

CHIME/FRB Catalog 1 results: statistical cross-correlations with large-scale structure

MASOUD RAFIEI-RAVANDI ,^{1,2} KENDRICK M. SMITH ,¹ DONGZI LI ,³ KIYOSHI W. MASUI ,^{4,5}
 ALEXANDER JOSEPHY ,^{6,7} MATT DOBBS ,^{6,7} DUSTIN LANG ,^{1,2} MOHIT BHARDWAJ ,^{6,7} CHITRANG PATEL ,^{6,8}
 KEVIN BANDURA ,^{9,10} SABRINA BERGER ,^{6,7} P. J. BOYLE ,^{6,7} CHARANJOT BRAR ,⁶ DANIELA BREITMAN ,^{8,11,12}
 TOMAS CASSANELLI ,^{8,12} PRAGYA CHAWLA ,^{6,7} FENGQIU ADAM DONG ,¹³ EMMANUEL FONSECA ,^{6,7}
 B. M. GAENSLER ,^{8,12} UTKARSH GIRI ,^{1,2} DEBORAH C. GOOD ,¹³ MARK HALPERN ,¹³ JANE KACZMAREK ,¹⁴
 VICTORIA M. KASPI ,^{6,7} CALVIN LEUNG ,^{4,5} HSIU-HSIEN LIN ,¹⁵ JUAN MENA-PARRA ,⁴ B. W. MEYERS ,¹³
 D. MICHELLI ,^{6,7} MORITZ MÜNCHMEYER ,¹⁶ CHERRY NG ,⁸ EMILY PETROFF ,^{6,7,17,18} ZIGGY PLEUNIS ,^{6,7}
 MUBDI RAHMAN ,¹⁹ PRANAV SANGHAVI ,^{9,10} PAUL SCHOLZ ,⁸ KAITLYN SHIN ,^{4,5} INGRID H. STAIRS ,¹³
 SHRIHARSH P. TENDULKAR ,^{20,21} KEITH VANDERLINDE ,^{8,12} AND ANDREW ZWANIGA ,⁶

¹Perimeter Institute for Theoretical Physics, 31 Caroline Street N, Waterloo, ON N25 2YL, Canada

²Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada

³Cahill Center for Astronomy and Astrophysics, California Institute of Technology, 1216 E. California Boulevard, Pasadena, CA 91125, USA

⁴MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA

⁵Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA

⁶Department of Physics, McGill University, 3600 rue University, Montréal, QC H3A 2T8, Canada

⁷McGill Space Institute, McGill University, 3550 rue University, Montréal, QC H3A 2A7, Canada

⁸Dunlap Institute for Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

⁹Lane Department of Computer Science and Electrical Engineering, 1220 Evansdale Drive, PO Box 6109, Morgantown, WV 26506, USA

¹⁰Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA

¹¹Department of Physics, University of Toronto, 60 St. George Street, Toronto, ON M5S 1A7, Canada

¹²David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

¹³Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1 Canada

¹⁴Dominion Radio Astrophysical Observatory, Herzberg Research Centre for Astronomy and Astrophysics, National Research Council Canada, PO Box 248, Penticton, BC V2A 6J9, Canada

¹⁵Canadian Institute for Theoretical Astrophysics, 60 St. George Street, Toronto, ON M5S 3H8, Canada

¹⁶Department of Physics, University of Wisconsin-Madison, 1150 University Ave., Madison, WI 53706, USA

¹⁷Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

¹⁸Veni Fellow

¹⁹Sidrat Research, PO Box 73527 RPO Wychwood, Toronto, ON M6C 4A7, Canada

²⁰National Centre for Radio Astrophysics, Post Bag 3, Ganeshkhind, Pune, 411007, India

²¹Department of Astronomy and Astrophysics, Tata Institute of Fundamental Research, Mumbai, 400005, India

ABSTRACT

The CHIME/FRB Project has recently released its first catalog of fast radio bursts (FRBs), containing 492 unique sources. We present results from angular cross-correlations of CHIME/FRB sources with galaxy catalogs. We find a statistically significant (p -value $\sim 10^{-4}$, accounting for look-elsewhere factors) cross-correlation between CHIME FRBs and galaxies in the redshift range $0.3 \lesssim z \lesssim 0.5$, in three photometric galaxy surveys: WISE \times SCOS, DESI-BGS, and DESI-LRG. The level of cross-correlation is consistent with an order-one fraction of the CHIME FRBs being in the same dark matter halos as survey galaxies in this redshift range. We find statistical evidence for a population of FRBs with large host dispersion measure (~ 400 pc cm $^{-3}$), and show that this can plausibly arise from gas

in large halos ($M \sim 10^{14} M_{\odot}$), for FRBs near the halo center ($r \lesssim 100$ kpc). These results will improve in future CHIME/FRB catalogs, with more FRBs and better angular resolution.

Keywords: Radio transient sources (2008), Large-scale structure of the universe (902), High energy astrophysics (739), Cosmology (343)

1. INTRODUCTION

Fast radio bursts (FRBs) are millisecond flashes of radio waves whose dispersion is beyond what we expect from Galactic models along the line of sight. The origin of FRBs is still a mystery, despite over a decade of observations and theoretical exploration (see, e.g. Cordes & Chatterjee 2019; Petroff et al. 2019; Platts et al. 2019). The Canadian Hydrogen Intensity Mapping Experiment / Fast Radio Burst Project (CHIME/FRB; CHIME/FRB Collaboration 2018) has recently released its first catalog of FRBs containing 492 unique sources (CHIME/FRB Collaboration 2021), increasing the number of known FRBs by a factor ~ 4 .¹ This unprecedented sample size is a new opportunity for statistical studies of FRBs.

The angular resolution of CHIME/FRB is not sufficient to associate FRBs with unique host galaxies, except for some FRBs at very low DM, for example a repeating CHIME FRB associated with M81 (Bhardwaj et al. 2021). This appears to put some science questions out of reach, such as determining the redshift distribution of CHIME FRBs.

However, with large enough catalogs of both FRBs and galaxies, it is possible to associate FRBs with galaxies statistically, using angular cross-correlations. Intuitively, if the angular resolution θ_f of an FRB experiment is too large for unique host galaxy associations, there will still be an excess probability (relative to a random point on the sky) to observe FRBs within distance $\sim \theta_f$ of a galaxy. Formally, this corresponds to a cross-correlation between the FRB and galaxy catalogs, which we will define precisely in §3. By measuring the correlation as a function of galaxy redshift and FRB dispersion measure (DM) (defined below), the redshift distribution and related properties of the FRB population can be constrained, even in the absence of per-object associations.

FRB-galaxy cross-correlations have been proposed in a forecasting context (McQuinn 2014; Masui & Sigurdsson 2015; Shirasaki et al. 2017; Madhavacheril et al. 2019; Rafiei-Ravandi et al. 2020; Reischke et al. 2021a;

Alonso 2021; Reischke et al. 2021b), and applied to the ASKAP and 2MPZ/HIPASS catalogs by Li et al. (2019). In this paper, we will use machinery developed by Rafiei-Ravandi et al. (2020) for modeling the FRB-galaxy cross-correlation, and disentangling it from propagation effects. This machinery uses the halo model for cosmological large-scale structure (LSS); for a review see Cooray & Sheth (2002).

Before summarizing the main results presented here, we recall the definition of FRB DM. FRBs are dispersed: the arrival time at radio frequency ν is delayed, by an amount proportional to ν^{-2} . The dispersion is proportional to the DM, defined as the free electron column density along the line of sight:

$$\text{DM} \equiv \int n_e(x) dx. \quad (1)$$

Since FRBs have not been observed to have spectral lines, FRB redshifts are not directly observable. However, the DM is a rough proxy for redshift (Macquart et al. 2020). We write the total DM as the sum of contributions from our Galaxy and halo (DM_{gal}), the IGM (DM_{IGM}), and the FRB host galaxy and halo (DM_{host}):

$$\text{DM} = \text{DM}_{\text{gal}} + \text{DM}_{\text{IGM}}(z) + \text{DM}_{\text{host}}. \quad (2)$$

The IGM contribution $\text{DM}_{\text{IGM}}(z)$ is given by the Macquart relation:

$$\text{DM}_{\text{IGM}}(z) = n_{e,0} \int_0^z dz' f_d(z') \frac{1+z'}{H(z')}, \quad (3)$$

where $f_d(z)$ is the mean electron ionization fraction at redshift z , $n_{e,0} = 2.13 \times 10^{-7} \text{ cm}^{-3}$ is the comoving electron density, and $H(z)$ is the Hubble expansion rate. If f_d is assumed independent of redshift, then Eq. (3) has the following useful approximation:

$$\text{DM}_{\text{IGM}}(z) \approx (1000 \text{ pc cm}^{-3}) f_d z. \quad (4)$$

We checked that this approximation is accurate to 6% for $z \leq 3$, assuming that helium reionization is complete by $z = 3$. By default, we assume $f_d = 0.9$, which implies $\text{DM}_{\text{IGM}}(z) \approx 900z \text{ pc cm}^{-3}$.

We briefly summarize the main results of the paper. We find a statistically significant correlation between CHIME FRBs and galaxies in the redshift range

¹ For a complete list of known FRBs, see <https://www.herta-experiment.org/frbstats> (Spanakis-Misirlis 2021) or the Transient Name Server (TNS, Petroff & Yaron 2020).

$0.3 \lesssim z \lesssim 0.5$. The correlation is seen in three photometric galaxy surveys: WISE \times SCOS, DESI-BGS, and DESI-LRG (described in §2.2). The statistical significance of the detection in each survey is $p \sim (2.7 \times 10^{-5})$, (3.1×10^{-4}) , and (4.1×10^{-4}) , respectively. These p -values account for look-elsewhere effects, in both angular scale and redshift range. The observed level of correlation is consistent with an order-one fraction of CHIME FRBs inhabiting the same dark matter halos as galaxies in these surveys. CHIME/FRB does not resolve halos, so we cannot distinguish between FRBs in survey galaxies and FRBs in the same halos as survey galaxies.

We study the DM dependence of the FRB-galaxy correlation and find a correlation between high-DM (extragalactic DM $\geq 785 \text{ pc cm}^{-3}$) FRBs and galaxies at $z \sim 0.4$. This implies the existence of an FRB subpopulation with host DM $\gtrsim 400 \text{ pc cm}^{-3}$. Such large host DMs have not yet been seen in observations that directly associate FRBs with host galaxies. To date, 14 FRBs (excluding a Galactic magnetar, see CHIME/FRB Collaboration (2020a); Bochenek et al. (2020)) have been localized to host galaxies, all of which have $\text{DM}_{\text{host}} \lesssim 200 \text{ pc cm}^{-3}$. In §4.2, we explain why these observations are not in conflict. We also show that host DMs $\gtrsim 400 \text{ pc cm}^{-3}$ can arise from ionized gas in large ($M \gtrsim 10^{14} M_{\odot}$) dark matter halos, if FRBs are located near the halo center ($r \lesssim 100 \text{ kpc}$).

This paper is structured as follows. In §2, we describe the observations and data reduction. Clustering results are presented in §3 and interpreted in §4. We conclude in §5. Throughout, we adopt a flat Λ CDM cosmology with Hubble expansion rate $h = 0.67$, matter abundance $\Omega_m = 0.315$, baryon abundance $\Omega_b = 0.048$, initial power spectrum amplitude $A_s = 2.10 \times 10^{-9}$, spectral index $n_s = 0.965$, neutrino mass $\sum_{\nu} m_{\nu} = 0.06 \text{ eV}$, and CMB temperature $T_{\text{CMB}} = 2.726 \text{ K}$. These parameters are consistent with Planck results (Aghanim et al. 2020).

2. DATA

2.1. FRB catalog

The first CHIME/FRB catalog is described in (CHIME/FRB Collaboration 2021). In order to maximize localization precision and to simplify selection biases, we include only a single burst with the highest significance for each repeating FRB in this analysis. This treats repeating and nonrepeating FRBs as a single population. In future CHIME/FRB catalogs with more repeaters, it would be interesting to analyze the two populations separately. In CHIME/FRB, there is currently no evidence that repeaters and nonrepeaters have different sky distributions (CHIME/FRB Collaboration 2021). We

also exclude three sidelobe detections (FRB20190210D, FRB20190125B, FRB20190202B), leaving a sample of 489 unique sources. We do not exclude FRBs with `excluded_flag=1`, indicating an epoch of low sensitivity, since we expect the localization accuracy of such FRBs to be similar to the main catalog.

Throughout this paper, *all DM values are extragalactic*. That is, before further processing of the CHIME FRBs, we subtract the Galactic contribution DM_{gal} from the observed DM. The value of DM_{gal} is estimated using the YMW16 (Yao et al. 2017) model. In §4.2, we show that using the NE2001 (Cordes & Lazio 2002) model does not affect results qualitatively. The CHIME/FRB extragalactic DM distribution is shown in Figure 1.

We do not subtract an estimate of the Milky Way halo DM, since the halo DM is currently poorly constrained by observations. The range of allowed values is roughly $10 \lesssim \text{DM}_{\text{halo}} \lesssim 100 \text{ pc cm}^{-3}$, and the (dipole-dominated) anisotropy is expected to be small (Prochaska & Zheng 2019; Keating & Pen 2020). The results of this paper are qualitatively unaffected by the value of DM_{halo} .

The CHIME/FRB pipeline assigns a nominal sky location to each FRB based on the observed signal-to-noise ratio (SNR) in each of 1024 formed beams. In the simplest case of an FRB that is detected only in a single formed beam, the nominal location is the center of the formed beam. For multibeam detections, the nominal location is roughly a weighted average of the beam centers (CHIME/FRB Collaboration 2019, 2021). Statistical errors on CHIME/FRB locations are difficult to model, since they depend on both the details of the CHIME telescope and selection biases that depend on the underlying FRB population. We discuss this further in §3.1 and Appendix A.

2.2. Galaxy catalogs

On the galaxy side, we have chosen five photometric redshift catalogs: 2MPZ, WISE \times SCOS, DESI-BGS, DESI-LRG, and DESI-ELG. Note that the DESI catalogs are the photometric target samples for forthcoming spectroscopic DESI surveys with the same names. Table 1 summarizes key properties of our reduced samples for the cross-correlation analysis, and the redshift distributions are shown in Figure 1.

The 2MASS Photometric Redshift (2MPZ) catalog (Bilicki et al. 2013) contains ~ 1 million galaxies with $z \lesssim 0.3$ (redshift error $\sigma_z \sim 0.02$), enabling the construction of a 3D view of LSS at low redshifts (see, e.g. Alonso et al. 2015; Balaguera-Antolínez et al. 2018). In this work, we use the mask made by Alonso et al. (2015)

Survey	f_{sky}	$[z_{\min}, z_{\max}]$	z_{med}	N_{gal}	N_{FRB}
2MPZ	0.647	[0.0, 0.3]	0.08	670,442	323
WISE×SCOS	0.638	[0.0, 0.5]	0.16	6,931,441	310
DESI-BGS	0.118	[0.05, 0.4]	0.22	5,304,153	183
DESI-LRG	0.118	[0.3, 1.0]	0.69	2,331,043	183
DESI-ELG	0.055	[0.6, 1.4]	1.09	5,314,194	62
BGS+LRG	0.118	[0.05, 1.0]	0.28	7,690,819	183

Table 1. Galaxy survey parameters: sky fraction f_{sky} (not accounting for CHIME/FRB coverage), redshift range $[z_{\min}, z_{\max}]$, median redshift z_{med} , total number of unmasked galaxies N_{gal} , and number of FRBs N_{FRB} overlapping the survey. The “BGS+LRG” catalog is used only in §3.5, and consists of all unique objects from the DESI-BGS and DESI-LRG catalogs.

for the 2MPZ catalog. Following Bilicki et al. (2013), we discard galaxies whose K_s -band magnitude is below the completeness limit $m_{K_s} = 13.9$.

The WISE×SuperCOSMOS photometric redshift catalog (WISE×SCOS, Bilicki et al. 2016) contains ~ 20 million point sources with $z \lesssim 0.5$ ($\sigma_z \sim 0.03$) over 70% of the sky, making it a versatile dataset for cross-correlation studies. In this work, we use a slightly modified catalog (Krakowski et al. 2016), which includes probabilities ($p_{\text{gal}}, p_{\text{star}}, p_{\text{qso}}$) for each object to be a galaxy, star, or quasar, respectively. We use objects with $p_{\text{gal}} \geq 0.9$, which is consistent with the weighted mean purity of identified galaxies across the $W1$ band (Krakowski et al. 2016). We use a standard mask² to remove the Galactic foreground, Magellanic Clouds and bright stars. Additionally, we mask out regions that are contaminated visually owing to their proximity to the Galactic plane:

$$\begin{aligned} &(|b| \leq 20^\circ) \text{ and } ((0^\circ \leq l \leq 30^\circ) \text{ or } (330^\circ \leq l \leq 360^\circ)), \\ &(|b| \leq 18^\circ) \text{ and } ((30^\circ \leq l \leq 60^\circ) \text{ or } (300^\circ \leq l \leq 330^\circ)), \\ &(|b| \leq 17^\circ) \text{ and } (0^\circ \leq l \leq 360^\circ). \end{aligned} \quad (5)$$

The Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys (Dey et al. 2019) were designed to identify galaxies for spectroscopic follow-up. We use the catalogs from the DR8 release, with photometric redshifts from Zhou et al. (2020a). Following DESI, we consider three samples: the Bright Galaxy Survey (BGS), the Luminous Red Galaxy (LRG) sample, and the Emission Line Galaxy (ELG) sample, corresponding to redshift ranges $0.05 \leq z \leq 0.4$, $0.3 \leq z \leq 1$, and $0.6 \leq z \leq 1.4$ respectively (Figure 1).

For each of the three DESI samples, we define survey geometry cuts as follows. For simplicity, we restrict to the northern part of the survey ($\text{Dec} > 32^\circ 375$, $b \geq +17^\circ$), which contains ~ 2 times as many CHIME/FRB sources as the southern part. Note that the northern and southern DESI surveys are obtained from different telescopes and may have different systematics. For the DESI-ELG sample, we impose the additional constraint $b \geq +45^\circ$ in order to mitigate systematic depth variations. We restrict to sky regions that were observed at least twice in each of the $\{g, r, z\}$ bands (Zhou et al. 2020a). We mask bad pixels, bright stars, large galaxies, and globular clusters using the appropriate DESI bitmask.³

In addition to these geometric cuts, we impose per-object cuts on the DESI catalogs by removing point-like objects (TYPE=PSF), and applying the appropriate color cuts for each of the three surveys. Color cuts for the BGS, LRG, and ELG catalogs are defined by Ruiz-Macias et al. (2020), Zhou et al. (2020b), and Raichoor et al. (2020) respectively. For BGS, we include both “faint” ($19.5 < r < 20$) and “bright” ($r < 19.5$) galaxies (terminology from Ruiz-Macias et al. 2020). For BGS and LRG, we exclude objects with poorly constrained photometric redshifts ($z_{\text{phot, std}} > 0.08$). Our final BGS, LRG and ELG samples have typical redshift error $\sigma_z \sim 0.03, 0.04$, and 0.15 respectively.

3. FRB-GALAXY CORRELATION RESULTS

In this section, we describe our pipeline for computing the FRB-galaxy cross power spectrum. The pipeline consists of mapping sources onto a sky grid and then computing the spherical harmonic transform and the angular power spectrum. Error bars are assigned using mock FRB catalogs.

3.1. Pipeline overview

Our central statistic is the angular power spectrum C_ℓ^{fg} , a Fourier-space statistic that measures the level of correlation between the FRB catalog f and galaxy catalog g , as a function of angular wavenumber ℓ . Formally, C_ℓ^{fg} is defined by

$$\langle a_{\ell m}^f a_{\ell' m'}^{g*} \rangle = C_\ell^{fg} \delta_{\ell \ell'} \delta_{mm'}, \quad (6)$$

where $a_{\ell m}^Y$ is the spherical harmonic transform of catalog $Y \in \{f, g\}$ (the all-sky analog of the Fourier transform on the flat sky). Intuitively, a detection of nonzero C_ℓ^{fg}

² <http://ssa.roe.ac.uk/WISeSCOS.html>

³ MASKBITS 1, 5–9, and 11–13, defined here: <https://www.legacysurvey.org/dr8/bitmasks/>

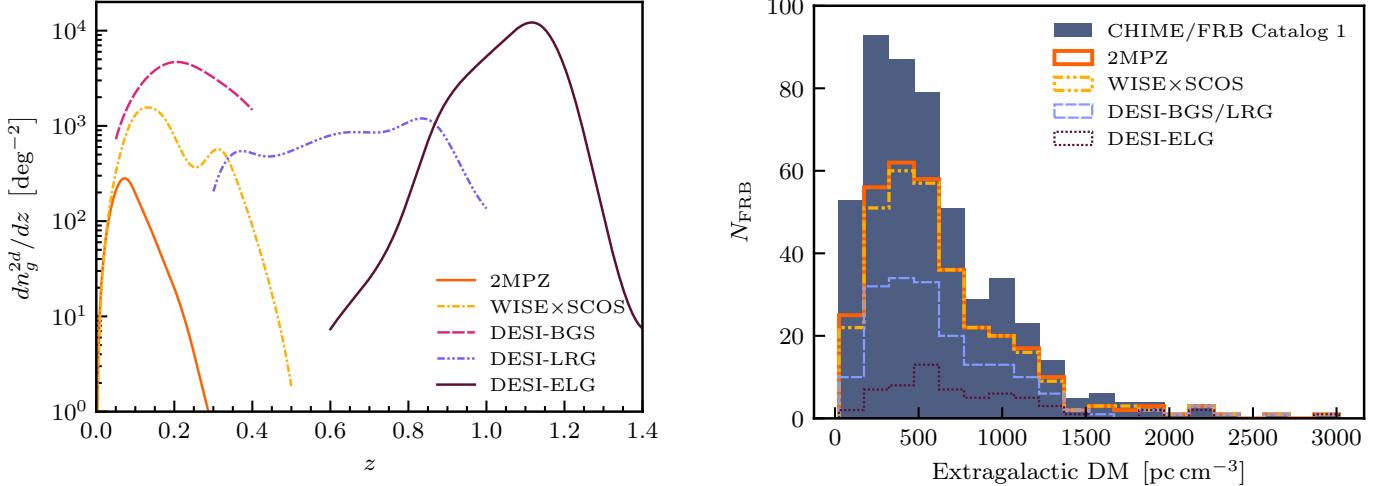


Figure 1. *Left panel:* Redshift distributions for the five galaxy samples in this paper (§2.2). *Right panel:* FRB extragalactic DM distributions for the CHIME/FRB catalog (solid) and for the subset of the CHIME/FRB catalog that overlaps spatially with each galaxy survey.

at wavenumber ℓ corresponds to a pixel-space angular correlation at separation $\theta \sim \ell^{-1}$.

The power spectrum C_ℓ^{fg} is not the only way of representing a cross-correlation between catalogs as a function of scale. Another possibility is the correlation function $\zeta(\theta)$, obtained by counting pairs of objects whose angular separation θ lies in a set of nonoverlapping bins. This method was used by Li et al. (2019) to correlate ASKAP FRBs with nearby galaxies. The power spectrum C_ℓ^{fg} and correlation function $\zeta(\theta)$ are related to each other by the Legendre transform $\zeta(\theta) = \sum_\ell (2\ell + 1)/(4\pi) C_\ell^{fg} P_\ell(\cos \theta)$. Therefore, C_ℓ^{fg} and $\zeta(\theta)$ contain the same information, and the choice of which one to use is a matter of convenience. We have used the power spectrum C_ℓ^{fg} , since it has the property that nonoverlapping ℓ -bins are nearly uncorrelated, making it straightforward to infer statistical significance from plots.

Throughout the paper, it will be useful to have a model FRB-galaxy power spectrum C_ℓ^{fg} in mind. In Figure 2, we show C_ℓ^{fg} for a galaxy population at $z \sim 0.4$, calculated using the “high- z ” FRB model from Rafie-Ravandi et al. (2020), with median FRB redshift $z = 0.76$. The main features of C_ℓ^{fg} are as follows:

- The leftmost peak at $\ell \sim 10^2$ is the *two-halo term* $C_\ell^{fg(2h)}$, which arises from FRBs and galaxies in different halos. The two-halo term does not probe the details of FRB-galaxy associations; it arises because FRBs and galaxies both inhabit halos, and halos are clustered on ~ 100 Mpc scales (the correlation length of the cosmological density field).

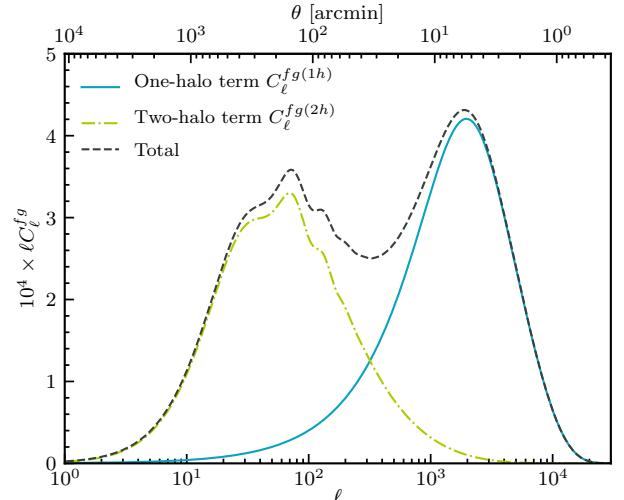


Figure 2. Model FRB-galaxy power spectrum C_ℓ^{fg} from §3.1, for a galaxy population near $z \sim 0.37$ and FRB angular resolution $1'$. Note that we have plotted (ℓC_ℓ^{fg}) , for consistency with later plots in the paper. In this and later plots in the paper, the angular scale on the top axis is $\theta = \pi/\ell$, and is intended to provide an intuitive mapping between angular multipole ℓ and an angular scale.

- The rightmost peak at $\ell \sim 10^3$ is the *one-halo term* $C_\ell^{fg(1h)}$, which is sourced by (FRB, galaxy) pairs in the same dark matter halo.
- For completeness, we note that for $\ell \gtrsim 10^4$, there is a “Poisson” term (not shown in Figure 2) that is sourced by FRBs in catalog galaxies (not elsewhere in the halo). CHIME/FRB’s limited angular resolution suppresses C_ℓ^{fg} at high ℓ , hiding the Poisson term. Intuitively, this is because CHIME/FRB cannot resolve different galaxies in the same dark matter halo.

Although the one-halo and two-halo terms look comparable in Figure 2, the SNR of the one-halo term is a few times larger. In this paper, we do not detect the two-halo term with statistical significance (see Figure 7). Therefore, throughout the paper we will often neglect the two-halo term, and make the approximation $C_\ell^{fg} \approx C_\ell^{fg(1h)}$.

The one-halo term $C_\ell^{fg(1h)}$ is constant in ℓ for $\ell \lesssim 10^3$, and suppressed for $\ell \gtrsim 10^3$. (Note that in Figure 2, we have plotted ℓC_ℓ^{fg} , for consistency with later figures in the paper.) The high- ℓ suppression arises from two effects: (1) statistical errors on FRB positions (the CHIME/FRB “beam”), and (2) displacements between FRBs and galaxies in the same dark matter halo.

Within the statistical errors of the C_ℓ^{fg} measurement in this paper, both effects can be modeled as Gaussian, i.e. high- ℓ suppression of the form $e^{-\ell^2/L^2}$:

$$C_\ell^{fg(1h)} = \alpha e^{-\ell^2/L^2}, \quad (7)$$

where we have omitted the two-halo term since we do not detect it with statistical significance. In principle, the value of L in Eq. (7) is computable, given models for statistical errors on CHIME FRB sky locations and FRB/galaxy profiles within dark matter halos. However, FRB halo profiles are currently poorly constrained, and CHIME FRB location errors are difficult to model, since they depend on both instrumental selection effects and details of the FRB population. In Appendix A, we explore modeling issues in detail and show that a plausible (but conservative, i.e. wide) range of L -values is $315 \leq L \leq 1396$.

Summarizing the above discussion, our pipeline works as follows. We measure the angular power spectrum C_ℓ^{fg} from the FRB and galaxy catalogs, and fit the ℓ -dependence to the template form $C_\ell^{fg} = \alpha e^{-\ell^2/L^2}$ in Eq. (7). We treat the amplitude α as a free parameter, and vary the template scale L over the range $315 \leq L \leq 1396$, to evaluate the correlation amplitude as a function of scale.

3.2. Overdensity maps

Turning now to implementation, the first step in our pipeline is to convert the FRB and galaxy catalogs into “overdensity” maps $\delta_f(\mathbf{x}), \delta_g(\mathbf{x})$, defined by

$$\delta_Y(\mathbf{x}) = \frac{1}{n_Y^{2d} \Omega_{\text{pix}}} \left(N_{Y \in \mathbf{x}} - \bar{N}_{Y \in \mathbf{x}} \right). \quad (8)$$

Here, $Y \in \{f, g\}$ denotes a catalog, \mathbf{x} denotes an angular pixel, $N_{Y \in \mathbf{x}}$ denotes the number of catalog objects in pixel \mathbf{x} , and $\bar{N}_{Y \in \mathbf{x}}$ denotes the expected number of catalog objects in pixel \mathbf{x} due to the survey geometry. The prefactor $1/(n_Y^{2d} \Omega_{\text{pix}})$ is conventional, where n_Y^{2d} is

the 2D number density and Ω_{pix} is the pixel area. For CHIME/FRB, the expected number density $\bar{N}_{f \in \mathbf{x}}$ depends on declination (Dec). The definition (8) of $\delta_f(\mathbf{x})$ weights each pixel \mathbf{x} proportionally to the expected number of FRBs. This weighting is optimal since the FRB field is Poisson noise dominated ($C_\ell^{ff} \approx 1/n_f^{2d}$).

The difference between a density map and an overdensity map is the second term \bar{N} in Eq. (8), which removes spurious density fluctuations due to the survey geometry. We compute the \bar{N} -term differently for different catalogs as follows.

For the three DESI catalogs, we estimate \bar{N} using “randoms” from the DESI-DR8 release, i.e. simulated catalogs that encode the survey geometry, with no spatial correlations between objects. We use random catalogs from the DESI-DR8 data release (source density $n_g^{2d} = 5000 \text{ deg}^{-2}$), and apply the DESI “geometry” cuts from the previous section.

For the other two galaxy surveys (2MPZ and WISE×SCOS), random catalogs are not readily available, so we represent the survey geometry by an angular HEALPix (Gorski et al. 2005) mask, and assume uniform galaxy density outside the mask:

$$\bar{N}_{g \in \mathbf{x}} = \begin{cases} n_g^{2d} \Omega_{\text{pix}} & \text{if } \mathbf{x} \text{ is unmasked} \\ 0 & \text{if } \mathbf{x} \text{ is masked} \end{cases} \quad (9)$$

The mask geometries for 2MPZ and WISE×SCOS were described previously in §2.2.

Finally, for the CHIME/FRB catalog, computing \bar{N} deserves some discussion. The CHIME/FRB number density \bar{N} is inhomogeneous, peaking near the north celestial pole. To an excellent approximation, the number density is azimuthally symmetric in equatorial coordinates, i.e. independent of right ascension (RA) at fixed declination, because CHIME is a cylindrical drift-scan telescope oriented north-south (CHIME/FRB Collaboration 2021). Therefore, we make random FRB catalogs that represent \bar{N} by randomizing RAs of the FRBs in the observed catalog, leaving declinations fixed. When making randoms, we also loop over 1000 copies of the CHIME/FRB catalog, so that the random catalogs are much larger than the data catalog (appropriately rescaling \bar{N} and n_f^{2d} in Eq. 8).

In Figure 3, we show overdensity maps $\delta_Y(\mathbf{x})$ for the CHIME/FRB sources and the galaxies. These maps are useful as visual checks for systematic effects, before catalogs are cross-correlated. For example, if the Galactic mask is not conservative enough, the overdensity map may show visual artifacts with $\delta_g < 0$, since Galactic extinction will suppress the observed catalog density N , relative to \bar{N} . No visual red flags are seen in either the CHIME/FRB or galaxy maps, even without a Galactic

mask for CHIME/FRB. This is consistent with [Josephy et al. \(2021\)](#), who found no evidence for Galactic latitude dependence in the CHIME/FRB number density after correcting for selection effects. As described in §2.2, we do apply a Galactic mask in our pipeline, so even if the FRB catalog does contain low-level biases in the Galactic plane, they should be mitigated.

3.3. Estimating the power spectrum C_ℓ^{fg}

We estimate C_ℓ^{fg} in our pipeline by taking spherical transforms of the overdensity maps $\delta_f(\mathbf{x}), \delta_g(\mathbf{x})$, to get spherical harmonic coefficients $a_{\ell m}^f$ and $a_{\ell m}^g$. Then, we estimate the power spectrum C_ℓ^{fg} as

$$\hat{C}_\ell^{fg} = \frac{1}{f_{\text{sky}}^{fg}} \sum_{m=-\ell}^{\ell} \frac{1}{2\ell+1} a_{\ell m}^{f*} a_{\ell m}^g, \quad (10)$$

where f_{sky}^{fg} is the fractional sky area subtended by the intersection of the FRB and galaxy surveys. The f_{sky}^{fg} prefactor normalizes the power spectrum estimator to have the correct normalization on the partial sky. Throughout the main analysis, we represent overdensities as HEALPix maps with 1° resolution ($N_{\text{side}} = 2048$), and estimate the power spectrum to a maximum multipole of $\ell_{\text{max}} = 2000$, corresponding to angular scale $\theta = \pi/\ell_{\text{max}} = 5.4'$.

We assign error bars to the power spectrum C_ℓ^{fg} using Monte Carlo techniques, simulating mock FRB catalogs and cross-correlating them with the real galaxy catalogs. We simulate mock FRB catalogs by keeping FRB declinations the same as in the real catalog, but randomizing right ascensions. This mimics the logic used to construct random FRB catalogs in §3.2. In fact, the only difference in our pipeline between a “mock” and a “random” FRB catalog is the number of FRBs: a mock catalog has the same number of FRBs as the data, whereas a random catalog has a much larger number. Conceptually, there is another difference between mocks and randoms: mocks should include any spatial clustering signal present in the real data, whereas randoms are unclustered and only represent the survey geometry. For FRBs, spatial clustering is small compared to Poisson noise ($C_\ell^{ff} \approx 1/n_f^{2d}$, see Figure 5), so we can make the approximation that clustering is negligible.

3.4. Statistical significance and *look-elsewhere effect*

In Figure 4, we show the angular power spectrum C_ℓ^{fg} for a set of nonoverlapping ℓ bins. A weak positive FRB-galaxy correlation is seen at $500 \lesssim \ell \lesssim 1000$ in some of the galaxy surveys. In this subsection, we will address the question of whether this correlation is statistically significant.

As explained in §3.1, we will fit the FRB-galaxy correlation to the template $C_\ell^{fg} = \alpha e^{-\ell^2/L^2}$, treating the amplitude α as a free parameter, and varying the template scale L over the range $315 \leq L \leq 1396$. Let us temporarily assume that L is known in advance. In this case, an optimal estimator for α is

$$\hat{\alpha}_L = \frac{1}{\mathcal{N}_L} \sum_{\ell \geq \ell_{\min}} (2\ell+1) \frac{e^{-\ell^2/L^2}}{C_\ell^{gg}} \hat{C}_\ell^{fg}, \quad (11)$$

where \hat{C}_ℓ^{fg} was defined in Eq. (10), and the normalization \mathcal{N}_L is defined by

$$\mathcal{N}_L = \sum_{\ell \geq \ell_{\min}} (2\ell+1) \frac{e^{-2\ell^2/L^2}}{C_\ell^{gg}}. \quad (12)$$

We have included a cutoff at $\ell_{\min} = 50$ to mitigate possible large-scale systematics. This is a conservative choice, since Figure 5 does not show evidence for systematic power in the auto power spectrum C_ℓ^{ff} for $\ell \gtrsim 15$. Eq. (11) is derived by noting that

$$\begin{aligned} \text{Var}(\hat{C}_\ell^{fg}) &\propto \frac{C_\ell^{ff} C_\ell^{gg} + (C_\ell^{fg})^2}{2\ell+1} \\ &\propto \frac{C_\ell^{gg}}{2\ell+1}, \end{aligned} \quad (13)$$

where the first line follows from Wick’s theorem, and the second line follows since $(C_\ell^{fg})^2 \ll C_\ell^{ff} C_\ell^{gg}$, and C_ℓ^{ff} is nearly constant in ℓ .

We define the quantity

$$\text{SNR}_L = \frac{\hat{\alpha}_L}{\text{Var}(\hat{\alpha}_L)^{1/2}}, \quad (14)$$

which is the statistical significance of the C_ℓ^{fg} detection in “sigmas”, for a fixed choice of L . In Figure 6, we show the quantity SNR_L , as a function of scale L .

We pause for a notational comment: throughout the paper, ℓ denotes a multipole (as in C_ℓ^{fg}), and L denotes the template scale defined in Eq (7). The value of SNR_L (or $\hat{\alpha}_L$) is obtained by summing C_ℓ^{fg} over $\ell \lesssim L$, as in Eq. (11). When C_ℓ^{fg} is computed as a function of ℓ (Figure 4), neighboring ℓ bins are nearly uncorrelated, whereas when SNR_L is computed as a function of L (Figure 6), nearby L -values are highly correlated.

In Figure 6, it is seen that SNR_L can be as large as 2.67, for a certain choice of L and galaxy survey (namely DESI-BGS at $L = 1396$). However, it would be incorrect to interpret this as a 2.67σ detection, since the value of L has been cherry-picked to maximize the signal.

To quantify statistical significance in a way that accounts for the choice of L (the “look-elsewhere effect”), we restrict the search to $315 \leq L \leq 1396$ and define

$$\text{SNR}_{\text{max}} = \max_{315 \leq L \leq 1396} \text{SNR}_L. \quad (15)$$

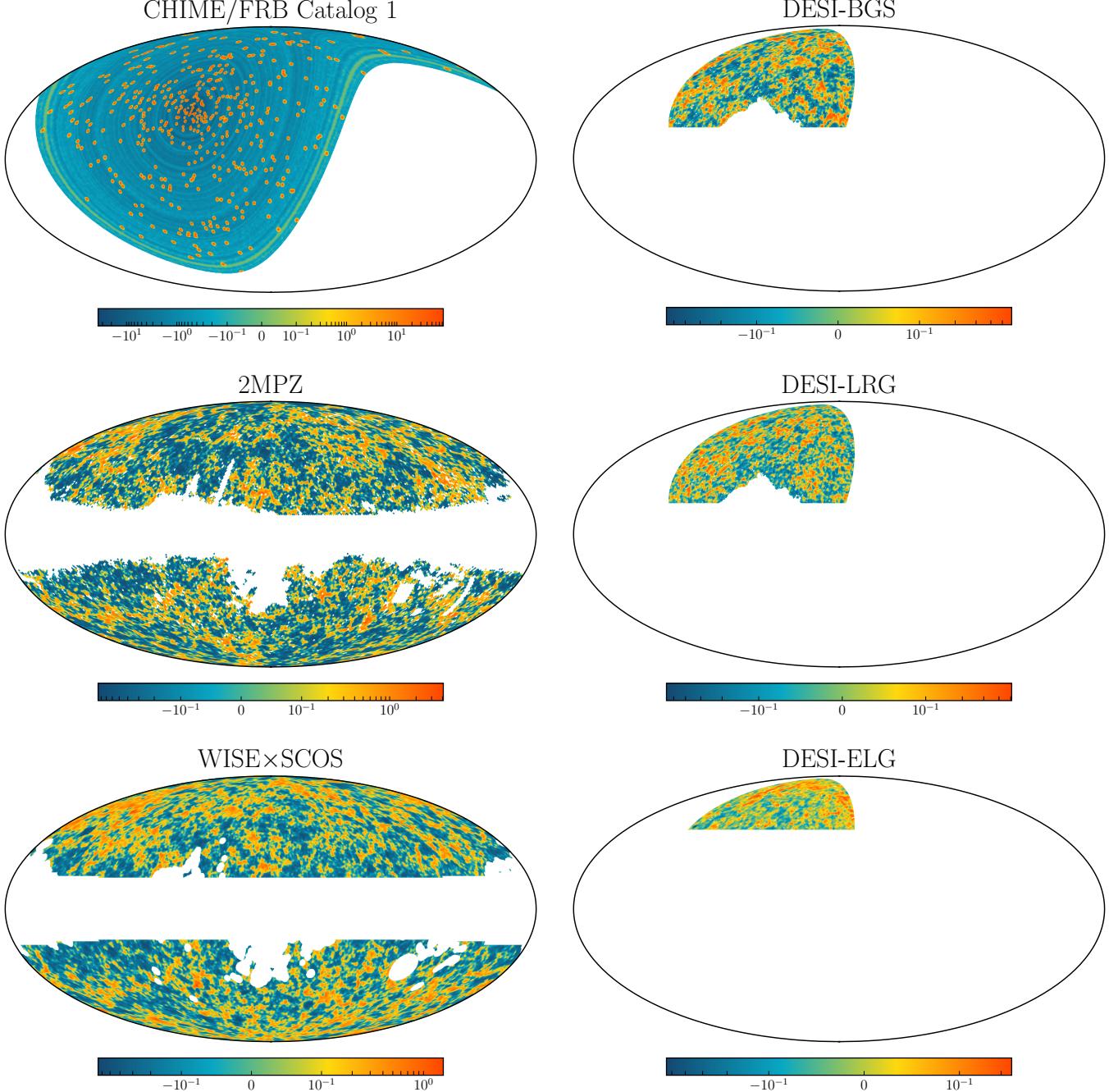


Figure 3. CHIME/FRB overdensity map $\delta_f(\mathbf{x})$, and galaxy overdensity maps $\delta_g(\mathbf{x})$ for each galaxy survey. Maps are shown in Mollweide projection, centered on $l = 180^\circ$ in the Galactic coordinate system, after applying the angular masks used in the analysis pipeline. To interpret the color scale, note that by Eq. (8), each object in a pixel contributes $1/(n_Y^{2d}\Omega_{\text{pix}})$ to the overdensity δ_Y .

For fixed L , SNR_L is approximately Gaussian distributed, and represents statistical significance in “sigmas”. Since SNR_{\max} is obtained by maximizing over trial L -values, SNR_{\max} is non-Gaussian, and we assign statistical significance by Monte Carlo inference.

In more detail, we compare the “data” value of SNR_{\max} (e.g. $\text{SNR}_{\max} = 2.67$ for DESI-BGS) to an ensemble of Monte Carlo simulations, obtained by cross-

correlating mock FRB catalogs with the real galaxy catalog as in §3.3. We assign a p -value by computing the fraction of mocks with $\text{SNR}_{\max}^{(\text{mock})} \geq \text{SNR}_{\max}^{(\text{data})}$. We find $p = 0.0166$ for DESI-BGS, i.e. evidence for a correlation at 98.34% CL after accounting for the look-elsewhere effect in L . The p -values for the other galaxy surveys are shown in Figure 6.

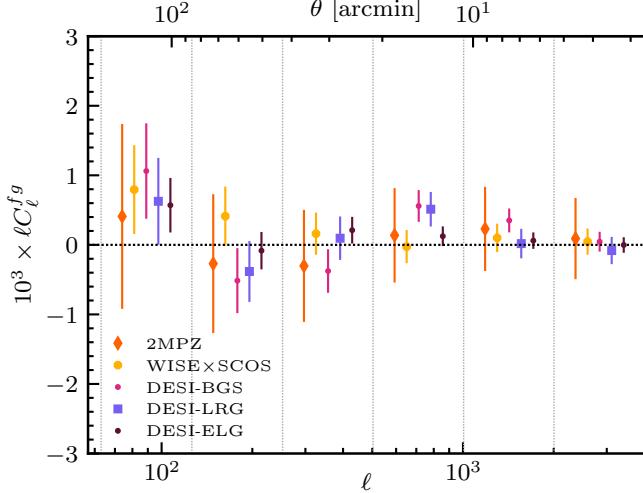


Figure 4. FRB-galaxy cross power spectrum C_ℓ^{fg} in a set of nonoverlapping ℓ bins delimited by vertical lines, with 1σ error bars. Data points are shifted slightly from the center of corresponding ℓ bins for visual clarity. Here, we have used all galaxies in the catalogs; if we restrict the redshift ranges, then the correlation is more significant (Figure 7).

Our interpretation is that this level of evidence is intriguing, but not high enough to be conclusive. Therefore, we do not interpret the FRB-galaxy correlation in Figures 4 and 6 as a detection. However, in the next subsection we will restrict the redshift range of the galaxy catalog (accounting for the look-elsewhere effect in choice of redshift range) and find a high-significance detection.

3.5. Redshift dependence

To illustrate our method for studying redshift dependence, we will use the WISE \times SCOS galaxy catalog as a running example. Suppose we cross-correlate CHIME FRBs with WISE \times SCOS galaxies above some minimum redshift z_{\min} , where z_{\min} is a free parameter that will be varied. For each z_{\min} , we repeat the analysis of the previous subsection. The power spectrum $C_\ell^{fg}(z_{\min})$ and quantity $\text{SNR}_L(z_{\min})$ (defined in Eq. 14) are now functions of two parameters: z_{\min} and template scale L .

In the top panels of Figure 7, we show the power spectrum $C_\ell^{fg}(z_{\min})$ for the fixed choice of redshift $z_{\min} = 0.3125$, and $\text{SNR}_L(z_{\min})$ as a function of L and z_{\min} . For specific parameter choices, we see a large FRB-galaxy correlation, e.g. $\text{SNR}_L(z_{\min}) = 4.88$ at $L = 543$ and $z_{\min} = 0.3125$. As in the previous subsection, this would imply a 4.88σ cross-correlation for these cherry-picked values of (L, z_{\min}) , but does not account for the look-elsewhere effect in choosing these values.

To assign statistical significance in a way that accounts for the look-elsewhere effect, we use the same method as the previous subsection, except that we now

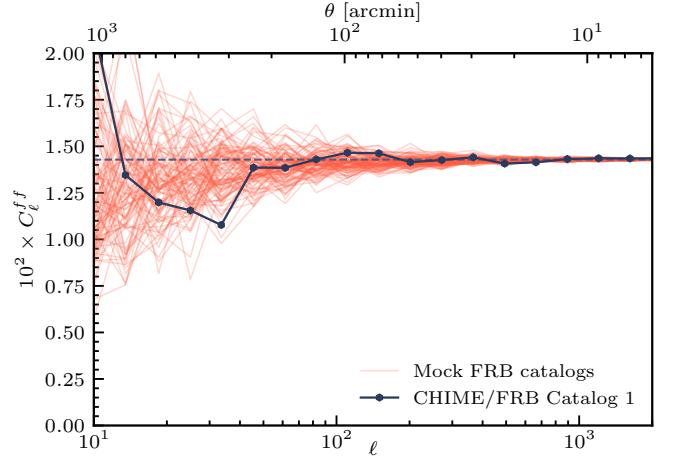


Figure 5. Angular auto power spectrum C_ℓ^{ff} for the CHIME/FRB catalog. Transparent lines represent 100 mock FRB catalogs that spatially model the real data. Throughout the analysis, we assume that the power spectrum C_ℓ^{ff} approaches a constant (dashed line) on small scales (high ℓ). Specifically, $C_\ell^{ff} \approx 1/n_f^{2d}$ for $315 \leq \ell \leq 1396$.

scan over two parameters (L, z_{\min}) rather than one (L) . Formally, we define

$$\text{SNR}_{\max} = \max_{0 \leq z_{\min} \leq 0.5} \max_{315 \leq L \leq 1396} \text{SNR}_L(z_{\min}), \quad (16)$$

analogously to Eq. (15) from the previous subsection. To assign bottom-line statistical significance, we would like to rank the “data” value $\text{SNR}_{\max} = 4.88$ within a histogram of SNR_{\max} values obtained by cross-correlating mock FRB catalogs with the galaxy catalog. However, with 10^4 simulations, we find that none of the mock catalogs actually exceed $\text{SNR}_{\max} = 4.88$, so we fit the tail of the SNR_{\max} distribution to an analytic distribution (a truncated Gaussian), and compute the p -value analytically. For details of the tail-fitting procedure, see Appendix C. We obtain **detection significance $p = 2.7 \times 10^{-5}$** for WISE \times SCOS with $z_{\min} = 0.3125$. This analysis “scans” over minimum redshift z_{\min} and scale L , and the significance fully accounts for the look-elsewhere effect in these parameters.

Similarly, we get $p = 3.1 \times 10^{-4}$ for DESI-BGS with $z_{\min} = 0.295$, scanning over z_{\min} and L . For DESI-LRG, we use a *maximum* redshift z_{\max} instead of a minimum redshift z_{\min} , since DESI-LRG is at higher redshift than WISE \times SCOS or DESI-BGS (Figure 1). Scanning over z_{\max} and scale L , we obtain $p = 4.1 \times 10^{-4}$ with $z_{\max} = 0.485$ for DESI-LRG. These results are shown in Figure 7.

Finally, we find borderline evidence $p = 0.0421$ ($L = 1396, z_{\max} = 0.86$) for a cross-correlation between DESI-ELG galaxies (varying z_{\max}) and CHIME FRBs with $\text{DM} \geq 500 \text{ pc cm}^{-3}$, where the choice of minimum DM

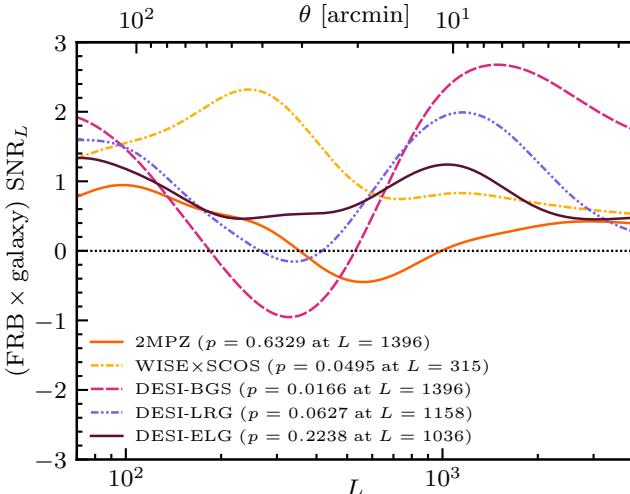


Figure 6. Quantity SNR_L , defined in Eq. (14), as a function of template scale L . As explained in §3.4, SNR_L is the statistical significance of the FRB-galaxy correlation in “sigmas”, for a fixed choice of L . The p -values in the legend are bottom-line detection significances after accounting for the look-elsewhere effect in L . Here, we have used all galaxies in the catalogs; if we restrict the redshift ranges, then the detection significance is higher (Figure 7).

is fixed. To justify this choice of DM_{min} , note that since host DMs must be positive, we do not expect a correlation between DESI-ELG galaxies ($z_{\text{min}} = 0.6$) and CHIME FRBs with $\text{DM} < 500 \text{ pc cm}^{-3}$ (allowing for statistical fluctuations in DM_{IGM} on the order of 40 pc cm^{-3}). We do not find any statistically significant detection with 2MPZ.

These results are consistent with a simple picture in which the FRB-galaxy correlation mainly comes from galaxies in redshift range $0.3 \lesssim z \lesssim 0.5$. For WISE×SCOS and DESI-BGS, the maximum survey redshifts are 0.5 and 0.4 respectively, and we find a strong detection when we impose a minimum redshift $z_{\text{min}} \sim 0.3$. For DESI-LRG, the minimum survey redshift is 0.3, and we find a strong detection when we impose a maximum redshift $z_{\text{max}} \sim 0.5$. The borderline detection in DESI-ELG and nondetection in 2MPZ are also consistent with this picture, in the sense that these catalogs do not overlap with the redshift range $0.3 \lesssim z \lesssim 0.5$.

As a direct way of seeing that the FRB-galaxy correlation is sourced by redshift range $0.3 \lesssim z \lesssim 0.5$, in Figure 8 we cross-correlate the FRB catalog with the combined BGS+LRG catalog (Table 1, bottom row) in nonoverlapping redshift bins with $0.05 \leq z \leq 1$. It is seen that the cross-correlation is driven by redshift range $0.3 \lesssim z \lesssim 0.5$. (The bin at $z \sim 0.75$ is nonzero at 2.2σ ,

which we interpret as borderline statistical significance, since there are 10 bins.)

In Appendix B, we examine the robustness of these results using null tests and do not find any evidence for systematic biases.

4. INTERPRETATION

So far, we have concentrated on establishing statistical significance of the FRB-galaxy correlation, in a Monte Carlo simulation pipeline that accounts for look-elsewhere effects. In this section, we will interpret the FRB-galaxy correlation, and explore implications for FRBs.

As explained in §3.1, the output of our pipeline is a constraint on the coefficient α in the template fit:

$$C_\ell^{fg} \approx C_\ell^{fg(1h)} \approx \alpha e^{-\ell^2/L^2}, \quad (17)$$

where the factor $e^{-\ell^2/L^2}$ is a Gaussian approximation to the high- ℓ suppression due to FRB/galaxy profiles and the instrumental beam.

At several points in this section, we will want to compare our FRB-galaxy correlation results to a model for $C_\ell^{fg(1h)}$. To do this, we interpret the low- ℓ limit of the model as a prediction for the coefficient α above. Formally, we define

$$\alpha \equiv \lim_{\ell \rightarrow 0} C_\ell^{fg(1h)} \quad (18)$$

and compare this model prediction for α to the value of $(\hat{\alpha}_L)_{L=1000}$, where the estimator $\hat{\alpha}_L$ was defined in Eq. (11). For simplicity we will fix $L = 1000$, since this gives a high-significance detection of the FRB-galaxy correlation in all three galaxy surveys (see Figure 7).

4.1. Link counting

In this subsection, we will interpret the amplitude of the FRB-galaxy correlation C_ℓ^{fg} in an intuitive way. First, we fix a galaxy catalog and redshift range. As a definition, we say that an FRB is *linked* to a galaxy if they are in the same dark matter halo. For each FRB f , we define the *link count* η_f by

$$\eta_f = \text{number of survey galaxies linked to FRB } f. \quad (19)$$

Given an FRB catalog, we define the *mean link count* η :

$$\eta = \langle \eta_f \rangle, \quad (20)$$

where the expectation value $\langle \cdot \rangle$ is taken over FRBs in the catalog.

To connect these definitions with our FRB-galaxy correlation results, we note that:

$$\alpha = \lim_{\ell \rightarrow 0} C_\ell^{fg(1h)} = \frac{\eta}{n_g^{2d}}, \quad (21)$$

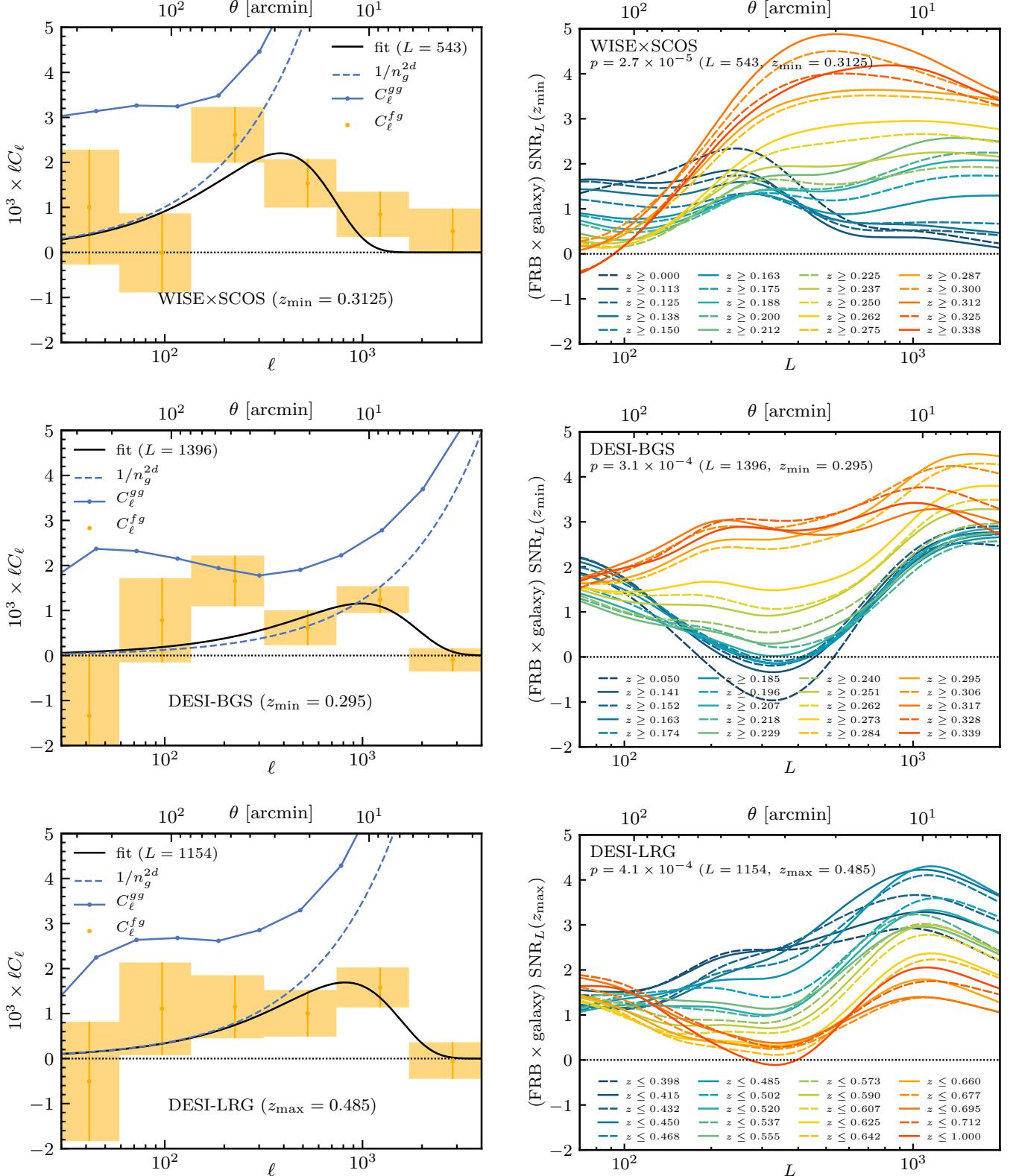


Figure 7. FRB-galaxy correlation analysis with two parameters: template scale L (defined in Eq. 7), and a redshift endpoint (either z_{\min} for WISE×SCOS and DESI-BGS, or z_{\max} for DESI-LRG). *Left column:* Angular cross power spectrum C_ℓ^{fg} and auto power spectrum C_ℓ^{gg} , for the fixed choice of redshift endpoint that maximizes FRB-galaxy correlation. The cross power “fit” is a best-fit template of the form $C_\ell^{fg} = \alpha e^{-\ell^2/L^2}$. *Right column:* Quantity SNR_L , defined in Eq. (14), as a function of L and redshift endpoint. As explained in §3.4, SNR_L is statistical significance of the FRB-galaxy correlation in “sigmas”, for a fixed choice of L and redshift endpoint. The p -values in the legend are bottom-line significance after accounting for the look-elsewhere effect in these choices (see §3.5).

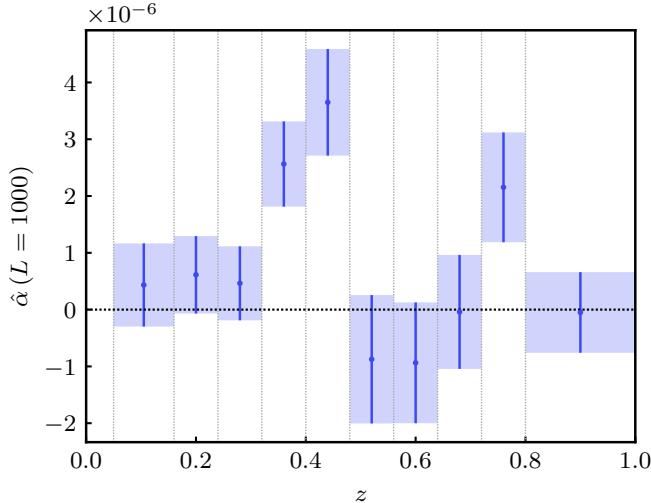


Figure 8. Redshift dependence of the FRB-galaxy correlation. We divide the BGS+LRG catalog into nonoverlapping redshift bins (dotted lines) and cross-correlate with CHIME FRBs. The quantity $(\hat{\alpha}_L)_{L=1000}$ on the y -axis is a measure of the level of cross-correlation, defined in Eq. (11).

where the first equality is Eq. (18), and the second equality follows from a short halo model calculation (Rafiei-Ravandi et al. 2020). That is, the amplitude α of the FRB-galaxy correlation (in the one-halo regime) is equivalent to a measurement of the mean link count η . This provides a more intuitive interpretation of the amplitude.

In each row of Table 2, we specify a choice of galaxy catalog and redshift range. The redshift ranges have been chosen to maximize C_ℓ^{fg} , as in §3.4. In the third column, we give the constraint on α obtained from the estimator $\hat{\alpha}_L$ at $L = 1000$. In the last column, we have translated this constraint of α to a constraint on η , using Eq. (21).

Taken together, the η measurements in Table 2 show that the CHIME/FRB catalog has mean link counts of order unity with galaxies in the range $0.3 \lesssim z \lesssim 0.5$. The precise value of η depends on the specific galaxy survey considered. Note that different galaxy surveys will have different values of η , since the number of galaxies per halo (and to some extent the population of halos that is sampled) will be different.

Since FRBs outside the redshift range of the galaxy catalog do not contribute to η , we write $\eta = p\tilde{\eta}$, where p is the probability that an FRB is in the catalog redshift range and $\tilde{\eta}$ is the mean link count of FRBs that are in the catalog redshift range.

For the galaxy surveys considered here, we expect $\tilde{\eta}$ to be of order unity, since dark matter halos rarely contain more than a few catalog galaxies. To justify this statement, we note that C_ℓ^{gg} is ~ 2 times larger than

the Poisson noise $1/n_g^{2d}$ in the one-halo regime (see Figure 7). By a link counting argument similar to Eq. (21), this implies that $\langle N_g^2 \rangle \sim 2\langle N_g \rangle$, where N_g is the number of galaxies in a halo, and the expectation values are taken over halos.

Since $\eta = p\tilde{\eta}$ is of order unity (by Table 2), and $\tilde{\eta}$ is of order unity (by the argument in the previous paragraph), we conclude that p is of order unity. That is, *an order-one fraction of CHIME FRBs are in the redshift range $0.3 \lesssim z \lesssim 0.5$.*

We have phrased this conclusion as a qualitative statement (“order-one fraction”) since it is difficult to assign a precise upper bound to $\tilde{\eta}$. More generally, it is difficult to infer the FRB redshift distribution (dn_f^{2d}/dz) from the FRB-galaxy correlation in the one-halo regime, since the level of correlation is proportional to $\tilde{\eta}(dn_f^{2d}/dz)$, with no obvious way of disentangling the two factors. Future CHIME/FRB catalogs should contain enough FRBs to detect the FRB-galaxy correlation on two-halo scales ($\ell \sim 100$) (Rafiei-Ravandi et al. 2020), which will help break the degeneracy and measure (dn_f^{2d}/dz) and $\tilde{\eta}$ separately.

4.2. DM dependence

In Figure 9, we divide the FRB catalog into extragalactic DM bins and explore the DM dependence of the FRB-galaxy cross-correlation.

A striking feature in Figure 9 is the nonzero correlation in the three highest-DM bins, corresponding to extragalactic DM $\geq 785 \text{ pc cm}^{-3}$.⁴ For reference, the last three bins represent 7%, 6%, and 15% of the CHIME/FRB catalog, respectively. At the redshift of the galaxy surveys ($z \sim 0.4$), the IGM contribution to the DM is $\text{DM}_{\text{IGM}}(z) \sim 360 \text{ pc cm}^{-3}$. Therefore, the observed FRB-galaxy correlation at $\text{DM} \geq 785 \text{ pc cm}^{-3}$ is evidence for a subpopulation of FRBs with host DMs of order $\text{DM}_{\text{host}} \sim 400 \text{ pc cm}^{-3}$.

This may appear to be in tension with recent direct associations between FRBs and host galaxies, which have typically been studied only for lower-redshift FRBs. At the time of this writing, 14 FRBs have been localized to host galaxies,⁵ all of which have $\text{DM}_{\text{host}} \lesssim 200 \text{ pc cm}^{-3}$ (Spitler et al. 2016; Bassa et al. 2017; Chatterjee et al. 2017; Kokubo et al. 2017; Tendulkar et al. 2017; Ban-

⁴ A technical comment here: for some DM bins in Figure 9, the large values of C_ℓ^{fg} lead to link counts η that are a few times larger than the link counts reported in Table 2 for the whole catalog, although statistical errors are large. However, the correlation coefficient between the FRB and galaxy fields is never larger than 1. In all cases, the field-level correlation $C_\ell^{fg}/(C_\ell^{ff} C_\ell^{gg})^{1/2}$ is of order 0.01 or smaller.

⁵ <https://frbhosts.org/#explore>

Survey	$[z_{\min}, z_{\max}]$	$(\alpha_L)_{L=1000}$	$n_g^{2d} [\text{sr}^{-1}]$	η
WISE×SCOS	$[0.3125, 0.5]$	$(4.35 \pm 0.97) \times 10^{-6}$	9.92×10^4	0.432 ± 0.096
DESI-BGS	$[0.295, 0.4]$	$(2.69 \pm 0.67) \times 10^{-6}$	7.94×10^5	2.13 ± 0.53
DESI-LRG	$[0.3, 0.485]$	$(3.94 \pm 0.93) \times 10^{-6}$	2.83×10^5	1.11 ± 0.26

Table 2. Clustering analysis in §4.1. The FRB-galaxy clustering statistic α_L (Eq. 11) can be translated to a constraint on η , the average number of survey galaxies in the same dark matter halo as a CHIME/FRB source (see text for details). The statistical error on η in each row is roughly proportional to n_g^{2d} .

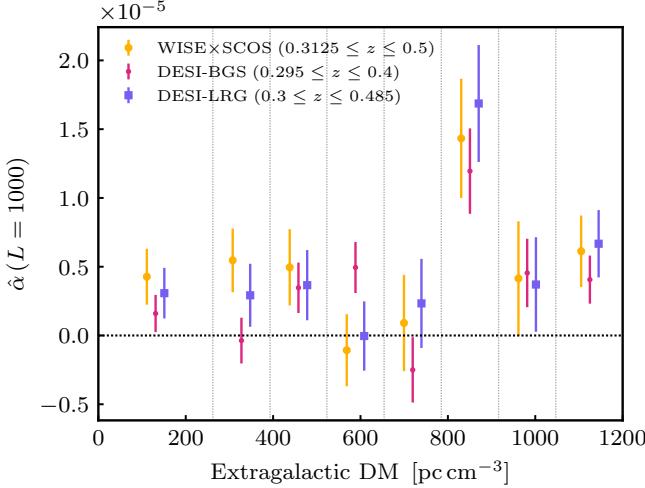


Figure 9. DM dependence of the FRB-galaxy correlation. We divide the CHIME/FRB catalog into DM bins (delimited by vertical lines) after subtracting the YMW16 estimate of the Galactic DM and cross-correlate each DM bin with the galaxy catalogs. The last DM bin extends to $\text{DM} = 3020 \text{ pc cm}^{-3}$. For each galaxy survey, we use the same redshift range (see legend) as in the left panel of Figure 7. The quantity $(\hat{\alpha}_L)_{L=1000}$ on the y -axis is defined in Eq. (11) and measures the level of FRB-galaxy correlation. This quantity is a per-object statistic that is derived from C_ℓ^{fg} . Hence, it does not necessarily follow number density variations in Figure 1.

nister et al. 2019; Prochaska et al. 2019; Ravi et al. 2019; Chittidi et al. 2020; Heintz et al. 2020; Law et al. 2020; Macquart et al. 2020; Mannings et al. 2021; Marcote et al. 2020; Simha et al. 2020; Bhandari et al. 2020a,b; CHIME/FRB Collaboration 2020b; Bhardwaj et al. 2021; James et al. 2021a). The rest of this section is devoted to interpreting this result further.

In Figure 9, the DM bin at $785 < \text{DM} < 916 \text{ pc cm}^{-3}$ is an outlier, suggesting a narrow feature in the DM dependence of the FRB-galaxy correlation. Given the error bars, it is difficult to say with statistical significance whether the apparent narrowness is real, or whether the true DM dependence is slowly varying. A crucial point here is that the three galaxy catalogs are highly correlated spatially (after restricting to the appropriate redshift ranges), which implies that the three measurements in Figure 9 have highly correlated statistical errors. Future CHIME/FRB catalogs will have smaller error bars

and can statistically distinguish a narrow feature from slowly varying DM dependence.

As a check, we remade Figure 9 using the NE2001 (Cordes & Lazio 2002) model for Galactic DM, instead of the YMW16 model. The effect of this change is small compared to the statistical errors in Figure 9.

We also performed the following visual check. The outlier bin with $785 < \text{DM} < 916 \text{ pc cm}^{-3}$ in Figure 9 only contains 12 FRBs in the DESI footprint. In Figure 10 we show the DESI-BGS galaxies in the vicinity of each FRB. The large FRB-galaxy correlation can be seen visually as an excess of galaxies (relative to random catalogs) within $7'$ of an FRB.⁶ None of the individual FRBs in Figure 9 give a statistically significant cross-correlation on its own, but the total FRB-galaxy correlation is significant at the 3σ – 4σ level. (We caution the reader that the galaxy counts in Figure 10 do not obey Poisson statistics, since the galaxies are clustered.) There are no visual red flags in Figure 10, such as a single FRB that gives an implausibly large contribution to the cross-correlation.

Finally, we address the question of whether the high-DM signal in Figure 9 is consistent with direct host associations. Consider the following two statements, in the context of FRB surveys with the CHIME/FRB sensitivity:

1. A random FRB with extragalactic DM $\geq 785 \text{ pc cm}^{-3}$ has an order-one probability of having redshift $z \sim 0.4$ (implying $\text{DM}_{\text{host}} \gtrsim 400 \text{ pc cm}^{-3}$).
2. A random FRB at redshift $z \sim 0.4$ has an order-one probability of having extragalactic DM $\geq 785 \text{ pc cm}^{-3}$.

The high-DM signal in Figure 9 implies statement 1, but not statement 2. We will now argue that statement 1 is actually consistent with direct associations.

⁶ The scale $\Theta \equiv 7'$ was obtained as $\Theta = \sqrt{8}/L$, where $L = 1396$ is the template scale where the DESI-BGS cross-correlation peaks in Figure 7. The factor $\sqrt{8}$ was derived by matching the variance $(\Theta^2/2)$ of a radius- Θ top hat to the variance $(4/L^2)$ of a Gaussian beam $e^{-\theta^2 L^2/4}$.

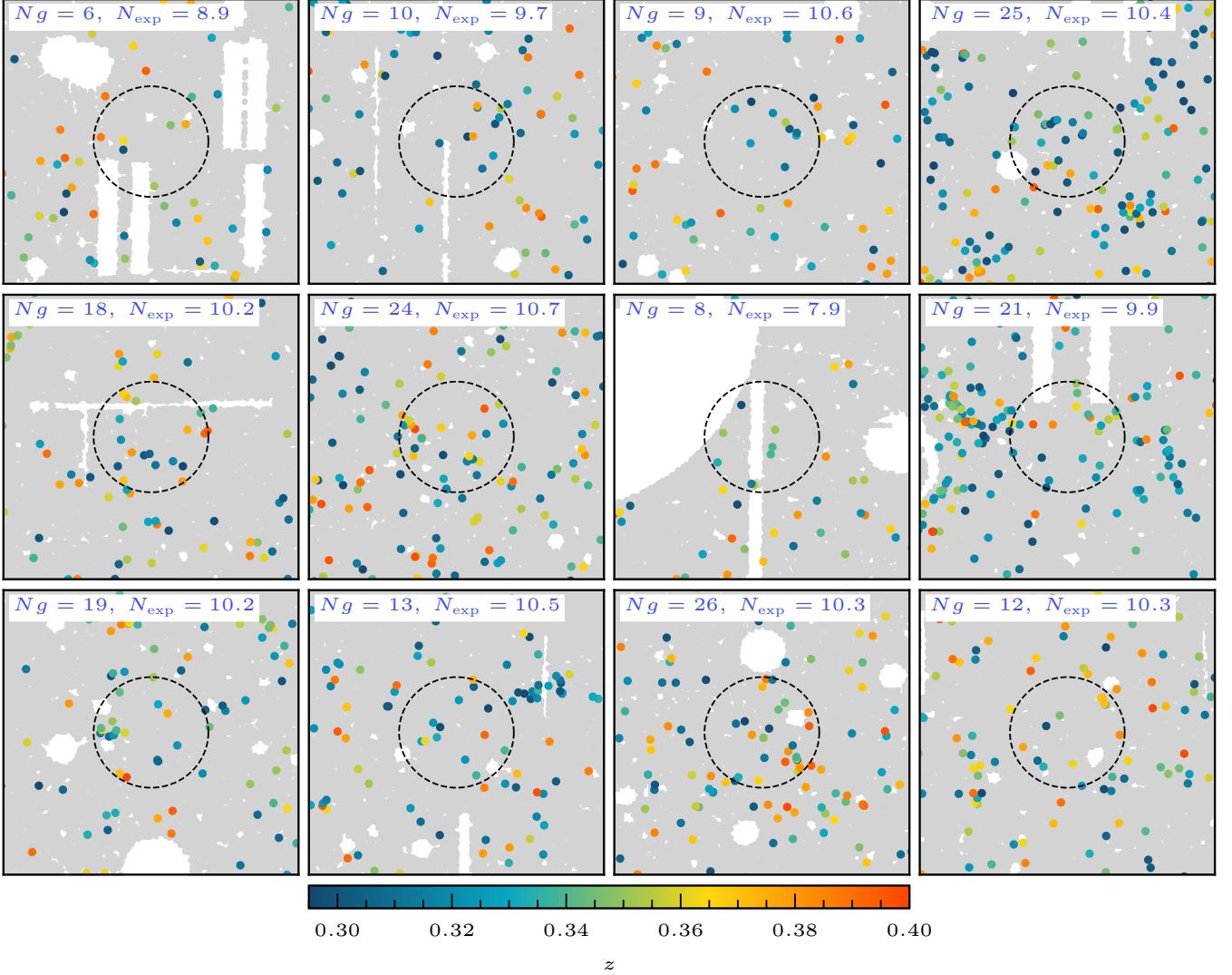


Figure 10. Visual representation of the cross-correlation between FRBs with $785 < \text{DM} < 916 \text{ pc cm}^{-3}$, and DESI-BGS galaxies. There are 12 FRBs in this DM range in the DESI footprint. For each such FRB, we plot the DESI-BGS galaxies in the redshift range $0.295 < z < 0.4$ in the vicinity of the FRB. We color-code galaxies by redshift, but note that redshift errors are comparable ($\sigma_z \sim 0.03$) to the redshift range shown. The gray points are objects in the DESI random catalog, to give a sense for the DESI mask geometry. The dashed circles are centered at FRBs, with radius $\Theta = 7'$ (see §4.2). The value of N_g in the upper left is the observed number of galaxies in the circle. The value of N_{exp} is the expected number of galaxies in the circle, inferred from randoms. The FRB-galaxy correlation appears as a statistical preference for $N_g > N_{\text{exp}}$.

The key point is that there are few direct associations at high DM. Out of the 14 direct associations to date, only one has extragalactic DM $\geq 785 \text{ pc cm}^{-3}$: an FRB with YMW16-subtracted DM 850 pc cm^{-3} at $z = 0.6$ (Law et al. 2020). Based on this one high-DM event, one cannot rule out statement 1 above (note that statement 2 would clearly be inconsistent with direct associations).

Therefore, there is no inconsistency between the high-DM FRB-galaxy correlation in Figure 9, and direct FRB host associations to date. The number of direct associations is rapidly growing, and we predict that FRBs with extragalactic DM $\geq 785 \text{ pc cm}^{-3}$ at $z \sim 0.4$ will be found in direct associations soon (see §5 for more discussion).

One final comment: we have presented statistical evidence that statement 1 is true in CHIME/FRB, but statement 1 depends to some extent on the selection function of the FRB survey. In particular, future surveys that are sensitive to fainter sources may detect larger numbers of high-redshift FRBs. In this scenario, it is possible that FRBs with extragalactic DM $\geq 785 \text{ pc cm}^{-3}$ will mostly come from $z \sim 0.8$, as expected from the Macquart relation.

4.3. Host halo DMs

In the previous subsection, we found statistical evidence for a population of FRBs at $z \sim 0.4$ with DM \gtrsim

400 pc cm⁻³. In this section we will propose a possible mechanism for generating such large host DMs. Note that for a Galactic pulsar, a DM of order 400 pc cm⁻³ would be unsurprising, but pulsar sight lines lie preferentially in the Galactic disk (boosting the DM), whereas FRBs are observed from a random direction.

Bright galaxies in cosmological surveys are usually found in large dark matter halos (Wechsler & Tinker 2018). Therefore, FRBs that correlate with such galaxies may have large host DMs, due to DM contributions from gas in the host halos. We refer to such a contribution as the *host halo DM* DM_{hh} , since the term “halo DM” is often used to refer to the contribution from the Milky Way halo.

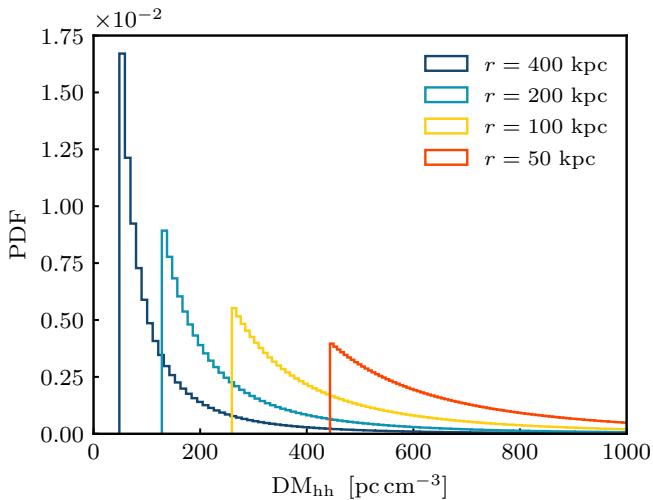


Figure 11. Host halo DM distributions for FRBs in a halo of mass $M = 10^{14} M_{\odot}$. The host halo DM is determined by two parameters: the distance r between the FRB and halo center, and viewing angle θ . Each histogram corresponds to one choice of r , with 10^5 values of θ . The halo gas profile is the “ICM” model from Prochaska & Zheng (2019).

Can host halo DMs plausibly be of order $\text{DM}_{\text{hh}} \gtrsim 400$ pc cm⁻³? To answer this question, in Figure 11, we show DM_{hh} histograms for simulated FRBs in a halo of mass $M = 10^{14} M_{\odot}$. The halo gas profile is the intra-cluster medium (ICM) model from Prochaska & Zheng (2019), based on X-ray observations from Vikhlinin et al. (2006)⁷. It is seen that FRBs near the centers ($r \lesssim 100$ kpc) of large ($M \sim 10^{14} M_{\odot}$) halos can have host halo DMs $\text{DM}_{\text{hh}} \gtrsim 400$ pc cm⁻³.

⁷ To calculate the host halo DM $\text{DM}_{\text{hh}} = \int dr n_e(r)$, we used a slightly modified version of the FRB software (github.com/FRBs/FRB) by Prochaska et al. We thank the authors for making their software public.

Thus, the high-DM signal in Figure 9 is plausibly explained by a small subpopulation of FRBs at redshift $0.3 \lesssim z \lesssim 0.5$ near the centers of large halos. Such a subpopulation could have $\text{DM}_{\text{host}} \gtrsim 400$ pc cm⁻³, and strongly correlate with galaxies, since bright galaxies are often in high-mass halos.

This mechanism is a proof of concept to show that $\text{DM}_{\text{hh}} \gtrsim 400$ pc cm⁻³ is plausible in some halo gas models. Other mechanisms may also be possible, such as augmentation by intervening foreground galaxies (James et al. 2021b). We emphasize that the statistical evidence for a population of FRBs with $\text{DM}_{\text{host}} \gtrsim 400$ pc cm⁻³, presented in the previous subsection, does not depend on the assumption of a particular model or mechanism.

4.4. Propagation effects

So far, we have assumed that the observed FRB-galaxy correlation is owing to spatial correlations between the FRB and galaxy populations. In this subsection, we will explore the alternate hypothesis that host DMs are always small (say $\text{DM}_{\text{host}} \sim 70$ pc cm⁻³), and that propagation effects are responsible for the observed correlation between $z \sim 0.4$ galaxies and high-DM FRBs.

“Propagation effects” is a catch-all term for what happens to radio waves during their voyage from source and observer due to intervening plasma. For example, dispersion, scattering, and plasma lensing are all propagation effects. Propagation effects can produce an apparent correlation between low-redshift galaxies and high-redshift FRBs, even when the underlying populations are not spatially correlated.

For example, low-redshift galaxies are spatially correlated with free electrons, which contribute to the DM of background FRBs. The DM contribution can either increase or decrease the probability of detecting a background FRB, depending on the selection function of the instrument. This effect can produce an apparent correlation or anticorrelation between low- z galaxies and high- z FRBs, in the absence of any spatial correlation between the galaxy and FRB populations.

Here, we will calculate contributions to C_{ℓ}^{fg} from propagation effects, using formalism from Rafiee-Ravandi et al. (2020). We will use a fiducial model in which host DMs are small ($\text{DM}_{\text{host}} \sim 70$ pc cm⁻³), implying negligible spatial correlation between $z \sim 0.4$ galaxies and high-DM FRBs. This is because we are interested in exploring the hypothesis that propagation effects (not large host DMs) are entirely responsible for the observed DM dependence in Figure 9. We describe the fiducial model in the next few paragraphs.

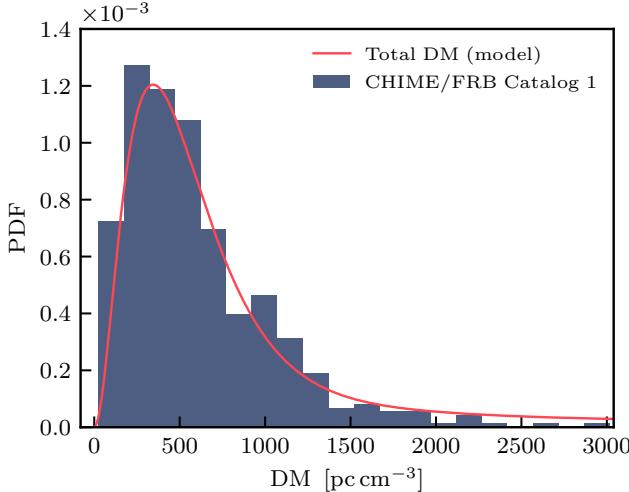


Figure 12. DM distribution (solid curve) for the fiducial FRB model used to study propagation effects in §4.4, with the CHIME/FRB DM distribution shown for comparison (histogram). In this model, host DMs are small, to explore the hypothesis that the correlation between $z \sim 0.4$ galaxies and high-DM FRBs is due to propagation effects, rather than large host DMs. The host DM distribution (not shown) is sharply peaked at $\text{DM}_{\text{host}} \sim 70 \text{ pc cm}^{-3}$.

First, we model the distribution of FRBs in redshift and DM. We assume that the FRB redshift distribution is

$$\frac{dn_f^{2d}}{dz} \propto z^2 e^{-\gamma z} \quad (22)$$

and that the host DM distribution is lognormal, and independent of redshift:

$$p(\text{DM}_{\text{host}}) \propto \frac{1}{\text{DM}_{\text{host}}} \exp\left(-\frac{(\log \text{DM}_{\text{host}} - \mu_{\log})^2}{2\sigma_{\log}^2}\right). \quad (23)$$

In Eqs. (22), (23), we choose parameters

$$\gamma = 6.7 \quad \mu_{\log} = 4.2 \quad \sigma_{\log} = 5. \quad (24)$$

The total DM is $\text{DM} = \text{DM}_{\text{IGM}}(z) + \text{DM}_{\text{host}}$. These parameters have been chosen so that the median FRB redshift is 0.4, the median host DM is 67 pc cm^{-3} , and the distribution of total DMs is similar to the observed DM distribution in Figure 12.

We will also need a fiducial model for $P_{ge}(k)$, the 3D galaxy-electron power spectrum at comoving wavenumber k . For reasons that we will explain shortly, we will need to know the one-halo contribution in the limit $k \rightarrow 0$, which is (Rafiei-Ravandi et al. 2020)

$$\lim_{k \rightarrow 0} P_{ge}^{1h}(k, z) = \frac{\langle N_e^{\text{ion}} \rangle}{n_{e,0}}, \quad (25)$$

where $\langle N_e^{\text{ion}} \rangle$ is the average (over survey galaxies) number of electrons in the halo containing a galaxy, and $n_{e,0}$

is the comoving electron number density. To compute $\langle N_e^{\text{ion}} \rangle$, we assume that survey galaxies are contained in dark matter halos whose mass M_h is lognormal-distributed, with parameters:

$$\langle \lambda \rangle = 13.4 \quad \sigma(\lambda) = 0.35, \quad (26)$$

where $\lambda \equiv \log_{10}(M_h/M_\odot)$. This distribution is a rough fit to the halo mass distribution shown in Figure 3 of Schaan et al. (2021) for SDSS-LOWZ, a well-characterized $z \sim 0.3$ galaxy survey similar to the ones considered here. We assume that these large halos have baryon-to-matter ratio equal to the cosmic average (Ω_b/Ω_m), with ionization fraction $f_b = 0.75$.

Finally, we model the CHIME/FRB selection function $S(\text{DM})$ in DM. This has been measured via Monte Carlo analysis of simulated events, and the result is shown in Figure 14 of the CHIME/FRB Catalog 1 paper (CHIME/FRB Collaboration 2021). Here, we will use the following rough visual fit:

$$\log S(\text{DM}) = 0.1 - 0.14 \left[\log \left(\frac{\text{DM}}{1000} \right) \right]^2. \quad (27)$$

The selection function $S(\text{DM})$ is, up to normalization, the probability that a random FRB with a given DM is detected by CHIME/FRB. As an aside, CHIME/FRB has a selection bias against detecting high-DM FRBs due to frequency channel smearing and a bias against detecting low-DM FRBs due to the details of the high-pass filtering used in radio frequency interference removal. (Scattering biases will be discussed later in this section.) This combination of biases results in the selection function (Eq. 27) with a local maximum at $\text{DM} \sim 1000 \text{ pc cm}^{-3}$.

With the fiducial model in the previous few paragraphs, we now proceed to calculate contributions to C_ℓ^{fg} from propagation effects.

The first propagation effect we will consider is “DM-completeness”, described schematically as follows. Consider a foreground population of galaxies, and a background (i.e. higher-redshift) population of FRBs. The galaxies are spatially correlated with ionized electrons, which increase DMs of the FRBs, by adding dispersion along the line of sight. This can either increase or decrease the apparent number density of FRBs, depending on whether $dS/d(\text{DM})$ is positive or negative. This combination of effects produces a correlation between number densities of FRBs and galaxies, i.e. a contribution to C_ℓ^{fg} that can be positive or negative.

In Rafiei-Ravandi et al. (2020), the contribution to C_ℓ^{fg} from DM-completeness is calculated:

$$C_\ell^{fg} = \frac{1}{n_g^{2d}} \int dz \frac{H(z)}{\chi(z)^2} \frac{dn_g^{2d}}{dz} W_f(z) P_{ge}\left(\frac{\ell}{\chi(z)}, z\right), \quad (28)$$

where the DM-completeness weight function W_f for DM bin $[\text{DM}_{\min}, \text{DM}_{\max}]$ is

$$W_f(z) = \frac{n_{e,0}}{n_f^{2d}} \frac{1+z}{H(z)} \int_z^\infty dz' \times \int_{\text{DM}_{\min}}^{\text{DM}_{\max}} d(\text{DM}) \frac{d^2 n_f^{2d}}{dz' d(\text{DM})} \frac{d \log S}{d(\text{DM})} \quad (29)$$

and $\chi(z)$ is comoving distance to redshift z . We convert this expression for C_ℓ^{fg} to an expression for our parameter α as follows:

$$\begin{aligned} \alpha &= \lim_{\ell \rightarrow 0} C_\ell^{fg(1h)} \\ &= \frac{1}{n_g^{2d}} \int dz \frac{H(z)}{\chi(z)^2} \frac{dn_g^{2d}}{dz} W_f(z) \frac{\langle N_e^{\text{ion}}(z) \rangle}{n_{e,0}}, \end{aligned} \quad (30)$$

where we have used Eq. (18) in the first line and Eqs. (25), (28) in the second line.

The second propagation effect we will consider is “DM-shifting”, which arises for an FRB catalog that has been binned in DM, as in Figure 9. Even in the absence of an instrumental selection function, DM fluctuations along the line of sight can shift FRBs across DM bin boundaries, either increasing or decreasing the observed number density of FRBs in a given bin. This effect is distinct from the DM-completeness effect described above, and also produces a contribution to C_ℓ^{fg} that can be positive or negative. Using results from Rafiei-Ravandi et al. (2020), the DM-shifting bias to α_L is given by the previous expression (30), but with the following expression for the DM-shifting weight function:

$$W_f(z) = -\frac{n_{e,0}}{n_f^{2d}} \frac{1+z}{H(z)} \int_z^\infty dz' \left[\frac{d^2 n_f^{2d}}{dz' d(\text{DM})} \right]_{\text{DM}_{\min}}^{\text{DM}_{\max}}. \quad (31)$$

In Figure 13, we show α_L -biases from the DM-completeness and DM-shifting propagation effects in our fiducial model, computed using Eqs. (29)–(31). For simplicity, we have approximated the precise z -dependence of the redshift-binned galaxy surveys in Figure 9 by assuming $dn_g^{2d}/dz = \text{const}$ for $0.3 \leq z \leq 0.4$. (The results are not very sensitive to the galaxy redshift distribution.)

Comparing to the FRB-galaxy correlation shown previously in Figure 9, we see that the total bias is $\sim 0.5\sigma$ in the second DM bin ($262 < \text{DM} < 393 \text{ pc cm}^{-3}$), and

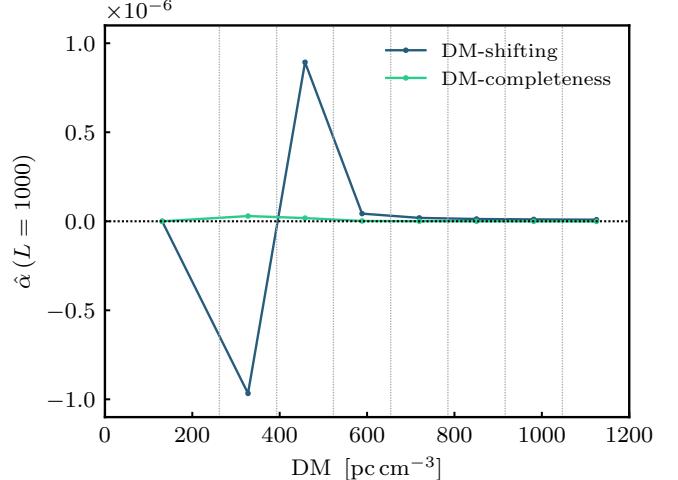


Figure 13. Predicted contribution to the FRB-galaxy correlation α_L (Eq. 11) from propagation effects, in the fiducial model from §4.4. The DM binning is the same as Figure 9. Comparing to the error bars in Figure 9, the DM-shifting contribution is $\sim 0.5\sigma$ in the second and third DM bins ($262 < \text{DM} < 393$ and $393 < \text{DM} < 523 \text{ pc cm}^{-3}$) and $\lesssim 0.1\sigma$ in the other bins. The DM-completeness contribution is very small.

$\lesssim 0.1\sigma$ in the other bins. These biases are too small, and have the wrong DM dependence, to explain the FRB-galaxy correlation shown previously in Figure 9.

So far, we have only considered propagation effects involving dispersion. The next propagation effect we might want to consider is scattering completeness, described intuitively as follows. Consider a foreground population of galaxies and a background population of FRBs. The galaxies are correlated with free electrons, which scatter-broaden FRBs and change their observed number density. Since scatter-broadening always decreases the probability that an FRB is detected, this effect always produces negative C_ℓ^{fg} .⁸ Therefore, scattering completeness cannot be responsible for the observed FRB-galaxy correlation, which is positive (as expected for clustering).

A final category of propagation effects is strong lensing (either plasma lensing or gravitational lensing) by foreground galaxies. Although strong lenses are rare, they can produce large magnification, increasing the detection rate of background FRBs by a large factor if the FRB luminosity function is sufficiently steep. A complete analysis of strong lensing in CHIME/FRB would

⁸ Formally, the selection function for scattering is a decreasing function of scattering width. This can be seen directly in Figure 15 of (CHIME/FRB Collaboration 2021).

be a substantial undertaking, and we defer it to a future paper.

5. SUMMARY AND CONCLUSIONS

In this paper, we find a cross-correlation between CHIME FRBs and galaxies at redshifts $0.3 \lesssim z \lesssim 0.5$. The correlation is statistically significant in three galaxy surveys: WISE \times SCOS, DESI-BGS, and DESI-LRG. The statistical significance of the detection in each survey is $p \sim 2.7 \times 10^{-5}$, 3.1×10^{-4} , and 4.1×10^{-4} , respectively. These p -values account for look-elsewhere effects in both angular scale L and redshift range.

The FRB-galaxy correlation is detected on angular scales ($\ell \sim 1000$) in the one-halo regime. In this regime, the amplitude of the correlation is proportional to the mean “link count” η of the FRB population, i.e. mean number of galaxies in the same halo as an FRB. Cross-correlating CHIME FRBs with $0.3 \lesssim z \lesssim 0.5$ galaxies, we find η of order unity.

This measurement of η cannot be directly translated to the probability p that an FRB is in the given redshift range. We can write $\eta = p\tilde{\eta}$, where $\tilde{\eta}$ is the mean link count of FRBs in the redshift range. Formally, we measure $(p\tilde{\eta})$ but not the individual factors $p, \tilde{\eta}$. However, in the bright galaxy surveys considered here, dark matter halos rarely contain more than a few catalog galaxies. We conclude that $\tilde{\eta}$ must be of order unity, implying that p is also of order unity. That is, an order-one fraction of CHIME FRBs are in redshift range $0.3 \lesssim z \lesssim 0.5$.

We have phrased this conclusion as a qualitative statement (“order-one fraction”), since it is difficult to assign a quantitative upper bound to $\tilde{\eta}$. This issue is a limitation of measuring FRB-galaxy correlations in the one-halo regime, where the FRB redshift distribution always appears multiplied by a linking factor $\tilde{\eta}$. Future CHIME/FRB catalogs should contain enough FRBs to detect the FRB-galaxy correlation on two-halo scales ($\ell \sim 100$) (Rafiei-Ravandi et al. 2020), which will help break this degeneracy.

We find statistical evidence for a population of FRBs with large host DMs, on the order of $\text{DM}_{\text{host}} \sim 400 \text{ pc cm}^{-3}$. More precisely, we detect a nonzero correlation between FRBs with $\text{DM} \geq 785 \text{ pc cm}^{-3}$ (after subtracting the YWM16 estimate of the Milky Way DM) and galaxies at $z \sim 0.4$, where the IGM contribution to the DM is $\text{DM}_{\text{IGM}}(z) \sim 360 \text{ pc cm}^{-3}$.

This may appear to be in tension with direct host galaxy associations. At the time of this writing, 14 FRBs have been localized to host galaxies, all of which have $\text{DM}_{\text{host}} \lesssim 200 \text{ pc cm}^{-3}$. However, FRBs with $\text{DM} \geq 785 \text{ pc cm}^{-3}$ are currently uncommon, and our FRB-galaxy correlation result must be interpreted carefully.

It implies that an order-one fraction of high-DM FRBs are at redshift $z \sim 0.4$ in CHIME/FRB, but it does not imply that an order-one fraction of FRBs at redshift $z \sim 0.4$ have high DM. These statements are actually consistent with the direct associations. Since there is currently only one direct association with YMW16-subtracted $\text{DM} \geq 785 \text{ pc cm}^{-3}$, one cannot currently rule out the possibility that an order-one fraction of high-DM FRBs are at $z \sim 0.4$.

The number of direct host associations is rapidly growing, and we predict that direct associations will soon find high-DM FRBs with $z \sim 0.4$. However, we note that most direct associations to date have been discovered by ASKAP at lower DM (on average) than the CHIME/FRB sample.

We briefly explore mechanisms for producing host DMs $\gtrsim 400 \text{ pc cm}^{-3}$, and show that contributions from gas in large halos provide a plausible mechanism. Quantitatively, we find that for FRBs near the centers ($r \lesssim 100 \text{ kpc}$) of large ($M \sim 10^{14} M_{\odot}$) halos the host halo DM can be $\gtrsim 400 \text{ pc cm}^{-3}$ (Figure 11), at least in one widely used ICM model (Prochaska & Zheng 2019). FRBs in such halos will strongly correlate with galaxies, since bright survey galaxies are often found in large halos. We show that line-of-sight propagation effects are unlikely to be a significant source of bias (§4.4).

Future measurements of FRB-galaxy cross-correlations will have higher SNR, and the results presented here could be extended in several ways. One could bin simultaneously in galaxy redshift and FRB DM, to explore the FRB-galaxy correlation strength as a function of two variables (z, DM). Cross-correlations can constrain the high- z tail of the FRB redshift distribution, where direct associations are difficult since individual galaxies are usually faint (Eftekhar & Berger 2017). Very high- z FRBs, if present, can be used to constrain cosmic reionization history (Caleb et al. 2019; Linder 2020; Zhang et al. 2021). Finally, line-of-sight propagation effects will eventually be detectable in C_{ℓ}^{fg} , and will be an interesting probe of the distribution of electrons in the universe.

This paper is based on FRBs from CHIME/FRB Catalog 1, which contains 489 unique sources and approximate angular sky positions. Future CHIME/FRB catalogs will include more FRB sources, many of which will have improved angular resolution through use of baseband data (Michilli et al. 2021). The FRB-galaxy correlation presented here should have much higher statistical significance in future CHIME/FRB catalogs and will be exciting to explore.

ACKNOWLEDGMENTS

We thank the Dominion Radio Astrophysical Observatory, operated by the National Research Council Canada, for gracious hospitality and expertise. CHIME is funded by a grant from the Canada Foundation for Innovation (CFI) 2012 Leading Edge Fund (Project 31170) and by contributions from the provinces of British Columbia, Québec and Ontario. The CHIME/FRB Project is funded by a grant from the CFI 2015 Innovation Fund (Project 33213), and by contributions from the provinces of British Columbia, Québec, and by the Dunlap Institute for Astronomy and Astrophysics at the University of Toronto. The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. Additional support was provided by the Canadian Institute for Advanced Research (CIFAR), McGill University and the McGill Space Institute via the Trottier Family Foundation, and the University of British Columbia. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research & Innovation. K.M.S. was supported by an NSERC Discovery Grant and a CIFAR fellowship. K.W.M. is supported by an NSF grant (2008031). M.D. is supported by a Killam Fellowship, NSERC Discovery Grant, CIFAR, and by the FRQNT Centre de Recherche en Astrophysique du Québec (CRAQ). M.B. is supported by an FRQNT Doctoral Research Award. K. B. is supported by an NSF grant (2006548). P.C. is supported by an FRQNT Doctoral Research Award. B.M.G. is supported by an NSERC Discovery grant (RGPIN-2015-05948), and by the Canada Research Chairs (CRC) program. D.C.G is supported by the John I. Watters Research Fellowship. V.M.K. holds the Lorne Trottier Chair in Astrophysics & Cosmology, a Distinguished James McGill Professorship and receives support from an NSERC Discovery Grant (RGPIN 228738-13) and Gerhard Herzberg Award, from an R. Howard Webster Foundation Fellowship from CIFAR, and from the FRQNT CRAQ. C.L. was supported by the U.S. Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. J.M.-P is a Kavli Fellow. D.M. is a Banting Fellow. E.P. acknowledges funding from an NWO Veni Fellowship. P.S. is a Dunlap Fellow and an NSERC Postdoctoral Fellow. K.S. is supported by the NSF Graduate Research Fellowship Program. FRB research at UBC is supported by an NSERC Discovery Grant and by CIFAR. This research was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). The Photometric Redshifts for the Legacy Surveys (PRLS) catalog used in this paper was produced thanks to funding from the U.S. Department of Energy Office of Science, Office of High Energy Physics via grant DE-SC0007914.

REFERENCES

- Aghanim, N., Akrami, Y., Ashdown, M., et al. 2020, A&A, 641, A6, doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910)
- Alonso, D. 2021, Physical Review D, 103, 123544, doi: [10.1103/PhysRevD.103.123544](https://doi.org/10.1103/PhysRevD.103.123544)
- Alonso, D., Salvador, A. I., Sánchez, F. J., et al. 2015, Monthly Notices of the Royal Astronomical Society, 449, 670, doi: [10.1093/mnras/stv309](https://doi.org/10.1093/mnras/stv309)
- Balaguera-Antolínez, A., Bilicki, M., Branchini, E., & Postiglione, A. 2018, Monthly Notices of the Royal Astronomical Society, 476, 1050, doi: [10.1093/mnras/sty262](https://doi.org/10.1093/mnras/sty262)
- Bannister, K. W., Deller, A. T., Phillips, C., et al. 2019, Science, 365, 565, doi: [10.1126/science.aaw5903](https://doi.org/10.1126/science.aaw5903)
- Bassa, C. G., Tendulkar, S. P., Adams, E. A. K., et al. 2017, The Astrophysical Journal, 843, L8, doi: [10.3847/2041-8213/aa7a0c](https://doi.org/10.3847/2041-8213/aa7a0c)
- Bhandari, S., Sadler, E. M., Prochaska, J. X., et al. 2020a, The Astrophysical Journal, 895, L37, doi: [10.3847/2041-8213/ab672e](https://doi.org/10.3847/2041-8213/ab672e)
- Bhandari, S., Bannister, K. W., Lenc, E., et al. 2020b, The Astrophysical Journal, 901, L20, doi: [10.3847/2041-8213/abb462](https://doi.org/10.3847/2041-8213/abb462)
- Bhardwaj, M., Gaensler, B. M., Kaspi, V. M., et al. 2021, The Astrophysical Journal, 910, L18, doi: [10.3847/2041-8213/abeaa6](https://doi.org/10.3847/2041-8213/abeaa6)
- Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., & Steward, L. 2013, The Astrophysical Journal Supplement Series, 210, 9, doi: [10.1088/0067-0049/210/1/9](https://doi.org/10.1088/0067-0049/210/1/9)
- Bilicki, M., Peacock, J. A., Jarrett, T. H., et al. 2016, The Astrophysical Journal Supplement Series, 225, 5, doi: [10.3847/0067-0049/225/1/5](https://doi.org/10.3847/0067-0049/225/1/5)
- Bochenek, C. D., Ravi, V., Belov, K. V., et al. 2020, Nature, 587, 59, doi: [10.1038/s41586-020-2872-x](https://doi.org/10.1038/s41586-020-2872-x)
- Caleb, M., Flynn, C., & Stappers, B. W. 2019, Monthly Notices of the Royal Astronomical Society, 485, 2281, doi: [10.1093/mnras/stz571](https://doi.org/10.1093/mnras/stz571)
- Chatterjee, S., Law, C. J., Wharton, R. S., et al. 2017, Nature, 541, 58, doi: [10.1038/nature20797](https://doi.org/10.1038/nature20797)
- CHIME/FRB Collaboration. 2018, The Astrophysical Journal, 863, 48, doi: [10.3847/1538-4357/aad188](https://doi.org/10.3847/1538-4357/aad188)
- . 2019, The Astrophysical Journal, 885, L24, doi: [10.3847/2041-8213/ab4a80](https://doi.org/10.3847/2041-8213/ab4a80)
- . 2020a, Nature, 587, 54, doi: [10.1038/s41586-020-2863-y](https://doi.org/10.1038/s41586-020-2863-y)
- . 2020b, Nature, 582, 351, doi: [10.1038/s41586-020-2398-2](https://doi.org/10.1038/s41586-020-2398-2)
- . 2021, Submitted to ApJS
- Chittidi, J. S., Simha, S., Mannings, A., et al. 2020, arXiv e-prints, arXiv:2005.13158.
- <https://ui.adsabs.harvard.edu/abs/2020arXiv200513158C>
- Cooray, A., & Sheth, R. 2002, Physics Reports, 372, 1, doi: [10.1016/s0370-1573\(02\)00276-4](https://doi.org/10.1016/s0370-1573(02)00276-4)
- Cordes, J. M., & Chatterjee, S. 2019, Annual Review of Astronomy and Astrophysics, 57, 417, doi: [10.1146/annurev-astro-091918-104501](https://doi.org/10.1146/annurev-astro-091918-104501)
- Cordes, J. M., & Lazio, T. J. W. 2002, arXiv e-prints, astro. <https://ui.adsabs.harvard.edu/abs/2002astro.ph..7156C>
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, The Astronomical Journal, 157, 168, doi: [10.3847/1538-3881/ab089d](https://doi.org/10.3847/1538-3881/ab089d)
- Eftekhari, T., & Berger, E. 2017, The Astrophysical Journal, 849, 162, doi: [10.3847/1538-4357/aa90b9](https://doi.org/10.3847/1538-4357/aa90b9)
- Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005, The Astrophysical Journal, 622, 759, doi: [10.1086/427976](https://doi.org/10.1086/427976)
- Heintz, K. E., Prochaska, J. X., Simha, S., et al. 2020, The Astrophysical Journal, 903, 152, doi: [10.3847/1538-4357/abb6fb](https://doi.org/10.3847/1538-4357/abb6fb)
- Hodges, J. L. 1958, Ark. Mat., 3, 469, doi: [10.1007/BF02589501](https://doi.org/10.1007/BF02589501)
- James, C. W., Prochaska, J. X., Macquart, J. P., et al. 2021a, arXiv e-prints, arXiv:2101.07998. <https://ui.adsabs.harvard.edu/abs/2021arXiv210107998J>
- . 2021b, arXiv e-prints, arXiv:2101.08005. <https://ui.adsabs.harvard.edu/abs/2021arXiv210108005J>
- Josephy, A., Chawla, P., Curtin, A. P., et al. 2021, Submitted to ApJ
- Keating, L. C., & Pen, U.-L. 2020, Monthly Notices of the Royal Astronomical Society: Letters, 496, L106, doi: [10.1093/mnrasl/slaa095](https://doi.org/10.1093/mnrasl/slaa095)
- Kokubo, M., Mitsuda, K., Sugai, H., et al. 2017, The Astrophysical Journal, 844, 95, doi: [10.3847/1538-4357/aa7b2d](https://doi.org/10.3847/1538-4357/aa7b2d)
- Krakowski, T., Małek, K., Bilicki, M., et al. 2016, A&A, 596. <https://doi.org/10.1051/0004-6361/201629165>
- Law, C. J., Butler, B. J., Prochaska, J. X., et al. 2020, The Astrophysical Journal, 899, 161, doi: [10.3847/1538-4357/aba4ac](https://doi.org/10.3847/1538-4357/aba4ac)
- Li, D., Yalinewich, A., & Breysse, P. C. 2019, arXiv e-prints, arXiv:1902.10120
- Linder, E. V. 2020, Physical Review D, 101, 103019, doi: [10.1103/PhysRevD.101.103019](https://doi.org/10.1103/PhysRevD.101.103019)
- Macquart, J. P., Prochaska, J. X., McQuinn, M., et al. 2020, Nature, 581, 391, doi: [10.1038/s41586-020-2300-2](https://doi.org/10.1038/s41586-020-2300-2)
- Madhavacheril, M. S., Battaglia, N., Smith, K. M., & Sievers, J. L. 2019, Physical Review D, 100, 103532, doi: [10.1103/PhysRevD.100.103532](https://doi.org/10.1103/PhysRevD.100.103532)
- Mannings, A. G., Fong, W.-f., Simha, S., et al. 2021, The Astrophysical Journal, 917, 75, doi: [10.3847/1538-4357/abff56](https://doi.org/10.3847/1538-4357/abff56)

- Marcote, B., Nimmo, K., Hessels, J. W. T., et al. 2020, *Nature*, 577, 190, doi: [10.1038/s41586-019-1866-z](https://doi.org/10.1038/s41586-019-1866-z)
- Masui, K. W., & Sigurdson, K. 2015, *Physical Review Letters*, 115, 121301, doi: [10.1103/PhysRevLett.115.121301](https://doi.org/10.1103/PhysRevLett.115.121301)
- McQuinn, M. 2014, *The Astrophysical Journal Letters*, 780, L33, doi: [10.1088/2041-8205/780/2/l33](https://doi.org/10.1088/2041-8205/780/2/l33)
- Michilli, D., Masui, K. W., Mckinven, R., et al. 2021, *The Astrophysical Journal*, 910, 147, doi: [10.3847/1538-4357/abe626](https://doi.org/10.3847/1538-4357/abe626)
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *Astrophys. J.*, 490, 493, doi: [10.1086/304888](https://doi.org/10.1086/304888)
- Petroff, E., Hessels, J. W. T., & Lorimer, D. R. 2019, *Astronomy and Astrophysics Review*, 27, 4, doi: [10.1007/s00159-019-0116-6](https://doi.org/10.1007/s00159-019-0116-6)
- Petroff, E., & Yaron, O. 2020, *Transient Name Server AstroNote*, 160, 1
- Platts, E., Weltman, A., Walters, A., et al. 2019, *Physics Reports*, 821, 1, doi: [10.1016/j.physrep.2019.06.003](https://doi.org/10.1016/j.physrep.2019.06.003)
- Prochaska, J. X., & Zheng, Y. 2019, *Monthly Notices of the Royal Astronomical Society*, 485, 648, doi: [10.1093/mnras/stz261](https://doi.org/10.1093/mnras/stz261)
- Prochaska, J. X., Macquart, J.-P., McQuinn, M., et al. 2019, *Science*, 366, 231, doi: [10.1126/science.aay0073](https://doi.org/10.1126/science.aay0073)
- Rafiei-Ravandi, M., Smith, K. M., & Masui, K. W. 2020, *Physical Review D*, 102, 023528, doi: [10.1103/PhysRevD.102.023528](https://doi.org/10.1103/PhysRevD.102.023528)
- Raichoor, A., Eisenstein, D. J., Karim, T., et al. 2020, *Research Notes of the American Astronomical Society*, 4, 180, doi: [10.3847/2515-5172/abc078](https://doi.org/10.3847/2515-5172/abc078)
- Ravi, V., Catha, M., D'Addario, L., et al. 2019, *Nature*, 572, 352, doi: [10.1038/s41586-019-1389-7](https://doi.org/10.1038/s41586-019-1389-7)
- Reischke, R., Hagstotz, S., & Lilow, R. 2021a, *Physical Review D*, 103, 023517, doi: [10.1103/PhysRevD.103.023517](https://doi.org/10.1103/PhysRevD.103.023517)
- . 2021b, arXiv e-prints, arXiv:2102.11554. <https://ui.adsabs.harvard.edu/abs/2021arXiv210211554R>
- Ruiz-Macias, O., Zarrouk, P., Cole, S., et al. 2020, *Research Notes of the American Astronomical Society*, 4, 187, doi: [10.3847/2515-5172/abc25a](https://doi.org/10.3847/2515-5172/abc25a)
- Schaan, E., Ferraro, S., Amodeo, S., et al. 2021, *Physical Review D*, 103, 063513, doi: [10.1103/PhysRevD.103.063513](https://doi.org/10.1103/PhysRevD.103.063513)
- Scholz, F. W., & Stephens, M. A. 1987, *Journal of the American Statistical Association*, 82, 918, doi: [10.1080/01621459.1987.10478517](https://doi.org/10.1080/01621459.1987.10478517)
- Shirasaki, M., Kashiyama, K., & Yoshida, N. 2017, *Physical Review D*, 95, 083012, doi: [10.1103/PhysRevD.95.083012](https://doi.org/10.1103/PhysRevD.95.083012)
- Simha, S., Burchett, J. N., Prochaska, J. X., et al. 2020, *The Astrophysical Journal*, 901, 134, doi: [10.3847/1538-4357/abafc3](https://doi.org/10.3847/1538-4357/abafc3)
- Spanakis-Misirlis, A. 2021, *Astrophysics Source Code Library*, ascl:2106.028. <https://ui.adsabs.harvard.edu/abs/2021ascl.soft06028S>
- Spitler, L. G., Scholz, P., Hessels, J. W. T., et al. 2016, *Nature*, 531, 202, doi: [10.1038/nature17168](https://doi.org/10.1038/nature17168)
- Tendulkar, S. P., Bassa, C. G., Cordes, J. M., et al. 2017, *The Astrophysical Journal*, 834, L7, doi: [10.3847/2041-8213/834/2/l7](https://doi.org/10.3847/2041-8213/834/2/l7)
- Vikhlinin, A., Kravtsov, A., Forman, W., et al. 2006, *The Astrophysical Journal*, 640, 691, doi: [10.1086/500288](https://doi.org/10.1086/500288)
- Wechsler, R. H., & Tinker, J. L. 2018, *Annual Review of Astronomy and Astrophysics*, 56, 435, doi: [10.1146/annurev-astro-081817-051756](https://doi.org/10.1146/annurev-astro-081817-051756)
- Yao, J. M., Manchester, R. N., & Wang, N. 2017, *The Astrophysical Journal*, 835, 29, doi: [10.3847/1538-4357/835/1/29](https://doi.org/10.3847/1538-4357/835/1/29)
- Zhang, Z. J., Yan, K., Li, C. M., Zhang, G. Q., & Wang, F. Y. 2021, *The Astrophysical Journal*, 906, 49, doi: [10.3847/1538-4357/abceb9](https://doi.org/10.3847/1538-4357/abceb9)
- Zhou, R., Newman, J. A., Mao, Y.-Y., et al. 2020a, *Monthly Notices of the Royal Astronomical Society*, doi: [10.1093/mnras/staa3764](https://doi.org/10.1093/mnras/staa3764)
- Zhou, R., Newman, J. A., Dawson, K. S., et al. 2020b, *Research Notes of the American Astronomical Society*, 4, 181, doi: [10.3847/2515-5172/abc0f4](https://doi.org/10.3847/2515-5172/abc0f4)

APPENDIX

A. STATISTICAL ERRORS ON FRB LOCATIONS

Statistical errors in CHIME/FRB sky locations suppress the FRB-galaxy power spectrum C_ℓ^{fg} on small scales (large ℓ). The suppression takes the form $C_\ell^{fg} \rightarrow b_\ell C_\ell^{fg}$, where b_ℓ is the “beam” transfer function. Throughout the paper, we have modeled statistical errors as Gaussian, which leads to a transfer function of the form $b_\ell = e^{-\ell^2/L^2}$.

In this appendix, we will study statistical errors in more detail, using toy models of the CHIME/FRB instrument and the FRB population. Our conclusions are as follows:

- Statistical errors are not strictly Gaussian, but a Gaussian transfer function $b_\ell = e^{-\ell^2/L^2}$ is a good approximation within the error bars of our C_ℓ^{fg} measurement.
- Calculating L from first principles is hard, since it depends on both the CHIME/FRB instrument and the FRB population. A plausible range of L -values is $315 \leq L \leq 1396$.

This justifies the methodology used throughout the paper, where a Gaussian transfer function $b_\ell = e^{-\ell^2/L^2}$ is used, but L is a free parameter that we fit to the data, varying L over the range $315 \leq L \leq 1396$.

A.1. Toy beam model 1: uniform density, center of nearest beam

CHIME FRBs are detected by searching a 4×256 regular array of formed beams independently in real time. A best-fit sky location is assigned to each detected FRB based on the detection SNR (or nondetection) in each beam, using the localization pipeline described by [CHIME/FRB Collaboration \(2019, 2021\)](#). For an FRB which is detected in a single beam, the localization pipeline assigns sky location equal to the beam center. For a multibeam detection, the assigned sky location is roughly a weighted average of the beams where the event was detected.

As a first attempt to model statistical errors in the localization pipeline, suppose that when an FRB is detected we assign it to the center of the closest FRB beam. This is a reasonable model for the single-beam detections as described above.

We neglect wavelength dependence of the beam and evaluate at central wavelength $\lambda = 0.5$ m. We also neglect FRBs in sidelobes of the primary beam, since these are a small fraction of the CHIME/FRB Catalog 1. Finally, we assume that FRBs detected by CHIME/FRB

are uniformly distributed over the sky. (This turns out to be a dubious approximation, as we will show in the next subsection.) What is b_ℓ in this toy model?

Let Θ_e be the elevation of the detected FRB (with the usual astronomical definition, i.e. $\Theta_e = 0$ for an FRB on the horizon, or $\Theta_e = \pi/2$ for an FRB at zenith). Let θ_x, θ_y be east-west and north-south sky coordinates in a coordinate system where the center of the formed beam is at $(0, 0)$. Let S be the set of points closer to $(0, 0)$ than any of the other beam centers:

$$S = \left[-\frac{\theta_0}{2}, \frac{\theta_0}{2} \right] \times \left[-\frac{\theta_0}{2 \sin \Theta_e}, \frac{\theta_0}{2 \sin \Theta_e} \right], \quad (\text{A1})$$

where $\theta_0 = 23.4^\circ$ in CHIME. If the detected FRBs are uniformly distributed on the sky, then the effective beam is

$$b_\ell = \frac{\int_S d^2\theta J_0(\ell\theta)}{\int_S d^2\theta 1}, \quad (\text{A2})$$

where $J_0(x)$ is a Bessel function. For the CHIME/FRB catalog, which contains FRBs with different elevations Θ_e , we average b_ℓ over Θ_e values in the catalog. It is straightforward to compute the elevation Θ_e for each FRB, using values of RA, Dec, and time of observation taken directly from the catalog. The resulting transfer function b_ℓ is shown in Figure 14, and agrees well with a Gaussian transfer function $b_\ell = e^{-\ell^2/L^2}$ with $L = 670$.

A.2. Toy beam model 2: including selection bias

In the previous subsection, we neglected a selection bias: an FRB is more likely to be detected if it is located at the center of the beam (where the instrumental response is largest). To account for this selection bias, we define the unnormalized intensity beam:

$$B(\theta_x, \theta_y) = \frac{\text{sinc}^2(\theta_x D_x / \lambda) \text{sinc}^2((\theta_y D_y \sin \Theta_e) / \lambda)}{\text{sinc}^2(\theta_x D_x / 4\lambda)}, \quad (\text{A3})$$

where $\theta_x, \theta_y, \Theta_e, \lambda$ are defined in §A.1, the CHIME aperture is modeled as a rectangle with dimensions $(D_x, D_y) = (80, 100)$ meters, and $\text{sinc}(x) = \sin(\pi x)/(\pi x)$.

Assuming a Euclidean FRB fluence distribution $N(\geq F) \propto F^{-3/2}$ (consistent with statistical analysis of the CHIME/FRB Catalog 1 ([CHIME/FRB Collaboration 2021](#))), the probability of detecting an FRB at sky location (θ_x, θ_y) is $\propto B(\theta_x, \theta_y)^{3/2}$. Therefore, the beam transfer function is

$$b_\ell = \frac{\int_S d^2\theta B(\theta)^{3/2} J_0(\ell\theta)}{\int_S d^2\theta B(\theta)^{3/2}} \quad (\text{A4})$$

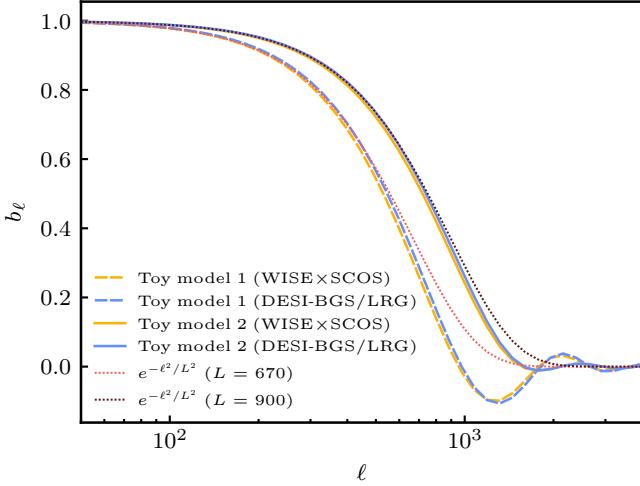


Figure 14. CHIME/FRB beam transfer function b_ℓ in a toy beam model, without (model 1, §A.1) and with (model 2, §A.2) selection bias included. Since b_ℓ is elevation dependent, the result is slightly different after averaging over FRBs in the WISE \times SCOS (orange) and DESI-BGS/LRG (blue) sky regions. For values of ℓ which are resolved by the beam (say, $b_\ell \gtrsim 0.25$), the beams are well approximated by Gaussians $b_\ell = e^{-\ell^2/L^2}$ (dotted curves).

averaged over catalog elevations Θ_e as in the previous subsection. The resulting transfer function b_ℓ is shown in Figure 14 and agrees well with a Gaussian transfer function $b_\ell = e^{-\ell^2/L^2}$ with $L = 900$.

A.3. Plausible range of L -values

Comparing the last two subsections, we see that the selection bias considered in §A.2 increases the effective value of L by 34%. This treatment of selection bias is incomplete, and a full study is outside the scope of this paper. For example, b_ℓ depends on wavelength λ , so there is a selection bias involving FRB frequency spectra. In addition, we have not attempted to model multibeam detections, which will be better localized than single-beam detections. Given these sources of modeling uncertainty, rather than trying to model the value of L precisely, we will assign a range of plausible L -values.

To assign a smallest plausible L -value, we make assumptions that lead to the largest plausible localization errors. We start with the toy beam model b_ℓ from §A.2, with $\lambda = 0.75$ m (the longest wavelength in CHIME). We then convolve with a halo profile ($b_\ell \rightarrow b_\ell u_\ell(M, z)^2$), where $u_\ell(M, z)$ is a Navarro-Frenk-White (NFW) density profile; Navarro et al. 1997), taking the halo mass M to be large ($M = 10^{14.5} h^{-1} M_\odot$) and the redshift to be small ($z = 0.05$). These specific values are somewhat arbitrary, but the goal is to establish a baseline plausible value of L_{\min} , not model a precise value of L . With the assumptions in this paragraph, we get $L_{\min} = 315$.

Similarly, to assign a largest plausible L -value, we make assumptions that lead to the smallest plausible localization errors. We use the smallest toy model from §A.2 with $\lambda = 0.375$ m (the shortest wavelength in CHIME). We assume that 40% of the events are multibeam detections, and that multibeam detections have localization errors that are smaller by a factor 3. As in the previous section, these specific values are somewhat arbitrary, but the goal is to establish a baseline plausible value of L_{\max} , not model a precise value of L . With the assumptions in this paragraph, we get $L_{\max} = 1396$.

B. NULL TESTS

As a general check for robustness of our FRB-galaxy correlation C_ℓ^{fg} , we would like to check that C_ℓ^{fg} does not depend on external variables, for example time of day (TOD). Our methodology for doing this is as follows. We divide the FRB catalog into low-TOD and high-TOD subcatalogs, cross-correlate each subcatalog with a galaxy sample, and compute the difference power spectrum:

$$d\hat{C}_\ell^{fg} \equiv \hat{C}_\ell^{fg}(\text{low}) - \hat{C}_\ell^{fg}(\text{high}). \quad (\text{B5})$$

Recall that for a non-null power spectrum \hat{C}_ℓ , we compressed the ℓ -dependence into a scalar summary statistic $\hat{\alpha}_L$ by taking a weighted ℓ -average (Eq. 11). Analogously, we compress the difference spectrum $d\hat{C}_\ell^{fg}$ into a summary statistic $\hat{\beta}_L$, defined by

$$\hat{\beta}_L = \sum_{\ell \geq \ell_{\min}} (2\ell + 1) \frac{e^{-\ell^2/L^2}}{C_\ell^{gg}} d\hat{C}_\ell^{fg}, \quad (\text{B6})$$

where L is an angular scale parameter. Next, by analogy with SNR_L (defined previously in Eq. 14), we define

$$\Delta_L = \frac{\hat{\beta}_L}{\text{Var}(\hat{\beta}_L)^{1/2}}. \quad (\text{B7})$$

The value of Δ_L quantifies consistency (in “sigmas”) between C_ℓ^{fg} for the low-TOD and high-TOD subcatalogs.

We fix $L = 1000$, and consider three choices of galaxy catalog: WISE \times SCOS with $z \geq 0.3125$, DESI-BGS with $z \geq 0.295$, and DESI-LRG with $z \leq 0.485$. These redshift ranges are “cherry-picked” to maximize the FRB-galaxy cross-correlation (see Figure 7), but this cherry-picking should not bias the difference statistic Δ_L . With these choices, we find $\Delta_L = \{1.22, -0.21, 1.30\}$ for WISE \times SCOS, DESI-BGS, and DESI-LRG respectively. Therefore, there is no statistical evidence for dependence of C_ℓ^{fg} on time of day, since a 1.22σ , 0.21σ , or 1.30σ result is not statistically significant.

This test can be generalized by splitting on a variety of external variables (besides TOD). In Table 3, we identify 12 such variables and denote the corresponding Δ_L values (with $L = 1000$) by Δ_i , where $i \in \{1, 2, 3, \dots, 12\}$. We note that these 12 tests are nonindependent, for example SNR is correlated with fluence. We also note that for many of these tests detection of a nonzero difference spectrum $d\hat{C}_\ell^{fg}$ does not necessarily indicate a problem. For example, DM dependence of C_ℓ^{fg} is expected at some level, since C_ℓ^{fg} is redshift dependent, and DM is correlated with redshift.

There are a few $\sim 2\sigma$ outliers in Table 3, but a few outliers are unsurprising, so it is not immediately clear whether the Δ_i values in Table 3 are statistically different from zero. To answer this question, we reduce the 12-component vector Δ_i into a scalar summary statistic, in a few different ways as follows.

Our first summary statistic is intended to test whether the most anomalous Δ_i -value in each column of Table 3 is statistically significant. We define

$$\Delta_{\max} = \max_i |\Delta_i|. \quad (\text{B8})$$

We then compare these values of Δ_{\max} to an ensemble of mocks. The mocks are constructed by randomizing the RA of each FRB in the catalog, keeping all other FRB properties (DM, SNR, etc.) fixed. This preserves any correlations which may be present between FRB properties. In the special case of the $|b| \geq 17^\circ$ null test, we recompute the value of b after randomizing RA.

In Table 4, we report the p -value for each Δ_{\max} , i.e. the fraction of mocks whose Δ_{\max} exceeds the “data” value. No statistically significant deviation from $\Delta_{\max} = 0$ is seen.

Our second summary statistic is intended to test whether the 12-component vector Δ_i is consistent with a multivariate Gaussian distribution. We define:

$$\chi^2 = \sum_{i,i'} \Delta_i \text{Cov}(\Delta_i, \Delta_{i'})^{-1} \Delta_{i'}, \quad (\text{B9})$$

where the covariance $\text{Cov}(\Delta_i, \Delta_{i'})$ is estimated from mock FRB catalogs, constructed as described above.

As before, to assign statistical significance, we compare the “data” value of χ^2 to an ensemble of mocks and report the associated p -value in Table 4. We find borderline evidence for $\chi^2 \neq 0$ for DESI-BGS ($p = 0.030$), but interpret this as inconclusive, since Table 4 contains six p -values, so one p -value as small as 0.03 is unsurprising (this happens with probability ≈ 0.18).

Finally, we compare the set of 12 Δ_i values to a jackknife distribution, obtained by randomly splitting the FRB catalog in half. We do this comparison using the 2-sample Kolmogorov-Smirnov (KS, Hodges 1958) and

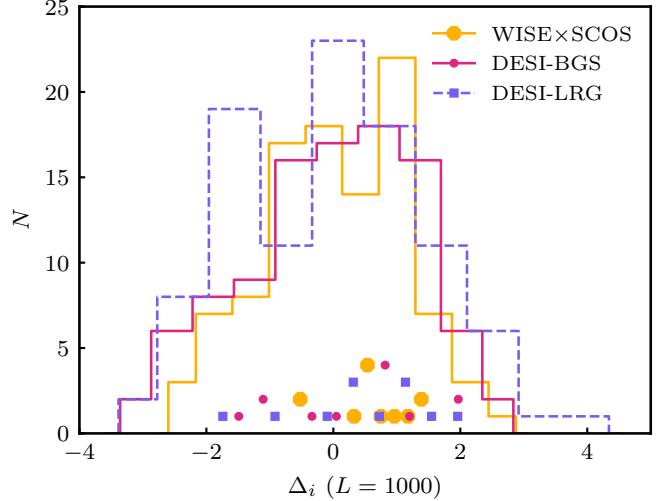


Figure 15. Histograms of the statistic Δ_i for the 12 null tests (filled markers) and 100 jackknives (lines). Using the KS and AD tests, the distributions are found to be consistent (Appendix B).

Anderson-Darling (AD, Scholz & Stephens 1987) tests. Figure 15 compares the two distributions for the three galaxy samples, and the last two columns of Table 4 summarize our results. As in the previous paragraph, there is one outlier: the WISE×SCOS KS p -value is 0.037, which we interpret as inconclusive, since it is one out of six p -values in the table (as in the previous paragraph).

Summarizing this appendix, we do not find statistically significant evidence that the FRB-galaxy clustering signal studied in this paper depends on any of the parameters in Table 3.

C. TAIL-FITTING PROCEDURE

In §3.5, we assign statistical significance of the FRB-galaxy detection, by defining a frequentist statistic SNR_{\max} , and ranking the “data” value $\text{SNR}_{\max}^{(\text{data})}$ within a histogram of simulated values $\text{SNR}_{\max}^{(\text{mock})}$. This procedure is conceptually straightforward, but there is a technical challenge: because $\text{SNR}_{\max}^{(\text{data})}$ turns out to be an extreme outlier, a brute-force approach requires an impractical number of simulations. Therefore, we fit the tail of the $\text{SNR}_{\max}^{(\text{mock})}$ distribution to an analytic distribution and assign statistical significance (or p -value) analytically.

Empirically, we find that the top 10% of the $\text{SNR}_{\max}^{(\text{mock})}$ distribution agrees well with the top 10% of a Gaussian distribution, as shown in Figure 16. The parameters of the Gaussian distribution were determined as follows. Let $p(x|\mu, \sigma)$ denote a Gaussian distribution with mean

Parameter	Median	Δ_i	Median	Δ_i	
	(WISE×SCOS)	WISE×SCOS	(DESI)	DESI-BGS	DESI-LRG
DM [pc cm ⁻³]	535.08	0.33	536.41	-1.68	-0.95
SNR	20.2	0.58	20.2	0.00	0.42
Scattering time [ms]	1.331	1.49	1.423	0.93	0.38
Pulse width [ms]	0.988	0.59	1.052	0.99	0.24
Spectral index	2.866	0.68	2.075	0.76	-0.25
Fluence [Jy ms]	3.503	1.28	3.115	-1.00	2.16
Bandwidth [MHz]	332.09	-0.44	358.09	0.80	1.58
Galactic $ b $	38°26'	0.59	38°24'	-1.27	-1.95
Catalog localization error	10°12'	0.52	9°53'	2.16	1.19
TOA - 58528 [MJD]	0.3686595	0.99	4.8473498	1.16	0.63
Peak frequency [MHz]	463.525	-0.63	449.036	1.97	1.30
Time of day [hr]	9.887	1.22	10.132	-0.21	1.30

Table 3. Null tests in Appendix B. For each parameter, we split the FRB catalog into “low” and “high” subcatalogs by comparing the parameter value to its median. (The median value is slightly different for FRBs in the WISE×SCOS and DESI footprints.) We correlate both subcatalogs with the galaxy surveys, and compute the statistic $\Delta = (\Delta_L)_{L=1000}$ (defined in Eq. B7), which measures consistency of the FRB-galaxy correlation in “sigmas”.

Galaxy sample	Δ_{\max}	p-value	χ^2	p-value	KS p-value	AD p-value
WISE×SCOS	1.49	0.779	9.26	0.659	0.037	0.067
DESI-BGS	2.16	0.270	22.96	0.030	0.381	0.250
DESI-LRG	2.16	0.274	17.26	0.145	0.113	0.171

Table 4. Summary statistics for the 12 null tests in Table 3. As described in Appendix B, we reduce the 12-component vector Δ_i into two scalar summary statistics Δ_{\max}, χ^2 , shown in the first four columns along with associated p-values from an ensemble of mocks. The last two columns compare the Δ_i values for each galaxy sample to a “jackknife” ensemble defined by randomly splitting the CHIME/FRB catalog.

μ and variance σ^2 :

$$p(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}. \quad (\text{C10})$$

Let $\Sigma_+ \subset \mathbb{R}$ be the top 10% of the simulated $\text{SNR}_{\max}^{(\text{mock})}$ values, and let Σ_- be the bottom 90%. Let $\Sigma_0 \in \mathbb{R}$ be the 90th percentile of the $\text{SNR}_{\max}^{(\text{mock})}$ distribution. Then, we choose parameters (μ, σ) to maximize the likelihood function:

$$\begin{aligned} \log \mathcal{L}(x|\mu, \sigma) &= \left(\sum_{x \in \Sigma_+} \log p(x|\mu, \sigma) \right) \\ &\quad + |\Sigma_-| \log \int_{-\infty}^{\Sigma_0} p(x|\mu, \sigma), \end{aligned} \quad (\text{C11})$$

where x denotes mock realizations. This likelihood function has been constructed to fit parameters to the details of the Σ_+ values, while putting all Σ_- values into a single coarse bin.

Figure 16 is a good visual test for goodness of fit, but as a more quantitative test, we compare the upper 10% of the simulated histogram with the Gaussian fit using

a KS test. We find that the two distributions agree to 1σ (and likewise for the other two cases, DESI-BGS and DESI-LRG).

In Table 5, we compute statistical significance for each of the three surveys, in two different ways. The “brute-force” p-value is obtained by counting the number of simulated $\text{SNR}_{\max}^{(\text{mock})}$ values (out of 10^4 total simulations) that exceed $\text{SNR}_{\max}^{(\text{data})}$. The “analytic” p-value is obtained by fitting the top 10% of the simulated $\text{SNR}_{\max}^{(\text{mock})}$ values to a Gaussian distribution, as described above, and evaluating the CDF of the distribution at $\text{SNR}_{\max}^{(\text{data})}$. The brute-force values are either uninformative (for WISE×SCOS), or have large Poisson uncertainties (for the other two surveys), so we have quoted the analytic p-values as our “bottom-line” detection significances throughout the paper.

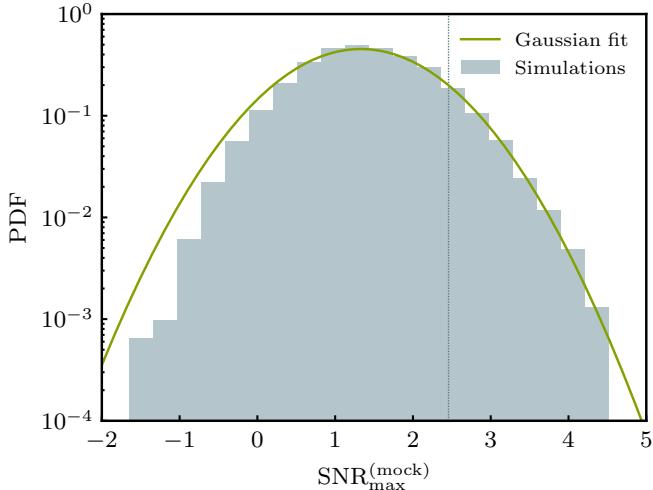


Figure 16. Gaussian fit to the tail of the $\text{SNR}_{\max}^{(\text{mock})}$ distribution from Appendix C. For the top $\sim 10\%$ of the samples (i.e. to the right of the dotted line) the agreement between the fit and the simulations is excellent. This plot is for WISE \times SCOS; the other two cases (DESI-BGS, DESI-LRG) are similar.

Survey	Brute-force	Analytic
WISE \times SCOS	0/10000	2.7×10^{-5}
DESI-BGS	4/10000	3.1×10^{-4}
DESI-LRG	5/10000	4.1×10^{-4}

Table 5. “Brute-force” and analytic p -values, computed as described in Appendix C.