{\partial\Delta}\\ &= -\frac{3t_d}{4(t_i^2+t_s^2+t_d^2)}\frac{\partial t_d}{\partial\Delta}\\ &= -\frac{3\mu(\Delta\nu)t_d}{2\nu^3(t_i^2+t_s^2+t_d^2)}.\end{aligned} \tag{34}$$

Here, the first line follows from toy model assumption 3, the second line follows from Eq. (32), and the last line follows from differentiating Eq. (33) with respect to DM.

To get the completeness term in the derivative $(\partial/\partial\Delta)(dn_f^{2d}/dz)$, we integrate Eq. (34) over D:

$$\begin{aligned}&\left(\frac{\partial}{\partial\Delta}\frac{dn_f^{2d}}{dz}\right)_{\mathrm{completeness}}\\ &\quad= \int dD\left(\frac{\partial}{\partial\Delta}\frac{d^2n_f^{2d}}{dz\,dD}\right)\\ &\quad= \int dD\left(\frac{d^2n_f^{2d}}{dz\,dD}\right)\left(-\frac{3\mu(\Delta\nu)t_d}{2\nu^3(t_i^2+t_s^2+t_d^2)}\right).\end{aligned} \tag{35}$$

In our toy model, the completeness term always gives a negative contribution to C_ℓ^{fg}, since increasing the DM of an FRB (at fixed fluence) decreases SNR. This is true under the assumptions of our toy model, but is not guaranteed to be true in general. For example, in the CHIME/FRB real-time search, the radiofrequency interference removal pipeline includes a filtering operation which detrends intensity data along its radiofrequency axis, removing the signal from low-DM events. In principle this gives a positive contribution to C_ℓ^{fg}, although end-to-end simulations of the CHIME/FRB triggering pipeline would be needed to determine whether the overall sign is positive or negative.

Summarizing, in this section we have calculated two contributions to C_ℓ^{fg} from propagation effects: a "DM-shifting" term and a "completeness" term. In both cases, the contribution to C_ℓ^{fg} is calculated as follows. We compute the intermediate quantity $(\partial/\partial\Delta)(dn_f^{2d}/dz)$ using Eq. (31) or Eq. (35), then the window function $W_f(z)$ using Eq. (29), and finally C_ℓ^{fg} using Eq. (24).

Finally, other studies have proposed to isolate these propagation effects to measure P_{ge} by cross-correlating galaxies with the 2D field $\bar{\Delta}(\boldsymbol{\theta})$ of DM averaged over all FRBs detected in a particular direction $\boldsymbol{\theta}$. Such statistics are related to the DM moment of C_ℓ^{fg},

$$C_\ell^{\bar{\Delta}g} \propto \sum_i D_i n_{f_i}^{2d} C_\ell^{f_i g}, \tag{36}$$

where f_i denotes the sample of FRBs in DM bin i centered on D_i. Since $C_\ell^{\bar{\Delta}g}$ is a moment of our clustering statistic C_ℓ^{fg}, the former contains a subset of the astrophysical information.

C. Numerical results

In this section, we numerically compare contributions to C_ℓ^{fg} from spatial clustering, and two propagation effects: DM shifting [Eq. (31)] and completeness [Eq. (35)]. For the completeness effect, we have used FRB intrinsic width $t_i = 10^{-3}$ s, and instrumental parameters matching CHIME/FRB: time sampling $t_s = 10^{-3}$ s, channel bandwidth $\Delta\nu = 400$ kHz, and central frequency $\nu = 600$ MHz.

To visualize contributions to C_ℓ^{fg}, we compress the power spectrum into two observables $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$, as described in Sec. IV. To compute these observables for propagation effects, we split the galaxy-electron power spectrum P_{ge} into 2-halo and 1-halo terms [see Eq. (A42) in Appendix A]. For $\ell \lesssim \chi/R_{\mathrm{halo}}$, these take the forms

$$P_{ge}^{2h}(k,z) \to b_g(z)b_e(z)P_{\mathrm{lin}}(k,z), \tag{37}$$

$$P_{ge}^{1h}(k,z) \to \frac{n_{ge}^{3d}(z)}{n_g^{3d}(z)n_e^{3d}(z)}, \tag{38}$$

where $n_e^{3d}(z)$ is the 3D number density of free electrons and $n_{ge}^{3d}(z)$ is defined by

$$n_{ge}^{3d}(z) = \int dM\, n_h^{3d}(M,z)\langle N_g N_e\rangle_{M,z} \tag{39}$$

similar to the definition of $n_{fg}^{3d}(z)$ in Eq. (18). Now a calculation combining Eqs. (16), (22) (25), (37), (38) shows that the contribution to the power spectrum observables $(b_f dn_f^{2d}/dz)$ and $(\gamma_f dn_f^{2d}/dz)$ from propagation effects is

$$\left(b_f\frac{dn_f^{2d}}{dz}\right)_{\mathrm{prop}} = W_f(z)(b_e(z)n_f^{2d}), \tag{40}$$

$$\left(\gamma_f \frac{dn_f^{2d}}{dz}\right)_{\rm prop} = W_f(z)(\gamma_e(z) n_f^{2d}). \tag{41}$$

Here, $b_e(z)$ is the large-scale clustering bias of free electrons, which we will take to be 1. The quantity $\gamma_e(z)$ is defined by

$$\gamma_e(z) = \frac{n_g^{3d}(z)}{n_e^{3d}(z)} \frac{n_{ge}^{3d}(z)}{n_{gg}^{3d}(z)} \tag{42}$$

similar to the definition of $\gamma_f(z)$ given previously.

In Figs. 6 and 7, we show power spectrum observables $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$ from clustering and both propagation effects, in the (DM, z) plane. It is seen that propagation effects are comparable in size to the clustering signal. However, it is qualitatively clear from Figs. 6 and 7 that there is some scope for separating the two based on their dependence on redshift and DM.

D. Ideas for separating spatial clustering from propagation effects

Propagation effects complicate interpretation of the FRB-galaxy cross spectrum C_ℓ^{fg}. For example, suppose a nonzero correlation is observed between high-DM FRBs and low-redshift galaxies. In the absence of propagation effects, this would mean that the FRBs and galaxies must overlap in redshift, implying a significant population of FRBs at low redshift and large host DM. However, in the presence of propagation effects, another possibility is that FRBs are at high redshift and correlated to low-redshift galaxies via propagation effects.

On the other hand, propagation effects add new information to C_ℓ^{fg}. By treating propagation effects as signal rather than noise, it may be possible to learn about the distribution of electrons in the IGM. In this section, we will consider the question of how the spatial clustering and propagation contributions to C_ℓ^{fg} might be separated. Rather than trying to anticipate every observational scenario that may arise, we will present some general ideas.

Propagation effects can sometimes be eliminated by changing the way the FRB catalog is selected. To take the case of dispersion, the DM-shifting term will be eliminated if the FRB catalog is unbinned in DM. Of course, this also throws away information since the DM dependence of the clustering signal is of interest. The completeness term will be eliminated if FRBs are selected in a fluence bin, rather than selecting FRBs above an SNR threshold. The fluence bin must be complete, in the sense that all FRBs in the bin are detected regardless of their dispersion. This may require restricting the cross-correlation to fairly large fluence and discarding low-fluence FRBs in the catalog.

Some propagation effects have a preferred sign; for example, the completeness term in Eq. (35) is negative, since adding dispersion makes FRBs harder to detect.[3] Scattering is another example of a propagation effect with a negative sign, for the same reason.

Propagation effects appear in the C_ℓ^{fg} power spectrum via the product $W_f(z)P_{ge}(\ell/\chi, z)$ [Eq. (24)]. We will discuss separately how the window function $W_f(z)$ and galaxy-electron power spectrum $P_{ge}(k, z)$ might be modeled.

The window function $W_f(z)$ may simplify in the limit of low z. As a concrete example, consider the DM-shifting effect, where the window function is

$$W_f(z) = n_{e,0} \frac{1+z}{H(z)} \times \int_z^\infty \frac{dz'}{n_f^{2d}} \left[\left(\frac{d^2 n_f^{2d}}{dz' dD}\right)_{D_{\rm min}} - \left(\frac{d^2 n_f^{2d}}{dz' dD}\right)_{D_{\rm max}}\right] \tag{43}$$

by combining Eqs. (29) and (31). In the limit of low z this becomes

$$\lim_{z\to 0} W_f(z) = \frac{n_{e,0}}{H_0} \frac{1}{n_f^{2d}} \left[\left(\frac{dn_f^{2d}}{dD}\right)_{D_{\rm min}} - \left(\frac{dn_f^{2d}}{dD}\right)_{D_{\rm max}}\right], \tag{44}$$

where the derivative (dn_f^{2d}/dD) can be estimated directly from data, since it is just the DM derivative of the observed DM distribution.

A similar comment applies to other propagation effects: the $z \to 0$ limit of the window function $W_f(z)$ can be estimated directly from the distribution of observed FRB parameters, plus a model of the instrumental selection. Away from the $z \to 0$ limit, the window function will depend on the FRB redshift distribution, which is not directly observable. On the other hand, this means that if the z dependence of $W_f(z)$ can be measured, it constrains the FRB redshift distribution.

Now we discuss modeling the galaxy-electron power spectrum $P_{ge}(k, z)$. On 2-halo dominated scales, where $P_{ge}(k, z) = b_g(z) b_e(z) P_{\rm lin}(k, z)$, this should be straightforward. The galaxy bias $b_g(z)$ can be determined from either the galaxy auto power spectrum or cross-correlations with gravitational lensing, and the electron bias $b_e(z)$ is expected to be very close to 1.

On 1-halo dominated scales, modeling $P_{ge}(k, z)$ is more difficult. One interesting near-future possibility is to measure $P_{ge}(k, z)$ through the kSZ (kinetic Sunyaev-Zeldovich) effect in the cosmic microwave background. Currently, the kSZ effect has been detected at a few sigma, but not constrained to high precision. However, measurements at the $\approx 10\sigma$ level are imminent, and future CMB experiments such as Simons Observatory and CMB-S4 will measure P_{ge}

[3] As discussed near Eq. (35), this is true for our toy instrumental model, but not guaranteed to be true for a real pipeline.

with percent-level accuracy [43,44]. These measurements will be very informative for modeling FRB propagation effects.

Less futuristically, the galaxy-*matter* power spectrum $P_{gm}(k,z)$ can be measured using cross-correlations between the galaxy catalog and gravitational lensing maps. On large scales, $P_{gm}(k,z)$ and $P_{ge}(k,z)$ are nearly equal, but on smaller scales they will differ since dark matter halo profiles are expected to be more compact than electron profiles. Nevertheless, measuring P_{gm} may be a useful starting point for modeling P_{ge}.

In a scenario where $P_{ge}(k,z)$ has been measured accurately as a function of k, the ℓ dependence of C_ℓ^{fg} is determined, even if the window function $W_f(z)$ is completely unknown. Therefore, it is possible to marginalize over propagation effects by fitting and subtracting a (z-dependent) multiple of $P_{ge}(\ell/\chi,z)$ from C_ℓ^{fg}. This marginalization will degrade clustering information to some extent. In the two-observable picture, statistical errors would increase on one linear combination of $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$.

Summarizing, there are several interesting ideas for modeling the separation of C_ℓ^{fg} into clustering and propagation signals. Which of these ideas proves to be most useful will depend on which observational scenario emerges, and what auxiliary information is available (e.g., kSZ).

VI. FORECASTS AND SIGNAL TO NOISE

A. Fisher matrix formalism

Our basic forecasting tool is the Fisher matrix, which we briefly review. Suppose we have M FRB fields $f_1,\ldots,f_M$ and N galaxy fields $g_1,\ldots,g_N$. We will always assume that galaxy fields are defined by narrow redshift bins, but FRB fields could be defined by binning in DM or a different quantity, or the FRB field could be unbinned ($M=1$).

We assume the FRB-galaxy cross power spectrum is of the form

$$C_\ell^{f_i g_j} = \sum_\mu A_\mu C_\ell^{f_i g_j(\mu)}, \tag{45}$$

where $\mu = 1,\ldots,P$. That is, the power spectrum is the sum of P terms whose ℓ, i, j dependence is fixed by a model, but whose coefficients A_μ are to be determined from data. For example, we could take $\mu \in \{1h, 2h\}$ with $P=2$, to forecast constraints on the overall amplitude of the 1-halo and 2-halo clustering terms. Propagation effects can similarly be included in the forecast.

Given this setup, the P-by-P Fisher matrix is

$$F_{\mu\nu} = \Omega \sum_{ij} \int \frac{\ell\, d\ell}{2\pi} \frac{C_\ell^{f_i g_j(\mu)} C_\ell^{f_i g_j(\nu)}}{C_\ell^{f_i f_i} C_\ell^{g_j g_j}}. \tag{46}$$

We assume that FRB auto power spectra are Poisson noise dominated, i.e.,

$$C_\ell^{f_i f_i} = (n_{f_i}^{2d})^{-1}, \tag{47}$$

but have written $C_\ell^{f_i f_i}$ in Eq. (46) for notational uniformity.

The Fisher matrix is the forecasted *inverse* covariance matrix of the amplitude parameters A_μ in Eq. (45). For example, if $P=1$, then the 1-by-1 Fisher "matrix" F is the SNR^2, and the statistical error on the amplitude parameter is $\sigma(A) = 1/\sqrt{F}$.

A few technical comments: The form of the Fisher matrix in Eq. (46) assumes that FRB and galaxy fields are each uncorrelated, i.e.,

$$C_\ell^{f_i f_j} = \delta_{ij} C_\ell^{f_i f_i}, \qquad C_\ell^{g_i g_j} = \delta_{ij} C_\ell^{g_i g_i}. \tag{48}$$

This assumption is satisfied for FRB fields, since we are assuming that autospectra are Poisson noise dominated. The galaxy fields will also be uncorrelated if they are defined by a set of nonoverlapping redshift bins. Equation (46) also assumes that $C_\ell^{fg} \ll (C_\ell^{ff} C_\ell^{gg})^{1/2}$ in the fiducial model. This will be a good approximation if the FRB number density is not too large. Finally, in Eq. (46) we have written the Fisher matrix as a double sum over (redshift, DM) bins for maximum generality, but for numerical forecasts we take the limit of narrow bins, by replacing the sum by an appropriate double integral.

B. Numerical results

In Table I, we show SNR forecasts for several FRB and galaxy surveys. We report SNR separately for six contributions to the power spectrum C_ℓ^{fg} as follows. First, we split the power spectrum into three terms from gravitational clustering, and the DM-shifting and completeness propagation effects described in Sec. V. We then split each of these terms into 1-halo and 2-halo contributions, for a total of six terms. Each SNR entry in Table I is given by $\sqrt{F_{ii}}$, where F_{ii} is the appropriate diagonal element of the 6-by-6 Fisher matrix. This corresponds to SNR of each contribution considered individually, without marginalizing the amplitude of the other terms in a joint fit.

The forecasts are extremely promising: a CHIME/FRB-like experiment which measures catalogs of $\sim 10^3$ FRBs with few-arcminute angular resolution can measure the clustering signal with high SNR. The precise value depends on the FRB redshift distribution and choice of galaxy survey, but can be as large as ≈ 100 in the low-z FRB model. As a consequence of the high total SNR, the FRB-galaxy correlation can be split up and measured in (z,D) bins, allowing the redshift distribution (or rather, the observables $b_f dn_f/dz$ and $\gamma_f dn_f/dz$) to be measured.

One interesting feature of Table I is that if FRBs do extend to high redshift, the cross-correlation with a

TABLE I. Forecasted SNR for FRB-galaxy cross-correlations. Each row corresponds to a choice of FRB model, galaxy survey, and FRB angular resolution θ_f. Each column corresponds to one contribution to the FRB-galaxy power spectrum. Each entry is total SNR after summing over angular wave number ℓ and a narrow set of redshift and DM bins. We have assumed a catalog with $N_{\rm FRB} = 1000$ FRBs ($D_{\rm max} = 10^4$); in general each SNR value scales as $N_{\rm FRB}^{1/2}$.

	Clustering		DM shifting		Completeness	
	$1h$	$2h$	$1h$	$2h$	$1h$	$2h$
High-z FRB model						
SDSS-DR8, $\theta_f = 1'$	25	6.1	18	5.8	1.2	0.4
SDSS-DR8, $\theta_f = 10'$	6.9	5.8	8.3	5.6	0.57	0.38
SDSS-DR8, $\theta_f = 30'$	2.4	4.9	5	4.9	0.34	0.33
2MPZ, $\theta_f = 1'$	8.2	1.8	10	2.8	0.72	0.2
2MPZ, $\theta_f = 10'$	4.8	1.7	7.4	2.8	0.51	0.2
2MPZ, $\theta_f = 30'$	2.2	1.7	4.8	2.8	0.32	0.19
DESI-ELG, $\theta_f = 1'$	12	4.6	5.4	3.4	0.34	0.22
DESI-ELG, $\theta_f = 10'$	1.9	4.2	0.85	3.1	0.055	0.2
DESI-ELG, $\theta_f = 30'$	0.49	3.2	0.22	2.4	0.014	0.15
Low-z FRB model						
SDSS-DR8, $\theta_f = 1'$	103	14	4.4	0.74	0.28	0.049
SDSS-DR8, $\theta_f = 10'$	87	14	4.1	0.74	0.26	0.049
SDSS-DR8, $\theta_f = 30'$	63	14	3.5	0.74	0.22	0.048
2MPZ, $\theta_f = 1'$	92	13	3.9	0.7	0.25	0.046
2MPZ, $\theta_f = 10'$	82	13	3.7	0.7	0.24	0.046
2MPZ, $\theta_f = 30'$	62	13	3.2	0.7	0.21	0.046

high-redshift galaxy sample is detectable (e.g., SNR = 12 for the high-z FRB model, DESI-ELG, and $\theta_f = 1$ arcminute). Angular cross-correlations should be a powerful tool for probing the high-z end of the FRB redshift distribution, where galaxy surveys are far from complete, and FRB host galaxy associations are difficult.

To get a sense for the level of correlation between different contributions to the FRB-galaxy power spectrum, we rescale the Fisher matrix to a correlation matrix $r_{ij} = F_{ij}/(F_{ii}F_{jj})^{1/2}$ whose entries are between -1 and 1. Using the SDSS-DR8 galaxy survey and high-z FRB model, we get

$$\begin{pmatrix} 1.00 & 0.20 & -0.76 & -0.17 & -0.10 & -0.03 \\ 0.20 & 1.00 & -0.14 & -0.83 & -0.02 & -0.14 \\ -0.76 & -0.14 & 1.00 & 0.19 & -0.22 & -0.04 \\ -0.17 & -0.83 & 0.19 & 1.00 & -0.04 & -0.23 \\ -0.10 & -0.02 & -0.22 & -0.04 & 1.00 & 0.19 \\ -0.03 & -0.14 & -0.04 & -0.23 & 0.19 & 1.00 \end{pmatrix}, \tag{49}$$

where the row ordering is the same as Table I. We see that there is not much correlation between 1-halo and 2-halo signals, but the clustering signal is fairly anticorrelated to the DM-shifting signal. The correlation is not perfect since there is some difference in the (redshift, DM) dependence, as can be seen directly by comparing the top and middle rows of Fig. 6. The correlation matrix depends to some degree on model assumptions. For example, in the low-z FRB model, the correlation matrix is

$$\begin{pmatrix} 1.00 & 0.17 & -0.02 & -0.00 & -0.78 & -0.16 \\ 0.17 & 1.00 & -0.00 & -0.02 & -0.13 & -0.86 \\ -0.02 & -0.00 & 1.00 & 0.19 & -0.19 & -0.04 \\ -0.00 & -0.02 & 0.19 & 1.00 & -0.04 & -0.20 \\ -0.78 & -0.13 & -0.19 & -0.04 & 1.00 & 0.19 \\ -0.16 & -0.86 & -0.04 & -0.20 & 0.19 & 1.00 \end{pmatrix}, \tag{50}$$

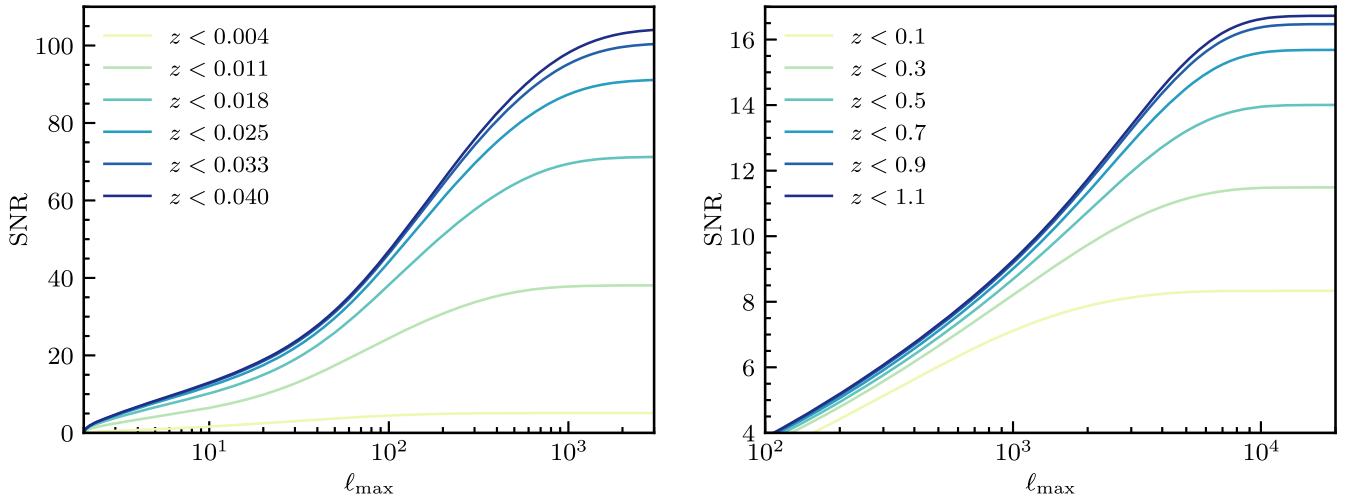

FIG. 8. Forecasted SNR of FRB-galaxy cross power, for varying choices of maximum redshift $z_{\rm max}$ and maximum angular wave number $\ell_{\rm max}$, after summing over narrow (D, z) bins. *Left panel:* Fiducial low-z FRB model and SDSS-DR8 galaxies. *Right panel:* Fiducial high-z FRB model and SDSS-DR8 galaxies.

where we have assumed the SDSS-DR8 galaxy survey. Here, there is a large correlation between clustering and completeness terms. (However, Table I shows that completeness terms are small in the low-z FRB model.)

Figure 8 shows the evolution of total SNR as a function of angular wave number and redshift. In the analysis of real data, large scales ($\ell \lesssim 20$) may be contaminated by Galactic systematic effects, such as dust extinction. Figure 8 shows that these scales make a small contribution to the total SNR, so our forecasts are robust against such systematics.

VII. SIMULATIONS

Our SNR forecasts in the previous section make the approximation that the FRB and galaxy fields are Gaussian. More precisely, we are assuming that the band power covariance of the FRB-galaxy power spectrum is given by the Gaussian (or disconnected) form

$$\mathrm{Cov}(C_b^{fg}, C_{b'}^{fg}) \approx \frac{C_b^{ff} C_b^{gg}}{f_{\mathrm{sky}}(\ell_{\mathrm{max}}^{(b)2} - \ell_{\mathrm{min}}^{(b)2})} \delta_{bb'}, \tag{51}$$

where C_b^{fg} denotes the estimated FRB-galaxy power in a set of nonoverlapping ℓ bands $\ell_{\mathrm{min}}^{(b)} \le \ell \le \ell_{\mathrm{max}}^{(b)}$ with $b = 1, \ldots, N_{\mathrm{bands}}$, and we have assumed $C_\ell^{fg} \ll (C_\ell^{ff} C_\ell^{gg})^{1/2}$.

In reality, FRB and galaxy fields are non-Gaussian. The FRB catalog consists of a modest number of objects that obey Poisson (not Gaussian) statistics. Galaxy catalogs are larger, but Poisson statistics of the underlying halos may be important, since the number of halos is smaller than the number of galaxies. The purpose of this section is to determine whether the Gaussian covariance (51) is a good approximation, by carrying out Monte Carlo simulations of galaxies and FRBs.

A. Description of simulation pipeline

Our simulation pipeline is based on the halo model from Sec. II and Appendix A. We use the high-z FRB model and SDSS-DR8 galaxy survey. Because non-Gaussian effects are expected to be largest for the 1-halo term, our simulation pipeline only includes 1-halo clustering. In particular, we do not simulate the Gaussian linear density field δ_{lin}, because it is not needed to simulate 1-halo clustering.

We use a 10×10 deg^2 sky patch, in the flat-sky approximation with periodic boundary conditions. We sample Poisson random halos in 100 redshift bins, and 500 logarithmically spaced mass bins between M_f and $M_{\mathrm{max}} = 10^{17}\ h^{-1}\ M_\odot$. For each halo, we assign an FRB and galaxy count by sampling a Poisson random variable whose expectation value is given by the HODs in Eqs. (A15) and (A20). For each FRB and galaxy, we assign a 3D location within the halo using the NFW profile [Eq. (A7)]. Angular positions are computed by projecting 3D positions onto the sky patch. In the case of FRBs, we convolve sky locations by the beam [Eq. (A34)]. Finally, FRBs are assigned a random DM, which is the sum of the IGM contribution $D_i(z)$ and a random host contribution D_h [see Eq. (A23)].

Next, we grid the FRB and galaxy catalogs onto a real-space 2049×2049 pixelization with resolution ≈ 0.3 arcminutes, using the cloud-in-cell (CIC) weighting scheme. We take the Fourier transform to obtain Fourier-space fields $\delta_f(\ell)$, $\delta_g(\ell)$. Then, following Eq. (A37), we estimate the angular cross power spectrum C_ℓ^{fg} by averaging the cross power $\langle \delta_f(\ell)^* \delta_g(\ell) \rangle$ in a nonoverlapping set of ℓ bins.

B. Numerical results

We run the pipeline for 10^5 Monte Carlo realizations and find that the cross power spectrum C_ℓ^{fg} of the simulations agrees with the numerical calculation of $C_\ell^{fg(1h)}$, for a few (DM, z) binning schemes. To compare the band power *covariance* to the Gaussian approximation in Eq. (51), we first estimate the covariance of the simulations as

$$\mathrm{Cov}(C_b^{fg}, C_{b'}^{fg}) = \left(\frac{1}{n_{\mathrm{sim}} - 1}\right) \times \sum_{i=1}^{n_{\mathrm{sim}}} (C_b^{fg,i} - \langle C_b^{fg} \rangle)(C_{b'}^{fg,i} - \langle C_{b'}^{fg} \rangle). \tag{52}$$

In Fig. 9, we show the band power *correlation* matrix $r_{bb'}$, obtained from the Monte Carlo covariance matrix $C_{bb'}$ in Eq. (51) by

$$r_{bb'} = \frac{C_{bb'}}{(C_{bb} C_{b'b'})^{1/2}}. \tag{53}$$

For a Gaussian field, $r_{bb'}$ is the identity (distinct band powers are uncorrelated). In our simulations, we do see off-diagonal correlations due to non-Gaussian statistics, but the correlations are small ($\approx 20\%$ for adjacent bands).

In Fig. 10, we compare the total SNR of the FRB-galaxy cross-correlation obtained from simulations to the Gaussian approximation. The total SNR was computed as

$$\mathrm{SNR}^2 = \sum_{b,b'} (C_b^{fg}) \mathrm{Cov}(C_b^{fg}, C_{b'}^{fg})^{-1} (C_{b'}^{fg}), \tag{54}$$

where $\mathrm{Cov}(C_b^{fg}, C_{b'}^{fg})$ is either the Monte Carlo covariance matrix in Eq. (52) or the Gaussian approximation in Eq. (51). From Fig. 10, the total SNR in the simulations agrees almost perfectly with the Gaussian forecast. This indicates that our forecasts in previous sections, which assume Gaussian statistics, are good approximations to the true non-Gaussian statistics of the FRB and galaxy fields.

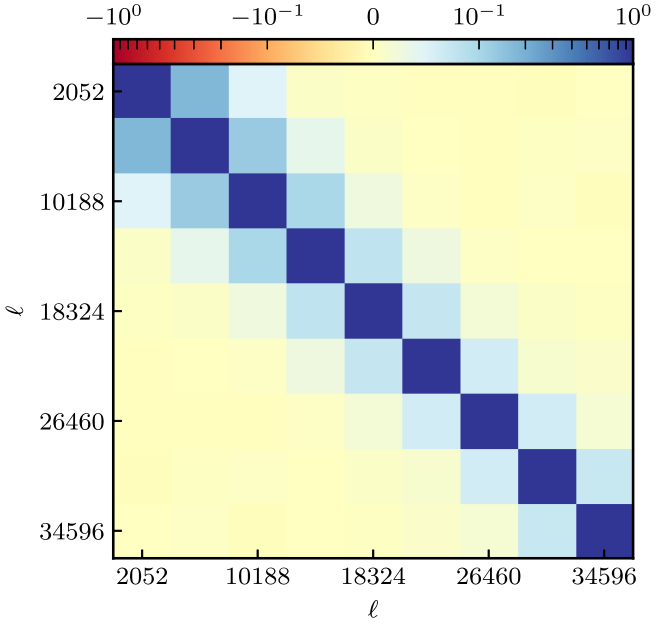

FIG. 9. Band power correlation matrix $r_{bb'}$ of the FRB-galaxy cross power spectrum $C_\ell^{fg(1h)}$, estimated from simulations [see Eq. (53)]. We have used the fiducial high-z FRB model, SDSS-DR8 galaxies, FRB angular resolution $\theta_f = 1'$, and maximum dispersion measure $D_{\rm max} = 10^4$. Correlations between band powers are ≈20% for adjacent ℓ bins, and decay rapidly after that. This is one way of quantifying the importance of non-Gaussian statistics, since off-diagonal correlations would be zero if the FRB and galaxy fields were Gaussian.

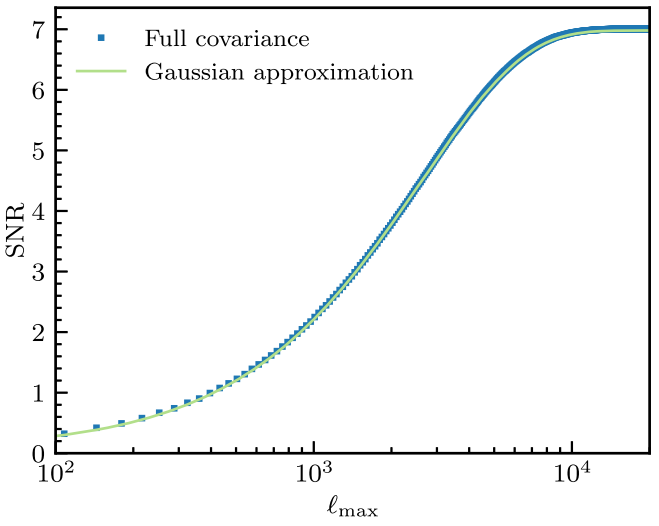

FIG. 10. Cumulative SNR for the FRB-galaxy cross power spectrum $C_\ell^{fg(1h)}$, using the Monte Carlo band power covariance [Eq. (52)], with the Gaussian approximation shown for comparison [Eq. (51)]. The two agree almost perfectly, justifying the Gaussian forecasts used throughout the paper. We have used the fiducial high-z FRB model, SDSS-DR8 galaxies, FRB angular resolution $\theta_f = 1'$, and maximum dispersion measure $D_{\rm max} = 10^4$.

VIII. DISCUSSION

In summary, the use of angular cross-correlations allows telescopes with high mapping speed and modest angular resolution to constrain quantities that appear to require host galaxy associations, such as the FRB redshift distribution. Angular cross-correlations may also be detectable at high redshift, where galaxy surveys are far from complete, and FRB host galaxy associations are difficult. This dramatically extends the scientific reach of instruments such as CHIME/FRB.

One complication is that the FRB redshift distribution (dn_f/dz) is not quite directly measurable. In Sec. IV we studied this issue and showed that there are two clustering observables $(b_f dn_f/dz)$ and $(\gamma_f dn_f/dz)$ in the 2-halo and 1-halo regimes, respectively. Here, b_f is the usual large-scale bias parameter, and the quantity γ_f [defined in Eq. (19)] depends on details of HODs.

Propagation effects can produce contributions to C_ℓ^{fg} that are comparable to the intrinsic clustering signal. This means, for example, that if a nonzero correlation is observed between FRBs and low-redshift galaxies, one cannot definitively conclude that a substantial population of FRBs exists at low z. The correlation could instead be due to the clustering of low-z galaxies with free electrons, which modulate the abundance of FRBs observed at higher z either through selection effects or by shifting FRBs between DM bins.

Propagation effects can be separated from clustering based on their dependence as functions of (z, D, ℓ). This is shown qualitatively in Figs. 6 and 7, where clustering and propagation signals have very different (z, D) dependence [after compressing the ℓ dependence into the two clustering observables $(b_f dn_f/dz)$ and $(\gamma_f dn_f/dz)$]. For a longer, more systematic discussion, see Sec. V D.

Propagation effects are both a potential contaminant of the clustering signal and a potential source of information about ionized electrons in the universe. Indeed, the "DM-shifting" propagation effect identified in Sec. V can be used to probe the distribution of electrons in the circumgalactic medium along the lines of [27–32].

We now interpret our forecasts in relation to the 3σ correlation between ASKAP-discovered FRBs and 2MPZ galaxies measured in [33]. Scaling to a sample of 21 galaxies, and noting the weak dependence on angular resolution, our low-z FRB model predicts an intrinsic clustering correlation SNR of roughly 12, a factor of 4 higher than that observed. While it is not straightforward to interpret SNR units—the difference could be one of signal amplitude, estimator optimality, or modeling—this would nonetheless seem to disfavor a completely nearby population. However, the measured SNR is far greater than what our high-z FRB model predicts and cannot be explained by DM shifting (the measurement was unbinned in DM) or completeness as modeled (wrong sign and too small of

an amplitude). As such, we suggest that the true FRB population may be somewhere between these two models, which could still be consistent with the three direct localizations (at high redshifts $z = 0.19$, 0.32, 0.66).

The results in this paper can be extended in several directions. We have not considered all possible propagation effects (e.g., scattering, plasma lensing), or fully explored the impact of various model assumptions (e.g., free electron profiles). We have explored the effect of binning the FRB catalog in DM, but not binning in other FRB observables. One particularly interesting possibility will be binning FRBs by observed flux F. By measuring the FRB distribution $d^2n_f/(dzdF)$ as a function of redshift and flux, the intrinsic luminosities of FRBs can be constrained.

The galaxy catalog can also be binned in different ways. As an interesting example which also illustrates subtleties in the interpretation, suppose we bin galaxies by estimated star formation rate, in order to determine whether FRBs are statistically associated with star formation. If the FRB-galaxy correlation is observed to be larger for star-forming galaxies, how should this be interpreted?

The answer depends on the angular scale ℓ where the power spectrum C_ℓ^{fg} is measured. On angular scales that are 2-halo dominated, FRBs and galaxies correlate via the observable $(b_f b_g dn_f/dz)$, so the observation just means that the galaxy bias b_g is larger for star-forming galaxies. On 1-halo dominated scales, the observation would imply that FRBs preferentially inhabit *halos* which contain star-forming galaxies, but this does not necessarily imply that FRBs inhabit the star-forming galaxies themselves. Finally, at very high ℓ where C_ℓ^{fg} is dominated by the Poisson term (a regime which we have mostly ignored in this paper, but see discussion in Sec. III), the observation would imply that FRBs do preferentially inhabit star-forming galaxies.

In this paper, we have developed tools for analysis and interpretation of FRB-galaxy cross-correlations. This work was largely motivated by analysis of CHIME/FRB data in progress, to be reported separately in the near future.

ACKNOWLEDGMENTS

K. M. S. was supported by an NSERC Discovery Grant, an Ontario Early Researcher Award, and a CIFAR fellowship. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. We thank Utkarsh Giri, Vicky Kaspi, Dustin Lang, Dongzi Li, and Ue-Li Pen for discussions.

APPENDIX A: HALO MODEL

In this appendix, we describe the model for spatial clustering of FRBs and galaxies used throughout the paper. We use a halo model approach: first we specify the clustering of dark matter halo, and then we specify how halos are populated by FRBs and galaxies.

1. Dark matter halos

We define $\sigma(R, z)$ to be the root-mean-square amplitude of the *linear* density field at redshift z, smoothed with a top hat filter of comoving radius R:

$$\sigma(R,z) = \left(\int \frac{d^3k}{(2\pi)^3} P_{\rm lin}(k,z) W(kR)^2 \right)^{1/2}, \quad \text{(A1)}$$

where $W(x)$ is the Fourier transform of a unit-radius top hat,

$$W(x) = 3\frac{\sin(x) - x\cos(x)}{x^3}, \quad \text{(A2)}$$

and $P_{\rm lin}(k,z)$ is the matter power spectrum in linear perturbation theory, which we compute numerically with CAMB [45]. Throughout, we adopt a flat ΛCDM cosmology with $h = 0.67$, $\Omega_m = 0.315$, $\Omega_b = 0.048$, $A_s = 2 \times 10^{-9}$, $n_s = 0.965$, $\sum_\nu m_\nu = 0.06$ eV, and $T_{\rm CMB} = 2.726$ K.

If M is a halo mass, we define

$$R_M = \left(\frac{3M}{4\pi\rho_m} \right)^{1/3}, \quad \text{(A3)}$$

where ρ_m is the comoving total matter density (dark matter + baryonic). Note that R_M is just the radius of a sphere which encloses mass M in a homogeneous universe. Abusing notation slightly, we define $\sigma(M, z)$ to be equal to $\sigma(R, z)$ evaluated at $R = R_M$.

Let $n_h(M, z)$ be the halo mass function, i.e., the number density of halos per comoving volume per unit halo mass. We use the Sheth-Tormen mass function [46,47], given by

$$n_h(M) = \frac{\rho_{m,0}}{M}\frac{d\log\sigma^{-1}}{dM} f(\sigma),$$
$$f(\sigma) = A\frac{\delta_c}{\sigma}\sqrt{\frac{2a}{\pi}}\left(1 + \left(\frac{\sigma^2}{a\delta_c^2}\right)^p\right)\exp\left(-\frac{a\delta_c^2}{2\sigma^2}\right), \quad \text{(A4)}$$

where $\sigma = \sigma(M, z)$ and

$$a = 0.707, \qquad \delta_c = 1.686, \qquad p = 0.3, \quad \text{(A5)}$$

and $A = 0.3222$ is the normalization that satisfies $\int d(\log\sigma) f(\sigma) = 1$, which means that all matter is formally contained in halos of some (possibly very small) mass M.

We assume that halos are linearly biased Poisson tracers of the cosmological linear density field $\delta_{\rm lin}$, i.e., the number of halos in comoving volume V and mass range $(M, M + dM)$ is a Poisson random variable with mean $dM(dn/dM)\int_V d^3x(1 + b_h(M)\delta_{\rm lin}(x))$. Here, $b_h(M)$ is the Sheth-Tormen halo bias:

$$b_h(M) = 1 + \frac{1}{\delta_c}\frac{d\log f}{d\log\sigma}. \tag{A6}$$

Note that σ, n_h, and b_h are functions of both M and z.

We assume that halos have NFW (Navarro-Frenk-White) density profiles [48]. Recall that the NFW profile $\rho(r)$ has two parameters: the virial radius $r_{\rm vir}$ where the profile is truncated, and the scale radius r_s which appears in the functional form of the profile. Sometimes, we reparametrize by replacing one of these parameters by the concentration $c = r_{\rm vir}/r_s$. The NFW profile $u(r)$ and its Fourier transform $\tilde{u}(k)$ are given by

$$u(r) = \frac{A}{(r/r_s)(1+r/r_s)^2} \quad (r \le r_{\rm vir}), \tag{A7}$$

$$\tilde{u}(k) = 4\pi A r_s^3\left(-\frac{\sin(\kappa c)}{\kappa(1+c)} + (\cos\kappa)[{\rm Ci}(\kappa(1+c)) - {\rm Ci}(\kappa)] + (\sin\kappa)[{\rm Si}(\kappa(1+c)) - {\rm Si}(\kappa)]\right), \tag{A8}$$

where $\kappa = kr_s$, and Si and Ci are the special functions,

$${\rm Si}(x) = \int_0^x dt\frac{\sin t}{t}, \tag{A9}$$

$${\rm Ci}(x) = -\int_x^\infty dt\frac{\cos t}{t} = \gamma + \log(x) + \int_0^x dt\frac{\cos t - 1}{t}, \tag{A10}$$

and $\gamma = 0.577216\ldots$ is Euler's constant. We choose the normalizing constant A in Eqs. (A7) and (A8) to be

$$A = \frac{1}{4\pi r_s^3}\left(\log(1+c) - \frac{c}{1+c}\right)^{-1}. \tag{A11}$$

With this value of A, the profile satisfies $\tilde{u}(0) = \int_0^{r_{\rm vir}} dr(4\pi r^2)u(r) = 1$.

To use the NFW profile, we need expressions for the virial radius $r_{\rm vir}(M,z)$ and halo concentration $c(M,z)$, as functions of halo mass and redshift. For the concentration, we use the fitting function from [49]:

$$\log_{10}c(M,z) = \alpha(z) + \beta(z)\log_{10}\left(\frac{M}{10^{12}h^{-1}\ M_\odot}\right),$$
$$\alpha(z) = 0.537 + 0.488\exp(-0.718z^{1.08}),$$
$$\beta(z) = -0.097 + 0.024z. \tag{A12}$$

For the virial radius, we reparametrize by defining a virial *density*,

$$\rho_{\rm vir} = \frac{3M(1+z)^3}{4\pi r_{\rm vir}^3}, \tag{A13}$$

and then use the fitting function for $\rho_{\rm vir}$ from [50]:

$$\rho_{\rm vir}(z) = 178\Omega_m(z)^{0.45}\rho_{\rm crit}(z) = 178\Omega_m(z)^{0.45}\left(\frac{3}{8\pi G}H(z)^2\right). \tag{A14}$$

The factor $(1+z)^3$ in Eq. (A13) arises because $\rho_{\rm vir}$ is a physical density, whereas $r_{\rm vir}$ is a comoving distance.

2. Galaxies

We assume that the number of galaxies in a halo of mass M is a Poisson random variable whose mean $\bar{N}_g(M,z)$ is given by

$$\bar{N}_g(M,z) = \begin{cases} (M/M_g(z)) & \text{if } M \ge M_g(z) \\ 0 & \text{if } M < M_g \end{cases}, \tag{A15}$$

where $M_g(z)$ is the minimum halo mass needed to host a galaxy.

For each galaxy survey considered in this paper, we compute $M_g(z)$ by matching to the redshift distribution dn_g^{2d}/dz, by numerically solving the equation

$$\frac{dn_g^{2d}}{dz} = \Omega\frac{\chi(z)^2}{H(z)}\int_{M_g(z)}^\infty dM\, n_h(M)\frac{M}{M_g(z)} \tag{A16}$$

for $M_g(z)$. [This procedure for reverse engineering a threshold halo mass $M_g(z)$ from an observed redshift distribution is sometimes called "abundance matching."] The redshift distribution dn_g^{2d}/dz is taken from [37–39] for SDSS-DR8, 2MPZ, and DESI-ELG, respectively. For each survey, the redshift distribution dn_g^{2d}/dz and threshold halo mass $M_g(z)$ are shown in Figs. 1 and 2.

3. FRBs

Similarly, we model the FRB population by starting with a redshift distribution dn_f/dz, which we take to be of the form

$$\frac{dn_f^{2d}}{dz} \propto z^2 e^{-\alpha z} \tag{A17}$$

for $0 \le z \le z_{\rm max}$, where the parameter α and maximum redshift $z_{\rm max}$ are given by

$$\alpha = \begin{cases} 3.5 & \text{(high-}z\text{ FRB model)} \\ 120 & \text{(low-}z\text{ FRB model)} \end{cases}, \tag{A18}$$

$$z_{\rm max} = \begin{cases} 5 & \text{(high-}z\text{ FRB model)} \\ 0.12 & \text{(low-}z\text{ FRB model)} \end{cases}, \tag{A19}$$

for our fiducial high-z and low-z FRB models, respectively. The FRB redshift distribution in both models is shown in Fig. 1.

We assume that the number of FRBs in a halo of mass M is a Poisson random variable whose mean $\bar{N}_f(M)$ is given by

$$\bar{N}_f(M,z) = \begin{cases} \eta(z)(M/M_f) & \text{if } M \geq M_f \\ 0 & \text{if } M < M_f \end{cases}, \quad \text{(A20)}$$

where M_f is the threshold halo mass for hosting an FRB, and $\eta(z)$ is an FRB event rate per threshold halo mass. In the FRB case, we take M_f to be a free parameter and determine $\eta(z)$ by abundance matching to the FRB redshift distribution in Eq. (A17). In detail, we take

$$M_f = 10^9\ h^{-1}\, M_\odot \quad \text{(A21)}$$

in both our fiducial high-z and low-z FRB models. The prefactor $\eta(z)$ is then determined by numerically solving the equation

$$\eta(z) = \frac{dn_f^{2d}}{dz}\left(\Omega\frac{\chi(z)^2}{H(z)}\int_{M_f}^{\infty} dM\, n_h(M)\frac{M}{M_f(z)}\right)^{-1}. \quad \text{(A22)}$$

Thus, our FRB redshift distribution and HOD are parametrized by $(\alpha, z_{\max}, M_f)$, and the total number of observed FRBs N_f which determines the proportionality constant in Eq. (A17).

We model dispersion measures by assuming that the host DM is a log-normal random variable. That is, the probability distribution is

$$p(D_h) = \frac{1}{D_h\sqrt{2\pi\sigma^2_{\log D}}}\exp\left(-\frac{(\log D_h - \mu_{\log D})^2}{2\sigma^2_{\log D}}\right), \quad \text{(A23)}$$

where the parameters $(\mu_{\log D}, \sigma_{\log D})$ are given by

$$\mu_{\log D} = \begin{cases} 4 & \text{(high-}z\text{ FRB model)} \\ 6.78 & \text{(low-}z\text{ FRB model)} \end{cases}, \quad \text{(A24)}$$

$$\sigma_{\log D} = \begin{cases} 1 & \text{(high-}z\text{ FRB model)} \\ 0.63 & \text{(low-}z\text{ FRB model)} \end{cases}. \quad \text{(A25)}$$

The FRB DM distribution in both models is shown in Fig. 1.

We assume that FRBs are observed with a Gaussian beam with FWHM θ_f. In the flat-sky approximation, statistical errors on FRB location (θ_x, θ_y) have the Gaussian probability distribution:

$$p(\theta_x, \theta_y) = \frac{4\log 2}{\pi\theta_f^2}\exp\left(-4\log 2\frac{\theta_x^2+\theta_y^2}{\theta_f^2}\right). \quad \text{(A26)}$$

By default, we take the FRB angular resolution to be $\theta_f = 1$ arcminute.

4. Power spectra

Given the model for halos, FRBs, and galaxies from the previous sections, we are interested in angular power spectra of the form C_ℓ^{XY}, where each 2D field X, Y could be either a galaxy field (denoted g) or an FRB field (denoted f). We are primarily interested in cross power spectra C_ℓ^{fg}, but autospectra (C_ℓ^{ff}, C_ℓ^{gg}) also arise when forecasting signal to noise [e.g., Eq. (46)].

For maximum generality, we assume binned FRB and galaxy fields. That is, the galaxy field is defined by specifying a redshift bin $(z_{\min}, z_{\max})$ and keeping only galaxies which fall in this range. Similarly, the FRB field is defined by keeping only galaxies in the DM bin $(D_{\min}, D_{\max})$, after subtracting the galactic contribution $\mathrm{DM}_{\mathrm{gal}}$. Note that the unbinned galaxy field can be treated as a special case, by taking the redshift bin large enough to contain all galaxies (and analogously for the FRB field).

Before computing the power spectrum C_ℓ^{XY}, we pause to define some new notation.

For each tracer field X, let $\bar{N}_X(M,z)$ denote the mean number of tracers in a halo of mass M at redshift z. If X is a binned galaxy field, in redshift bin $(z_{\min}, z_{\max})$, then $\bar{N}_X(M,z)$ is given by

$$\bar{N}_g(M,z) = \begin{cases} \frac{M}{M_g(z)} & \text{if } M \geq M_g(z) \quad \text{and} \quad z \in [z_{\min}, z_{\max}] \\ 0 & \text{otherwise} \end{cases}, \quad \text{(A27)}$$

generalizing Eq. (A15) for an unbinned galaxy field. If X is a binned FRB field, in DM bin $(D_{\min}, D_{\max})$, then

$$\bar{N}_f(M,z) = \begin{cases} \eta(z)\frac{M}{M_f}\int_{D_{\min}-D_i(z)}^{D_{\max}-D_i(z)} dD_h p(D_h) & \\ \qquad\qquad \text{if } M \geq M_f & \\ 0 \qquad\qquad \text{if } M < M_f & \end{cases}, \quad \text{(A28)}$$

generalizing Eq. (A20) for an unbinned FRB field. Here, $p(D_h)$ is the host DM probability distribution in Eq. (A23), and $D_i(z)$ is the IGM contribution to the DM at redshift z [Eq. (5)].

For each tracer field X, let $n_X^{3d}(z)$ be the 3D comoving number density, and let n_X^{2d} be the 2D angular number density. These densities can be written explicitly as follows:

$$n_X^{3d}(z) = \int dM\, n_h(M)\bar{N}_X(M,z), \quad \text{(A29)}$$

$$n_X^{2d} = \int dz \frac{\chi(z)^2}{H(z)} n_X^{3d}(z). \quad \text{(A30)}$$

Next, for a pair of tracer fields (X, Y), let n^{2d}_{XY} denote the angular number density of object pairs (x, y) which are colocated. In our fiducial model, each FRB and galaxy is randomly placed within its halo, so n^{2d}_{XY} is zero unless the fields X, Y contain the same objects. That is, if the galaxy fields in nonoverlapping redshift bins are denoted $g_1, \ldots, g_M$, and the FRB fields in nonoverlapping DM bins are denoted $f_1, \ldots, f_N$, then

$$n^{2d}_{f_i f_j} = n^{2d}_{f_i}\delta_{ij}, \qquad n^{2d}_{g_i g_j} = n^{2d}_{g_i}\delta_{ij}, \qquad n^{2d}_{f_i g_j} = 0. \tag{A31}$$

One final definition: For each tracer field X, let $u^X_\ell(M, z)$ denote the angular tracer profile sourced by a halo of mass M at redshift z, normalized to $u = 1$ at $\ell = 0$. The quantity $u^X_\ell(M, z)$ can be written explicitly as

$$u^g_\ell(M, z) = \tilde{u}(M, k, z)_{k=\ell/\chi(z)}, \tag{A32}$$

$$u^f_\ell(M, z) = b_\ell \tilde{u}(M, k, z)_{k=\ell/\chi(z)}, \tag{A33}$$

in the galaxy and FRB cases, respectively. Here, $\tilde{u}$ is the 3D NFW profile in Eq. (A8), and

$$b_\ell \equiv \exp\left(-\frac{\theta_f^2 \ell^2}{16 \log 2}\right) \tag{A34}$$

is the Fourier-transformed FRB error distribution from Eq. (A26).

Armed with the notation above, we can calculate the power spectrum C^{XY}_ℓ in a uniform way which applies to all choices of tracer fields X, Y. The calculation follows a standard halo model approach, and we present it in streamlined form.

Each tracer field X is derived from a catalog of objects at sky locations $\boldsymbol{\theta}^X_1, \ldots, \boldsymbol{\theta}^X_N$. The 2D field X is a sum of delta functions in real space, or a sum of complex exponentials in Fourier space:

$$X(\boldsymbol{\theta}) = \frac{1}{n^{2d}_X}\sum_j \delta^2(\boldsymbol{\theta} - \boldsymbol{\theta}^X_j), \tag{A35}$$

$$X(\boldsymbol{\ell}) = \frac{1}{n^{2d}_X}\sum_j e^{-i\boldsymbol{\ell}\cdot\boldsymbol{\theta}^X_j}, \tag{A36}$$

and likewise for Y. The power spectrum C^{XY}_ℓ is defined by the equation

$$\begin{aligned}\langle X(\boldsymbol{\ell})^* Y(\boldsymbol{\ell}')\rangle &= \frac{1}{n^{2d}_X n^{2d}_Y}\Big\langle \sum_{jk} e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}^X_j - i\boldsymbol{\ell}'\cdot\boldsymbol{\theta}^Y_k}\Big\rangle \\ &= C^{XY}_\ell (2\pi)^2 \delta^2(\boldsymbol{\ell} - \boldsymbol{\ell}').\end{aligned} \tag{A37}$$

The double sum $\sum_{jk}(\cdots)$ can be split into three terms: a sum over pairs (j, k) of objects in different halos, a sum over pairs (j, k) of noncolocated objects in the same halo, and a sum over colocated pairs (j, k). Correspondingly, the power spectrum C^{XY}_ℓ is the sum of "2-halo," "1-halo," and "Poisson" terms

$$C^{XY}_\ell = C^{XY(2h)}_\ell + C^{XY(1h)}_\ell + C^{XY(p)}_\ell, \tag{A38}$$

which are given explicitly as follows:

$$\begin{aligned}C^{XY(2h)}_\ell &= \frac{1}{n^{2d}_X n^{2d}_Y}\int dz \frac{\chi(z)^2}{H(z)} n^{3d}_X(z) n^{3d}_Y(z) \\ &\quad \times b_X(z, \ell) b_Y(z, \ell) P_{\rm lin}(k, z), \\ C^{XY(1h)}_\ell &= \frac{1}{n^{2d}_X n^{2d}_Y}\int dz\, dM \frac{\chi(z)^2}{H(z)} n_h(M, z) \\ &\quad \times \bar{N}_X(M, z)\bar{N}_Y(M, z) u^X_\ell(M, k) u^Y_\ell(M, k), \\ C^{XY(p)}_\ell &= \frac{n^{2d}_{XY}}{n^{2d}_X n^{2d}_Y},\end{aligned} \tag{A39}$$

where in the first line we have defined

$$\begin{aligned}b_X(z, \ell) \equiv \frac{1}{n^{3d}_X(z)}\int dM b_h(M, z) n_h(M, z) \\ \times \bar{N}_X(M, z) u^X_\ell(M, z).\end{aligned} \tag{A40}$$

On large scales (where $u_\ell = 1$), the quantity $b_X(z, \ell)$ reduces to the bias parameter $b_X(z)$ defined in Sec. III.

Throughout this paper, we have generally neglected the Poisson term in C^{fg}_ℓ, which arises if FRBs are actually located in survey galaxies (in contrast to the 1-halo term, which arises if FRBs are in the same halos as the survey galaxies). This is equivalent to our assumption in Eq. (A31) that $n^{2d}_{fg} = 0$. If this assumption is relaxed, then $C^{fg(p)}_\ell$ will be given by

$$C^{fg(p)}_\ell = b_\ell \frac{n^{2d}_{fg}}{n^{2d}_f n^{2d}_g}, \tag{A41}$$

where the FRB beam convolution b_ℓ has been inserted by hand into the general expression in Eq. (A39), since the FRB beam displaces FRBs relative to their host galaxies.

5. Free electrons

When modeling propagation effects (Sec. V), the 3D galaxy-electron power spectrum $P_{ge}(k, z)$ appears. This can also be computed in the halo model.

For simplicity, we will assume the approximation that all electrons are ionized. This is a fairly accurate approximation: the actual ionization fraction is expected to be ≈90%, with the remaining 10% of electrons in stars, or "self-shielding" HI (neutral hydrogen) regions in galaxies.

We will also make the approximation that electrons have the same halo profiles as dark matter. This is a good approximation on large scales, but may overpredict P_{ge} on

small scales by a factor of a few. This happens because dark matter is pressureless, whereas electrons have associated gas pressure, which "puffs out" the profile. In this paper our goal is modeling propagation effects at the order-of-magnitude level, and it suffices to approximate electron profiles by dark matter profiles. For a more precise treatment, fitting functions for electron profiles could be used [51].

Under these approximations, P_{ge} is the sum $P_{ge} = P_{ge}^{1h} + P_{ge}^{2h}$ of one-halo and two-halo terms, given by

$$\begin{aligned} P_{ge}^{1h}(k,z) &= \frac{1}{\rho_{m,0} n_g^{3d}(z)} \int dM\, M n_h(M,z) \\ &\quad \times \bar{N}_g(M,z)\tilde{u}(M,k,z)^2, \\ P_{ge}^{2h}(k,z) &= b_g(k,z) b_e(k,z) P_{\rm lin}(k,z), \end{aligned} \tag{A42}$$

where

$$\begin{aligned} b_e(k,z) &= \frac{1}{\rho_{m,0}} \int dM\, M b_h(M,z) n_h(M,z) \tilde{u}(M,k,z), \\ b_g(k,z) &= \frac{1}{n_g^{3d}(z)} \int dM\, b_h(M,z) n_h(M,z) \\ &\quad \times \bar{N}_g(M,z)\tilde{u}(M,k,z). \end{aligned} \tag{A43}$$

Note that $b_e(k,z) \to 1$ as $k \to 0$. Intuitively, the large-scale bias of free electrons is 1 in our model because electrons perfectly trace dark matter ($\delta_e = \delta_m$).

APPENDIX B: ACCURACY OF THE LIMBER APPROXIMATION

Throughout the paper, angular power spectra have been calculated using the Limber approximation [40–42]. Let X, Y be 2D fields which are obtained from the 3D density field $\delta_{\rm lin}(\mathbf{x})$ by line-of-sight integration:

$$\begin{aligned} X(\boldsymbol{\theta}) &= \int d\chi W_X(\chi) \delta_{\rm lin}(\chi\boldsymbol{\theta}), \\ Y(\boldsymbol{\theta}) &= \int d\chi W_Y(\chi) \delta_{\rm lin}(\chi\boldsymbol{\theta}), \end{aligned} \tag{B1}$$

where $W_X(\chi)$ and $W_Y(\chi)$ are radial weight functions. Then the Limber approximation is

$$C_\ell^{XY} \approx \int \frac{d\chi}{\chi^2} W_X(\chi) W_Y(\chi) P_{\rm lin}(k,\chi)_{k=\ell/\chi}. \tag{B2}$$

For example, the 2-halo power spectrum $C_\ell^{fg(2h)}$ in Eq. (13) was calculated by applying the Limber approximation with weight functions

$$W_X(z) = b_X(z) \frac{dn_X^{2d}}{dz} \quad (X \in \{f,g\}). \tag{B3}$$

How accurate is the Limber approximation? For a detailed analysis, including explicit calculation of subleading terms, see [52]. At back-of-the-envelope level, the Limber approximation is accurate if

$$\ell \gg \left(\frac{d\log W_X}{d\log\chi} \frac{d\log W_Y}{d\log\chi} \right)^{1/2}. \tag{B4}$$

In most of this paper, the factors $(d\log W/d\log\chi)$ are of order 1, and therefore the Limber approximation is accurate for $\ell \gg 1$. There is one exception: when we calculate C_ℓ^{fg} for a narrow FRB DM slice in our fiducial high-z FRB model (Sec. II), we have $(d\log W/d\log\chi) \approx z_f/(\Delta z_f)$, where the mean FRB redshift z_f can be as large as 3, and the width (Δz_f) of the FRB redshift distribution can be as small as 0.1. In this case, the Limber approximation will still be accurate for $\ell \gg (3/0.1)^{1/2} \approx 6$, which is sufficient for purposes of this paper.

One more subtle point. We sometimes consider the limit of narrow redshift bins, for example, when computing Fisher matrix forecasts in Sec. VI. Generally speaking, the Limber approximation for C_ℓ^{fg} breaks down when the redshift bin width (Δz) is taken to zero. However, the Fisher forecast converges as $(\Delta z) \to 0$: the Fisher matrix with $(\Delta z) = 0$ is nearly equal to the Fisher matrix with $(\Delta z) = 0.1$ (or smaller), and the Limber approximation is still accurate at $(\Delta z) = 0.1$. This is partly because the low-ℓ end of harmonic space contains a small area (see, e.g., Fig. 8). Therefore, the Limber-approximated narrow-bin Fisher matrix is a good approximation to the exact Fisher matrix.

[1] D. R. Lorimer, M. Bailes, M. A. McLaughlin, D. J. Narkevic, and F. Crawford, Science **318**, 777 (2007).
[2] J. I. Katz, Mod. Phys. Lett. A **31**, 1630013 (2016).
[3] E. Platts, A. Weltman, A. Walters, S. P. Tendulkar, J. E. B. Gordin, and S. Kandhai, Phys. Rep. **821**, 1 (2019).
[4] E. Petroff, J. W. T. Hessels, and D. R. Lorimer, Astron. Astrophys. Rev. **27**, 4 (2019).
[5] J. M. Cordes and T. J. W. Lazio, arXiv:astro-ph/0207156.
[6] J. M. Y. Yao, R. N. Manchester, and N. Wang, arXiv:1610.09448.

[7] E. Petroff, E. D. Barr, A. Jameson, E. F. Keane, M. Bailes, M. Kramer, V. Morello, D. Tabbara, and W. van Straten, Pub. Astron. Soc. Aust. **33**, e045 (2016).
[8] L. G. Spitler *et al.*, Nature (London) **531**, 202 (2016).
[9] P. Scholz *et al.*, Astrophys. J. **833**, 177 (2016).
[10] M. Amiri *et al.* (CHIME/FRB Collaboration), Nature (London) **566**, 235 (2019).
[11] B. C. Andersen *et al.* (CHIME/FRB Collaboration), Astrophys. J. **885**, L24 (2019).
[12] S. Chatterjee *et al.*, Nature (London) **541**, 58 (2017).
[13] B. Marcote *et al.*, Astrophys. J. **834**, L8 (2017).
[14] S. P. Tendulkar *et al.*, Astrophys. J. **834**, L7 (2017).
[15] K. W. Bannister *et al.*, arXiv:1906.11476.
[16] V. Ravi *et al.*, Nature (London) **572**, 352 (2019).
[17] T. Eftekhari and E. Berger, Astrophys. J. **849**, 162 (2017).
[18] K. W. Masui *et al.*, arXiv:1710.08591.
[19] M. Amiri *et al.* (CHIME/FRB Collaboration), arXiv:1803.11235.
[20] H. Aihara *et al.* (SDSS Collaboration), Astrophys. J. Suppl. Ser. **193**, 29 (2011).
[21] M. McQuinn and M. White, Mon. Not. R. Astron. Soc. **433**, 2857 (2013).
[22] B. Ménard *et al.*, arXiv:1303.4722.
[23] M. Rahman, B. Ménard, R. Scranton, S. J. Schmidt, and C. B. Morrison, Mon. Not. R. Astron. Soc. **447**, 3500 (2015).
[24] E. D. Kovetz, A. Raccanelli, and M. Rahman, Mon. Not. R. Astron. Soc. **468**, 3650 (2017).
[25] S. Passaglia, A. Manzotti, and S. Dodelson, Phys. Rev. D **95**, 123508 (2017).
[26] L. Hui, E. Gaztanaga, and M. LoVerde, Phys. Rev. D **76**, 103502 (2007).
[27] K. W. Masui and K. Sigurdson, Phys. Rev. Lett. **115**, 121301 (2015).
[28] M. McQuinn, Astrophys. J. **780**, L33 (2014).
[29] M. Shirasaki, K. Kashiyama, and N. Yoshida, Phys. Rev. D **95**, 083012 (2017).
[30] V. Ravi, Astrophys. J. **872**, 88 (2019).
[31] J. B. Muñoz and A. Loeb, Phys. Rev. D **98**, 103518 (2018).
[32] M. S. Madhavacheril, N. Battaglia, K. M. Smith, and J. L. Sievers, Phys. Rev. D **100**, 103532 (2019).
[33] D. Li, A. Yalinewich, and P. C. Breysse, arXiv:1902.10120.
[34] K. Bannister *et al.*, Astrophys. J. **841**, L12 (2017).
[35] R. M. Shannon *et al.*, Nature (London) **562**, 386 (2018).
[36] A. Cooray and R. K. Sheth, Phys. Rep. **372**, 1 (2002).
[37] E. S. Sheldon, C. Cunha, R. Mandelbaum, J. Brinkmann, and B. A. Weaver, Astrophys. J. Suppl. Ser. **201**, 32 (2012).
[38] M. Bilicki, T. H. Jarrett, J. A. Peacock, M. E. Cluver, and L. Steward, Astrophys. J. Suppl. Ser. **210**, 9 (2014).
[39] A. Aghamousa *et al.* (DESI Collaboration), arXiv:1611.00036.
[40] D. N. Limber, Astrophys. J. **117**, 134 (1953).
[41] N. Kaiser, Astrophys. J. **388**, 272 (1992).
[42] N. Kaiser, Astrophys. J. **498**, 26 (1998).
[43] K. M. Smith *et al.*, arXiv:1810.13423.
[44] P. Ade *et al.* (Simons Observatory), J. Cosmol. Astropart. Phys. 02 (2019) 056.
[45] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000).
[46] R. K. Sheth and G. Tormen, Mon. Not. R. Astron. Soc. **329**, 61 (2002).
[47] D. Reed, R. Bower, C. Frenk, A. Jenkins, and T. Theuns, Mon. Not. R. Astron. Soc. **374**, 2 (2007).
[48] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **490**, 493 (1997).
[49] A. A. Dutton and A. V. Macci, Mon. Not. R. Astron. Soc. **441**, 3359 (2014).
[50] V. R. Eke, J. F. Navarro, and C. S. Frenk, Astrophys. J. **503**, 569 (1998).
[51] N. Battaglia, J. Cosmol. Astropart. Phys. 08 (2016) 058.
[52] M. Loverde and N. Afshordi, Phys. Rev. D **78**, 123506 (2008).