Probing the large-scale clustering of fast radio bursts with CHIME/FRB

by

Masoud Rafiei-Ravandi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Vikram Ravi Assistant Professor, Dept. of Astronomy, Caltech
Supervisor(s):	Kendrick Smith Faculty, Perimeter Institute for Theoretical Physics Adjunct Faculty, Dept. of Physics & Astronomy, Univ. of Waterloo
	Avery Broderick Associate Professor, Dept. of Physics & Astronomy, Univ. of Waterloo Associate Faculty, Perimeter Institute for Theoretical Physics
Internal Member:	Michael Balogh Professor, Dept. of Physics & Astronomy, Univ. of Waterloo
Internal-External Member:	Ghazal Geshnizjani Research Associate Professor, Dept. of Applied Math, Univ. of Waterloo
Other Member(s):	Neal Dalal Faculty, Perimeter Institute for Theoretical Physics Adjunct Faculty, Dept. of Physics & Astronomy, Univ. of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This dissertation is partially the product of co-authored publications:

- In Chapter 1, I present an overview of results that are based on:

    1. Josephy, A. et al. (2021) *No Evidence for Galactic Latitude Dependence of the Fast Radio Burst Sky Distribution.* arXiv:2106.04353 (submitted to ApJ).
    2. Pleunis, Z. et al. (2021) *Fast Radio Burst Morphology in the First CHIME/FRB Catalog.* arXiv:2106.04356 (accepted to ApJ).
    3. CHIME/FRB Collaboration (2021) *The First CHIME/FRB Fast Radio Burst Catalog.* arXiv:2106.04352 (submitted to ApJS).
    4. Bhardwaj, M. et al. (2021) *A nearby repeating fast radio burst in the direction of M81.* ApJL, 910, L18.
    5. CHIME/FRB Collaboration et al. (2020) *A bright millisecond-duration radio burst from a Galactic magnetar.* Nature, 587, 54.
    6. CHIME/FRB Collaboration et al. (2020) *Periodic activity from a fast radio burst source.* Nature, 582, 351.
    7. Fonseca, E. et al. (2020) *Nine New Repeating Fast Radio Burst Sources from CHIME/FRB.* ApJL, 891, L6.
    8. Marcote, B., Nimmo, K., Hessels, J. W. T. et al. (2020) *A repeating fast radio burst source localised to a nearby spiral galaxy.* Nature, 577, 190.
    9. CHIME/FRB Collaboration (2019) *CHIME/FRB Detection of Eight New Repeating Fast Radio Burst Sources.* ApJL, 885, L24.
    10. Josephy, A., Chawla, P., Fonseca, E. et al. (2019) *CHIME/FRB Detection of the Original Repeating Fast Radio Burst Source FRB 121102.* ApJL, 882, L18.
    11. CHIME/FRB Collaboration (2019) *A second source of repeating fast radio bursts.* Nature, 566, 235.
    12. CHIME/FRB Collaboration (2019) *Observations of fast radio bursts at frequencies down to 400 megahertz.* Nature, 566, 230.
    13. CHIME/FRB Collaboration (2018) *The CHIME Fast Radio Burst Project: System Overview.* ApJ, 863, 16.

As a co-author of the listed publications, I contributed mainly to the CHIME/FRB software pipeline. In particular, I led the design and development of code for mitigating radio frequency interference in realtime data.

- Chapter 2 is based on: **Rafiei-Ravandi, M.**, Smith, K. M., and Masui, K. W. (2020) *Characterizing fast radio bursts through statistical cross-correlations.* PRD, 102, 023528.

  As the lead author of this publication, I contributed significantly to the writing, calculations and simulations.

- Chapter 3 is based on: **Rafiei-Ravandi, M.** et al. (2021) *CHIME/FRB Catalog 1 results: statistical cross-correlations with large-scale structure.* arXiv:2106.04354 (submitted to ApJ).

  As the lead author of this publication, I contributed significantly to the writing, analysis, calculations and interpretation.

Abstract

Explaining the nature of extragalactic fast radio bursts (FRBs) has puzzled astrophysicists since 2007. In this thesis, we introduce the CHIME/FRB instrument, an FRB search engine that could solve this puzzle through FRB population studies. We explore CHIME/FRB science results, with an emphasis on the FRB-galaxy correlation. First, we formulate a framework for characterizing FRBs (e.g. by constraining their redshift and host dispersion measure distributions) through angular cross-correlations with large-scale structure. Using this machinery we model, forecast and simulate the FRB-galaxy correlation for two distinct FRB models. Then, we apply this technique to real data, using the first CHIME/FRB catalog along with five photometric redshift catalogs of galaxies. Computing the FRB-galaxy cross power spectrum, 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 $0.3 \lesssim z \lesssim 0.5$. The strength and angular scale of the cross-correlation are consistent with an order-one fraction of CHIME FRBs being in this redshift range, and in the same dark matter halos as the survey galaxies. Finally, we find statistical evidence for a subpopulation of FRBs with large host dispersion measure ($\mathrm{DM_{host}} \sim 400\ \mathrm{pc\,cm^{-3}}$) at $z \sim 0.4$. We show that such large host DMs could be explained by a small population of FRBs near the centers ($r \lesssim 100$ kpc) of large ($M \sim 10^{14} M_\odot$) halos.

Acknowledgments

First and foremost, I would like to thank Kendrick Smith for kindly sharing his knowledge and wisdom with me. I would also like to thank Avery Broderick, Niayesh Afshordi, David Kubizňák, and members of the Examining Committee for their advice and support. I am grateful to the CHIME/FRB Collaboration for providing me with the opportunity to work on designing key components of the CHIME/FRB instrument. Finally, I am thankful to the Perimeter Institute for hosting an exceptional professional and educational environment.

Dedication

To my parents

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Fast radio bursts

Since the first discovery of cosmic radio waves by Karl G. Jansky in the 1930s [48, 49], radio telescopes have transformed our understanding of astrophysical phenomena. From the cosmic microwave background (CMB) at 2.726 K [3] to highly energetic pulsars with brightness temperature $\sim 10^{32}$ K [105], radio waves probe extreme physics in the Universe. Because of atmospheric opacity and quantum noise, ground-based radio observations are constrained to the frequency range $10\,\mathrm{MHz} \lesssim \nu \lesssim 1\,\mathrm{THz}$ [32]. Through this window, radio telescopes see hundreds of thousands of pulses from Galactic pulsars every day. These pulses are typically dispersed due to the intervening cold plasma in interstellar medium (ISM), which delays the arrival time (Δt) of a pulse as a function of frequency (ν):

$$\Delta t = (4.15\ \mathrm{ms}) \left(\frac{\mathrm{DM}}{1\ \mathrm{pc\ cm^{-3}}} \right) \left(\frac{\nu}{1\ \mathrm{GHz}} \right)^{-2} , \tag{1.1}$$

where $\mathrm{DM} = \int dx\, n_e(x)$, which is the *dispersion measure* or integrated column density of free electrons along the line of sight. Given a model for the distribution of ISM gas, dispersion measure can be used as a proxy for distance to pulsars. If the distance is known through other methods (e.g. parallax measurements) in advance, then the observed DM can constrain the ISM model.

Among many pulsars, there are occasionally a few radio bursts that happen to exhibit very high DMs with unknown origin. These random bursts have DMs beyond what we expect from Galactic models. Therefore, their source must be extragalactic! They are

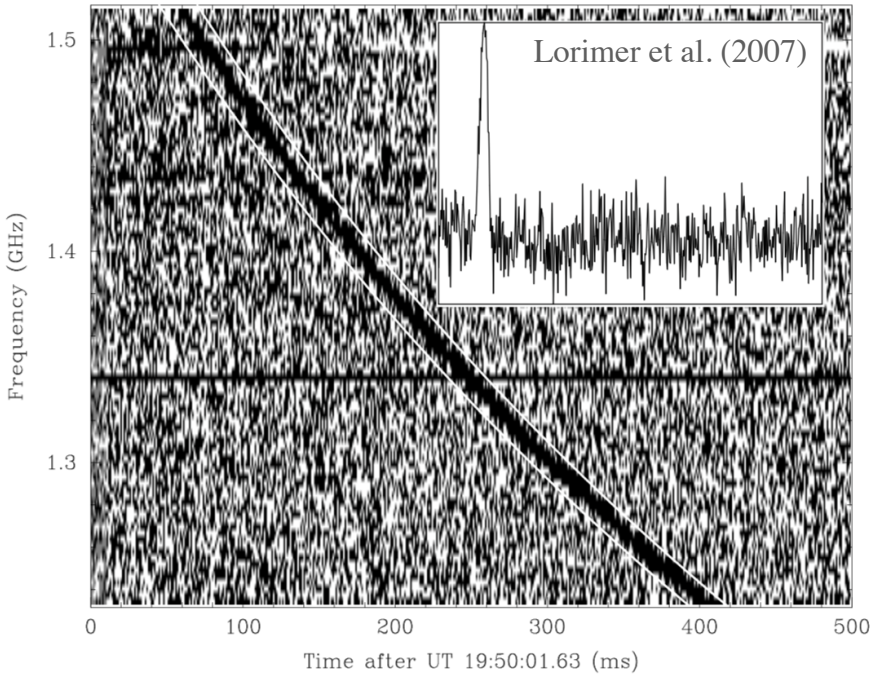

Figure 1.1: Intensity as a function of frequency and time for FRB010724, also known as the Lorimer burst [66]. This so-called *waterfall* plot clearly shows the delayed arrival ($\propto \nu^{-2}$) with DM $= 375 \pm 1$ pc cm^{-3}. The inset is the pulse profile after dedispersion, i.e. summing up the intensity at fixed DM.

extremely energetic ($\sim 10^{36}$ to 10^{42} ergs) and brief (~ 1 ms), which make them an exciting yet challenging target for follow-up observations. These mysterious signals are called fast radio bursts (FRBs). Figure 1.1 shows the first discovered FRB, which was found in archival data from 2001 [66]. We often write the observed DM as a sum of contributions from the Galactic ISM ($\mathrm{DM_{gal}}$), intergalactic medium or IGM ($\mathrm{DM_{IGM}}$), and an unknown host ($\mathrm{DM_{host}}$):

$$\mathrm{DM} \equiv \mathrm{DM_{gal}} + \mathrm{DM_{IGM}}(z) + \mathrm{DM_{host}}\,. \tag{1.2}$$

Assuming that free electrons are uniformly distributed in the IGM, the second term is:

$$\mathrm{DM_{IGM}}(z) = n_{e,0} \int_0^z dz' \, \frac{1+z'}{H(z')}\,, \tag{1.3}$$

where the comoving electron number density is $n_{e,0} = 2.13 \times 10^{-7}\,\mathrm{cm}^{-3}$ and $H(z)$ is the Hubble expansion rate. Thus, the observed DM places an upper bound on the source

redshift (e.g. FRBs with DM $= 1000$ pc cm^{-3} must originate from $z \lesssim 1$).

Besides dispersion, FRBs often exhibit some of the following observables:

- *Faraday rotation:* If the intervening plasma is magnetized, then a linearly polarized FRB can probe the magnetic field through Faraday rotation:

$$\mathrm{RM} \propto \int dx B_{\parallel}(x) n_e(x)\,, \tag{1.4}$$

  where RM is the rotation measure, and $B_{\parallel}(x)$ is the magnetic field strength parallel to the line of sight. If both DM and RM are induced by the same magnetized plasma, then the magnetic field can be estimated as follows:

$$\langle B_{\parallel} \rangle \propto \frac{\mathrm{RM}}{\mathrm{DM}} \tag{1.5}$$

- *Scattering:* A radio pulse can be scatter-broadened due to multipath propagation in a turbulent plasma. Let the inhomogeneous plasma be confined to a thin layer. Then, the scattering timescale (τ) is given by:

$$\tau \propto x \nu^{-4}\,, \tag{1.6}$$

  where x is the distance between the source and observer.

- *Scintillation:* Multipath propagation can also cause an interference pattern, and hence intensity variation at the observer plane.

- *Lensing:* FRB paths can be perturbed by mass and electron overdensities, which result in strong gravitational and plasma lensing respectively.

Figure 1.2 shows the range of spectral luminosity for radio transients. Occupying an unexplored region of the luminosity space, FRB progenitors have puzzled theorists for years; the underlying physical mechanism for generating FRBs is yet to be pinned down (see, e.g. [90] for a non-exhaustive list of models). At the time of this writing, 24 FRB sources have been found to be generating (multiple) fast radio bursts sporadically over long time intervals (e.g. a few pulses in a month). The discovery of repeating FRBs has ruled out progenitor theories involving cataclysmic events, at least for a subpopulation of FRBs. The fact that only a fraction of FRBs come from repeaters may simply be an observational effect. It is possible to imagine that all FRBs repeat with varying ranges of activity, so

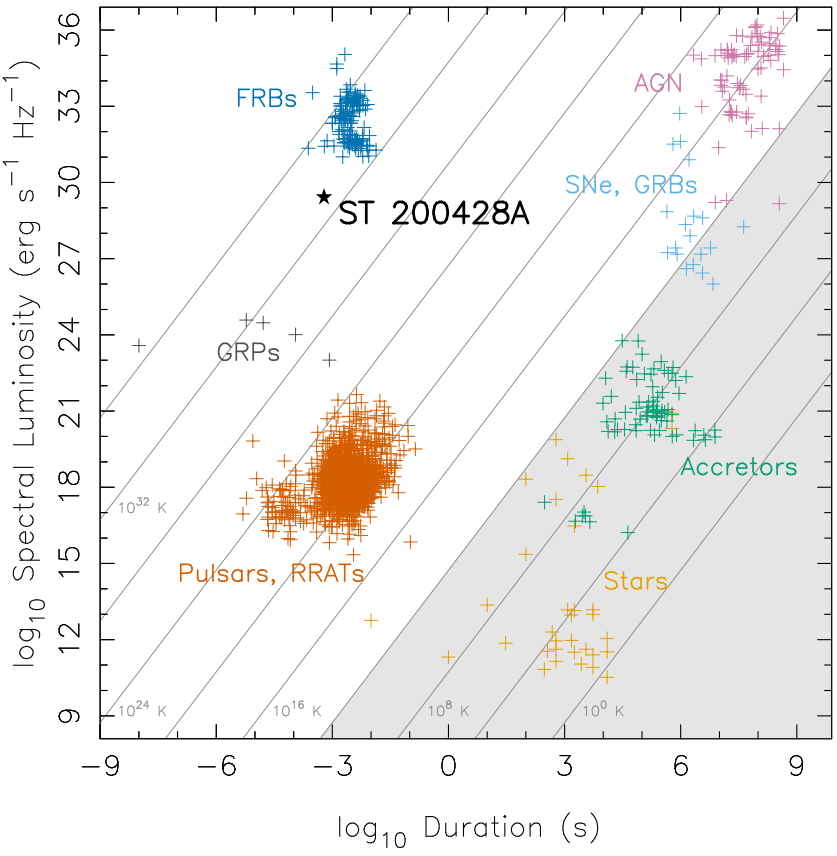

Figure 1.2: Spectral luminosity for radio transients [19], including a sample of FRBs with known distance. Diagonal lines indicate constant brightness temperatures at 1.4 GHz. The gray shaded area contains only the sources that could emit energy incoherently without a relativistic boost. "ST 200428A" marks a giant pulse from the Galactic magnetar SGR 1935+2154, the first example of an FRB in our Galaxy, discovered by CHIME/FRB, STARE2 and ARO in April 2020 (see §1.4).

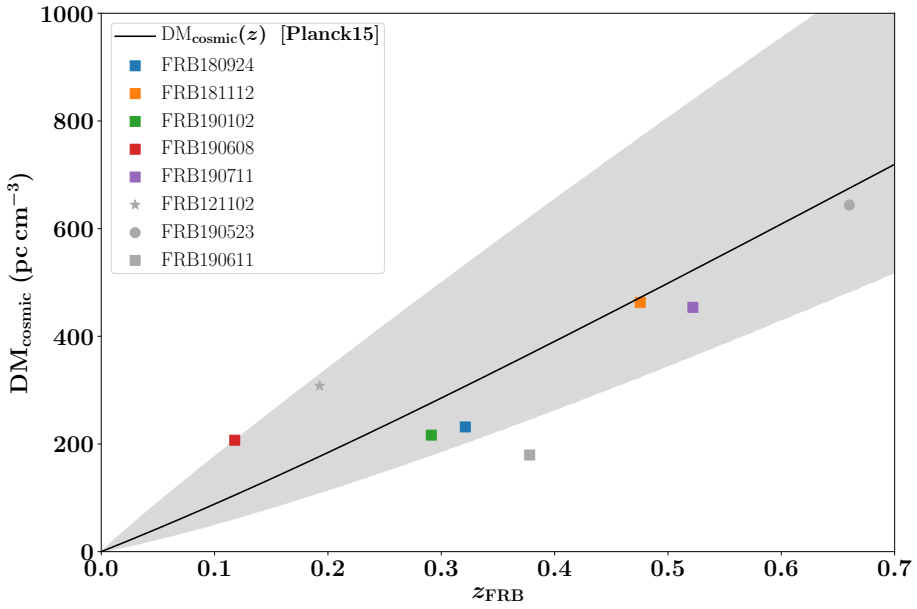

Figure 1.3: The relation between $\rm DM_{IGM}$ and redshift (also known as the Macquart relation; Eq. 1.3) for a sample of localized FRBs from [68]. The gray shaded area represents the uncertainty due to IGM anisotropy.

that after years of observation, all FRBs will eventually become a repeater. On the other hand, repeaters and non-repeaters may originate from two distinct populations of sources.

Repeaters allow for follow-up observations, which are key to understanding the host of FRBs. In the absence of any luminosity gap (see Figure 1.2), could FRBs be explained simply by a population of Galactic pulsars surrounded by highly dense plasma? Pulsars lie preferentially near the Galactic plane, where the expected DM is large along the line of sight. In contrast, FRBs have shown no correlation with any Galactic structures, reinforcing the hypothesis for an extragalactic origin. After over a decade of speculations, a handful of FRBs have finally been localized to other galaxies.

At the time of this writing, 607 FRB sources have been discovered, including 14 FRBs with known host galaxies. Most (492) of the known FRB sources are from CHIME, a new radio telescope which is the focus of this thesis. No FRB has shown any spectral lines that could be used to constrain its redshift. Therefore, FRB redshifts have been determined only through angular association with galaxies (see Figure 1.3). In general, compact radio telescopes, including those with a high mapping speed, are not capable of localizing FRBs to their hosts. Such precise association requires the use of very long baseline interferometry

(VLBI) and optical follow-ups.

Bright FRBs are distributed randomly at a rate of ≈ 800 per day over the full sky [29]. Therefore, they are a rare target for blind searches with a small field of view (e.g. ASKAP [15]). In addition, rapid follow-up observations are nearly impossible for non-repeating FRB sources, which are detected in radio surveys with a large field of view but low angular resolution. Despite all these challenges, the CHIME/FRB instrument (§1.2) has been very successful in increasing the total number of known FRBs by a factor ~ 5 in less than 2 years of operation, enabling FRB population studies for the first time. In §1.4, we will explore recent results from this instrument, which I have helped design, build and maintain during my PhD studies.

1.2 The CHIME/FRB instrument

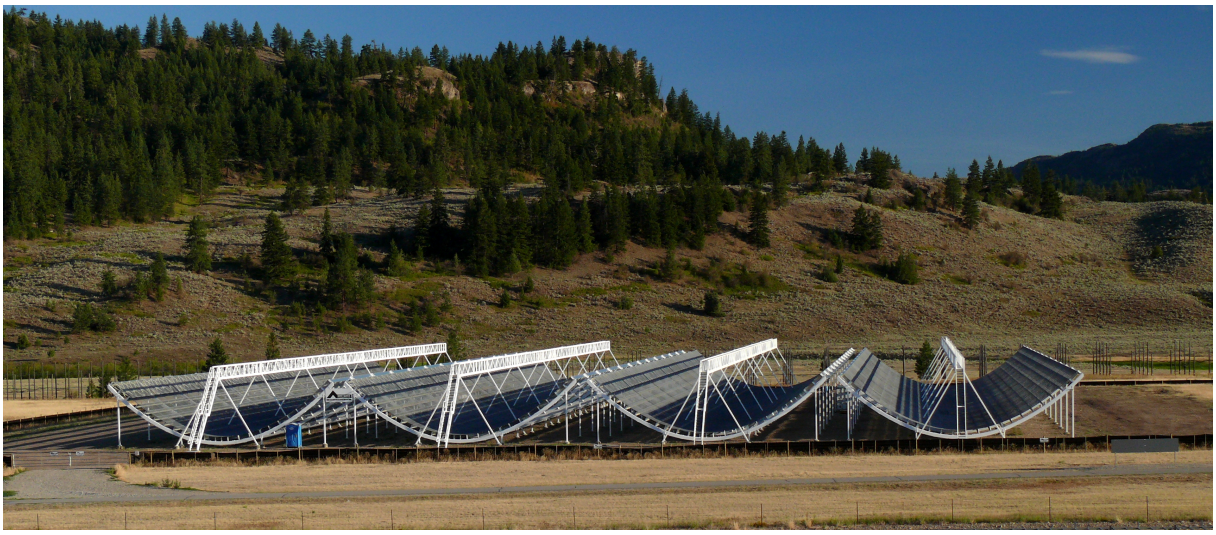

Figure 1.4: Photograph of the CHIME telescope, looking South.

The Canadian Hydrogen Intensity Mapping Experiment (CHIME)[1] was originally designed for mapping the 21-cm radio emission from neutral hydrogen in large-scale structure (LSS). CHIME is a transit telescope, comprising of four 80 m by 100 m cylinders with no moving parts (see Figure 1.4). It is located at the Dominion Radio Astrophysical Observatory (DRAO) in the Okanagan Valley, British Columbia (BC). DRAO is a radio silent

[1] https://chime-experiment.ca

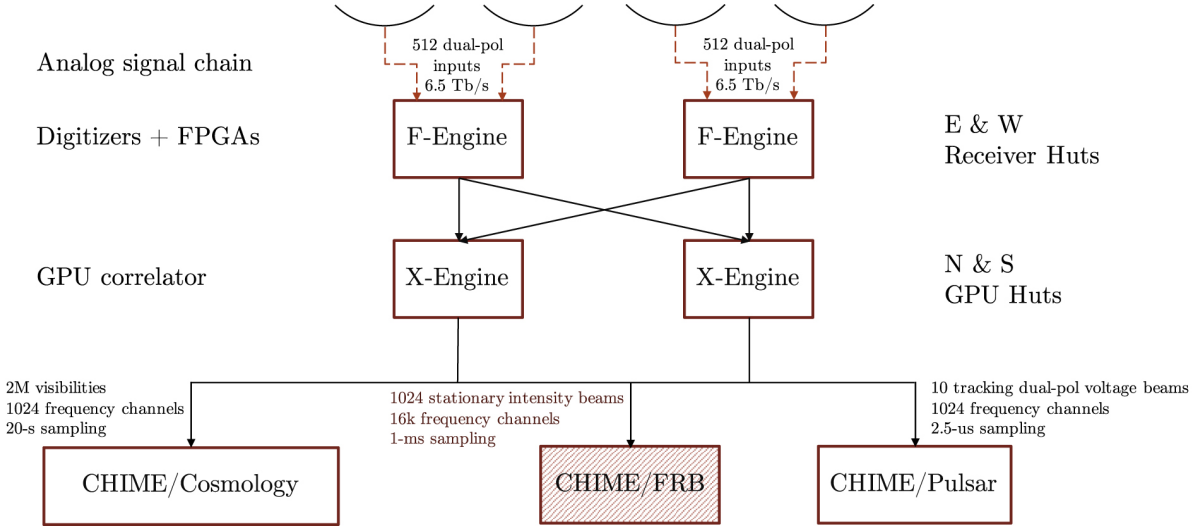

Figure 1.5: Signal path for the CHIME telescope. Each box depicts an RFI-shielded container with computer hardware that was built from scratch. The analog signal chain carries baseband (voltage) data through coaxial cables (dashed arrows) directly from 4×256 antennas, which are equipped with low-noise amplifiers, to the F-Engine. The X-Engine is a massive GPU-based correlator that sends various data products through network fibers (solid arrows) to the science backends at the bottom. The CHIME/FRB backend houses over 2,500 CPU cores primarily for processing frequency-channelized intensity data in L1 (see Figure 1.6).

area within a 30-min driving range from Penticton, BC. The CHIME telescope has three dedicated software backends for 21-cm cosmology, fast radio bursts (CHIME/FRB) and pulsars. Thanks to a coordinated effort by governments and locals, CHIME experiences minimal contamination ($\sim$ 20% of its 400-800 MHz bandwidth) due to radio frequency interference (RFI) from artificial devices (e.g. cellular phones, aircrafts and satellites). Nonetheless, traces of RFI can be extremely problematic for any real-time search in radio astronomy, since RFI is typically orders of magnitude brighter than natural emission from radio sources in the sky. In the first half of my PhD program, I led the design and development of RFI mitigation algorithms (see §1.3) for the CHIME/FRB instrument.

Figure 1.5 shows a diagram of signal chain from antennas to the digitizer (F-Engine), correlator (X-Engine) and the science backends (see [23] for a detailed description). As shown in Figure 1.6, the CHIME/FRB backend is divided into multiple levels. These levels have constantly been running (except during upgrade cycles and extreme weather

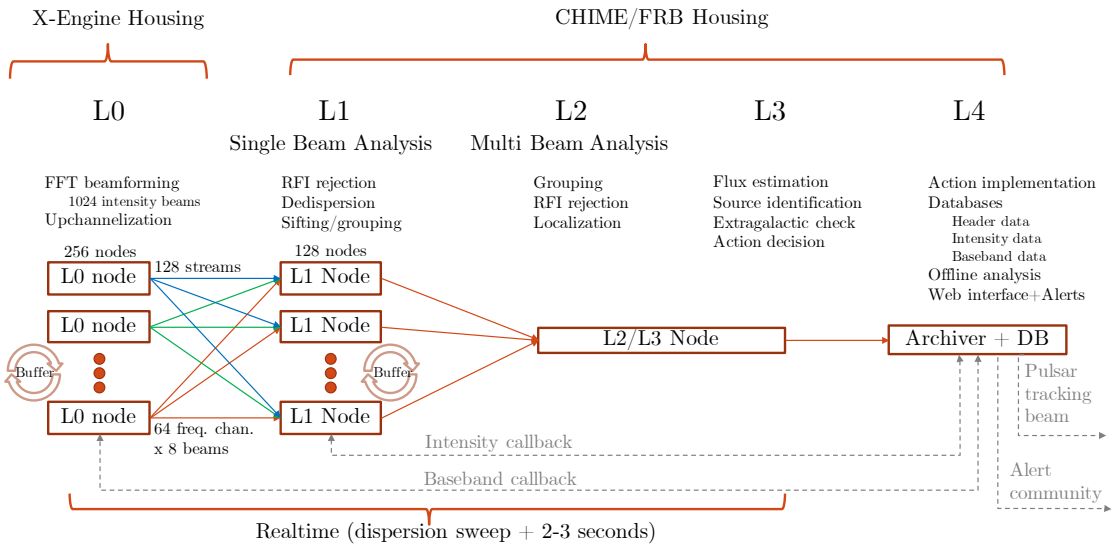

Figure 1.6: Schematic of the CHIME/FRB pipeline.

conditions) with minimal user intervention since July 2018:

- *L0:* Digitized signal is frequency-channelized and beamformed for 4×256 independent beams with $\sim 0.^{\circ}3$ resolution. L0 receives frequency-channelized data at 13 Tb/s from F-Engine, and sends intensity packets at 142 Gb/s to L1.

- *L1:* L0 packets are assembled in 2D intensity arrays with 16k frequency channels and 1ms time resolution. Then, RFI is masked by a chain of operations. Using the masked intensity, FRB search is performed through incoherent dedispersion up to DM = 13,000 $\mathrm{pc\,cm^{-3}}$. The L1 code[2] is state-of-the-art software, which has been optimized for the CHIME/FRB search with ~ 10 sec latency in real time. The output of L1 is a set of events with header information (e.g. DM, SNR, and sky coordinates) that are grouped into coarse-grained candidate events per beam.

- *L2-4:* Next, L1 candidates go through a multi-beam analysis, which groups events based on their (time, DM, sky coordinates). False positives are sent directly to the L4 database, while astrophysical candidates require further estimation and refinement of parameters in L3. L4 contains the CHIME/FRB database along with a user-friendly interface for validating astrophysical candidates.

[2] `rf_pipelines`: https://github.com/kmsmith137/rf_pipelines

1.3 RFI mitigation in L1

In §1.1, we defined the observed dispersion measure (DM) for astrophysical sources. We defined "dedispersion" as the process of summing up intensity values at fix DM (i.e. along a ν^{-2} curve in the (t, ν) space; see Figure 1.1). After dedispersion, an FRB appears as a vertically oriented burst with a single arrival time at all frequencies in waterfall plots. The dedispersion process can be contaminated by RFI, since the summation sees no difference between a real astrophysical signal and e.g. the Long Term Evolution (LTE) telecommunications band at ∼750 MHz. Unmasked RFI can easily swamp the output of dedispersion with thousands of false positives every second! These false positives have random shapes that can resemble a fast radio burst or pulsar in the (t, DM) space (see, e.g. [88]). Therefore, RFI must be removed prior to dedispersion (see Figure 1.7).

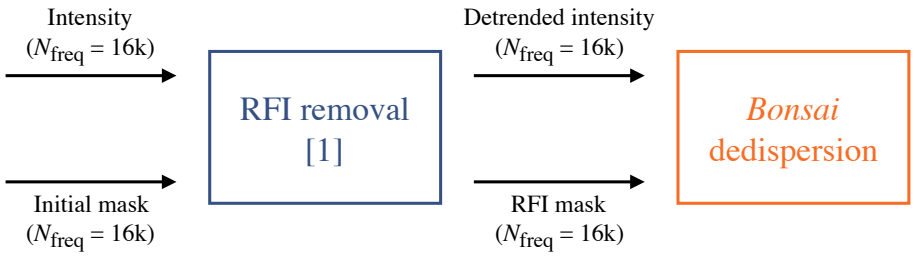

Figure 1.7: Top-level view of the CHIME/FRB L1 pipeline. The RFI removal is performed on an initial set of {intensity, mask} arrays prior to dedispersion, which takes an array of detrended (i.e. effectively high-pass filtered) intensity along with its RFI mask. At the top level, the entire sequence of operations assumes 16k frequency channels. However, the RFI mask is generated through a series of operations at a lower resolution (see Figure 1.8).

The L1 RFI removal pipeline is based on an object-oriented plugin-based library,[3] which allows users to quickly implement new code for processing data. The pipeline can be run online (in real time) or offline (for post-processing). In the CHIME/FRB system, the L1 server is responsible for running the pipeline, which depends on other libraries for various tasks (e.g. the `rf_kernels` library[4] contains a set of fast C++/assembly callables for RFI removal). The L1 framework has a hybrid interface: low-level C++ units can be called through Python. This makes the interface exteremly user-friendly, allowing programmers to exercise algorithmic trials (e.g. a new operation for removing a specfic type of RFI).

[3] `ch_frb_rfi`: https://github.com/mrafieir/ch_frb_rfi
[4] `rf_kernels`: https://github.com/kmsmith137/rf_kernels

Figure 1.8 illustrates the top-level view of RFI removal in L1, which contains a set of {intensity, mask} arrays with 16k frequency channels. The "initial mask" carries binary information about the state of L0 packets. RFI is masked inside a "sub-pipeline", which takes a set of downsampled {intensity, mask} arrays with 1k frequency channels. The output of the sub-pipeline is a mask that is subsequently upsampled to 16k frequency channels for a final "detrending" stage prior to dedispersion. Unlike operations on full 16k frequency channels (e.g. the final detrending stage), sub-pipeline operations are computationally fast on 1k frequency channels (see Figure 1.9). This down/upsampling logic is essential for the CHIME/FRB instrument, which requires ∼100 operations for mitigating RFI in real time.

Figure 1.9 shows the data flow through various operations inside the sub-pipeline. We define the fundamental unit of operation (e.g. an object that masks 5σ intensity outliers) as a "transform". The first transform is a "static mask", which simply flags RFI-contaminated intensity data with known frequency (e.g. 730.1–755.9 MHz), by setting corresponding channels to zero in the mask array. Using the updated mask, we then start clipping intensity outliers, by dynamically assigning zeros to the running mask array. Next, we high-pass filter intensity by fitting and subtracting polynomials (i.e. detrending) along the time and frequency axes. Finally, we update the mask by applying the same set of clipping transforms to the detrended intensity.

The output of the CHIME/FRB "sub-pipeline" is an RFI mask based on clipping and detrending through ∼100 transforms. Since each transform has a set of internal parameters, the grand design of an entire sub-pipeline becomes a complicated problem with hundreds of variables. Our goal is to design an RFI algorithm that can simultaneously minimize the false positive rate, false negative rate, and computational cost. Ideally, one would find the best solution by minimizing a global cost function, based on fully automated transforms. Nevertheless, we took an empirical approach for the CHIME/FRB instrument; we designed the CHIME/FRB L1 RFI transform chain by trial and error, based on inspecting hundreds of hours of saved intensity data. Initially, we designed a sequence of transforms that could effectivly eliminate false positives in days of incoherent-beam data, while signals from bright pulsars (e.g. PSR B0329+54) were recovered. Then, we fine-tuned the transform chain using beamformed data. Thus far, our results have been extremely promising. We are still in the process of improving our solution, planning to automate and optimize the algorithm through machine learning.

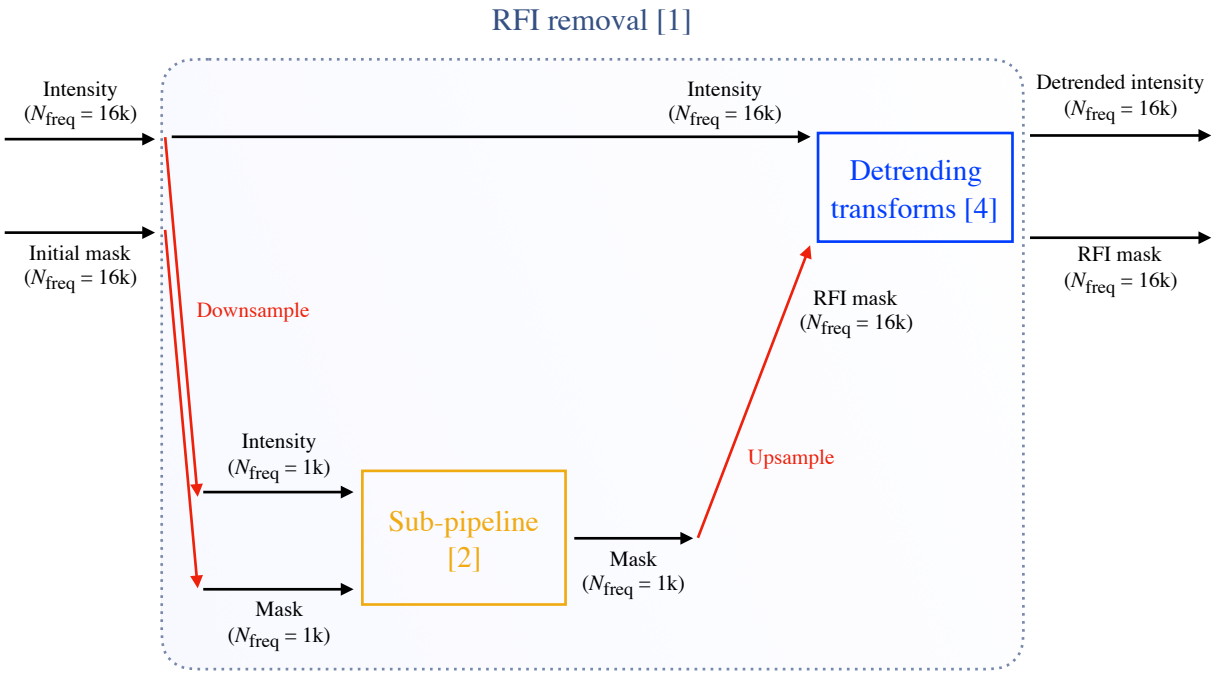

Figure 1.8: RFI removal in L1. In order to minimize the overall computational cost, input data with 16k frequency channels are downsampled to N_{freq}=1k. The "sub-pipeline" contains a sequence of transforms such that its initial input is the downsampled {intensity, mask} arrays, and its final output is a mask whose zeroed entries match with RFI patterns in the downsampled intensity data. Once the sub-pipeline converges, its output mask is upsampled to match (in frequency resolution) with the detrended intensity data (16k) for the dedispersion transform.

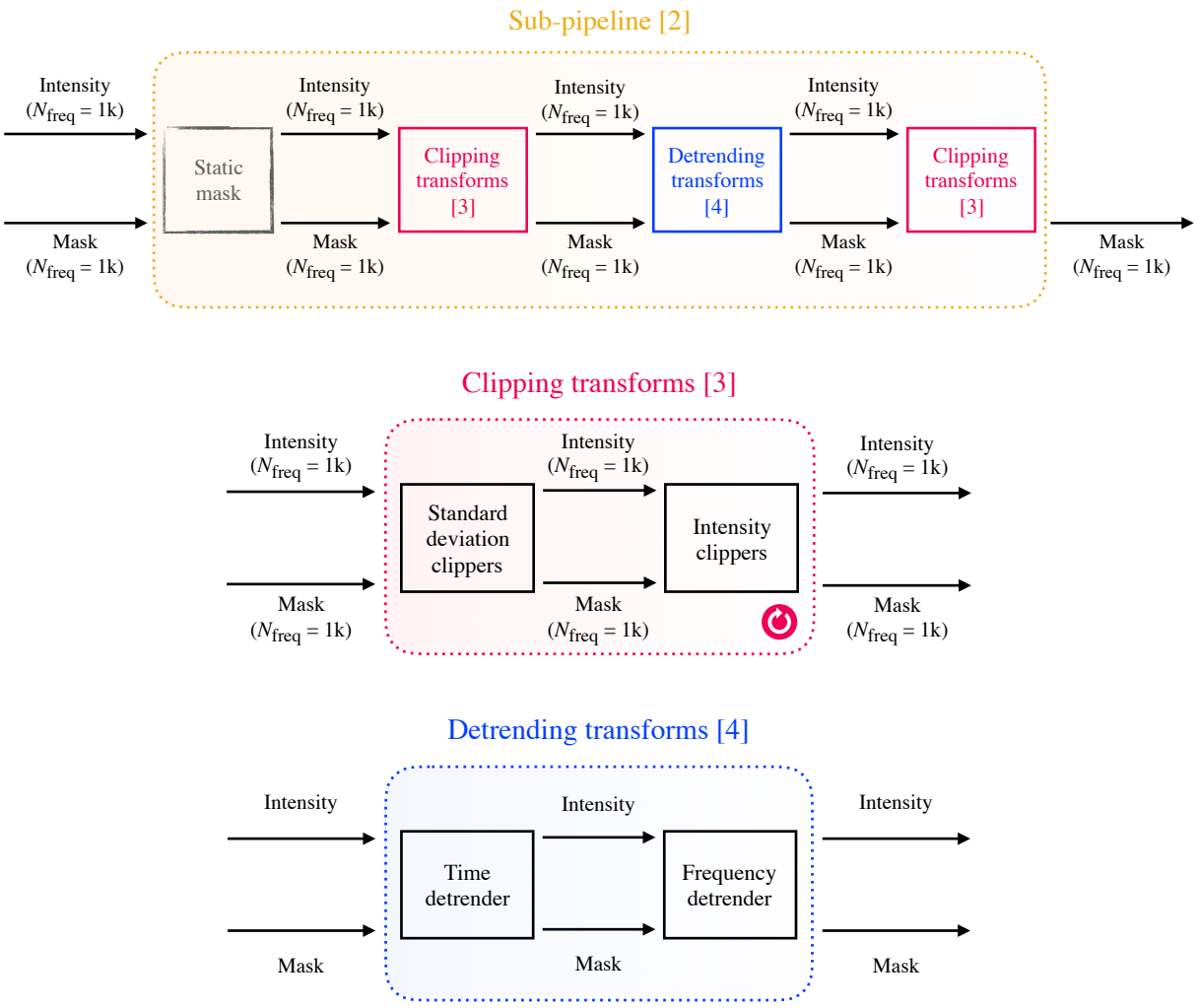

Figure 1.9: Sequence of transforms inside the CHIME/FRB L1 sub-pipeline. "Clipping transforms" flag intensity outliers, by assigning zeros to the running mask. The clipping transform chain is iterated a few times before passing processed {intensity, mask} arrays to the next operation block in a long chain of transforms. "Detrending transforms" effectively high-pass filter the intensity array along time and frequency.

1.4 CHIME/FRB science results

This section presents a brief overview of science results that I have co-authored with the CHIME/FRB Collaboration. The CHIME/FRB instrument saw its first light in 2017. In 2018, we reported on the detection of the first FRB in the CHIME band (400–800 MHz) [22]. At the time, only ~ 50 FRBs had been discovered at frequencies $\nu \gtrsim 750$ MHz. The first CHIME/FRB discovery proved the presence of FRBs at frequencies $\nu \lesssim 750$ MHz.

- In 2019, we published two papers [26, 24] on the detection of 13 new FRBs, including a new repeating source, in the CHIME band while the instrument was still in its pre-commissioning phase. Figure 1.10 shows these detections, based on which we ruled out spectral cuttoffs, e.g. due to free-free absorption [100], for FRBs in the CHIME band. Operating at such low frequencies (~ 400 MHz), we made the most precise measurements of FRB scattering tails (see Eq. 1.6).

- Figure 1.11 shows the new repeating source. It is the second repeating FRB (R2) after FRB121102 (R1), which was found by [119, 120] in survey data from 2012. The detection of R2 proved that repeaters do exist in the CHIME band. Later in 2019, we detected a single burst from R1 at frequencies ∼600 MHz [51]. The burst consisted of downward drifting subbursts, with an effective DM of 563.06 ± 0.5 pc cm^{-3}, which was ∼1% larger than its value from 2012. This change could be either intrinsic (e.g. DM perturbations at the host of emission) or due to systematic biases (e.g. different fitting routines for a morphologically complex FRB).

- Later in [25, 41], we reported on the discovery of 17 new repeating FRB sources in the DM range 103.5–1380 pc cm^{-3}. Using the relatively large population of CHIME FRBs, we found that repeaters are generally wider (i.e. larger effective pulse width) than non-repeaters, and that they often have complex morphologies, including downward drifting (see, e.g. Figure 1.12).

- Among all CHIME/FRB repeaters, FRB180916 (R3) has been the most fruitful source. R3 was the second repeating FRB (after R1) with available polarization data [25]. Figure 1.13 shows its Faraday rotation measure (Eq. 1.4), which is modest compared to R1. In 2020, using the European VLBI Network (EVN) we localized R3 to a star-forming region inside a spiral galaxy (see Figure 1.14) at redshift $z = 0.0337$ [72].

- CHIME/FRB scans the sky constantly, making it an excellent instrument for identifying temporal patterns in radio transients. Monitoring R3 for more than a year, we

discovered periodic activity from this special source [28]: R3 exhibits periodicity in 5-day phase windows, with a period of 16.35 days (see Figure 1.15). This is the first discovery of a periodic phenomenon in an FRB.

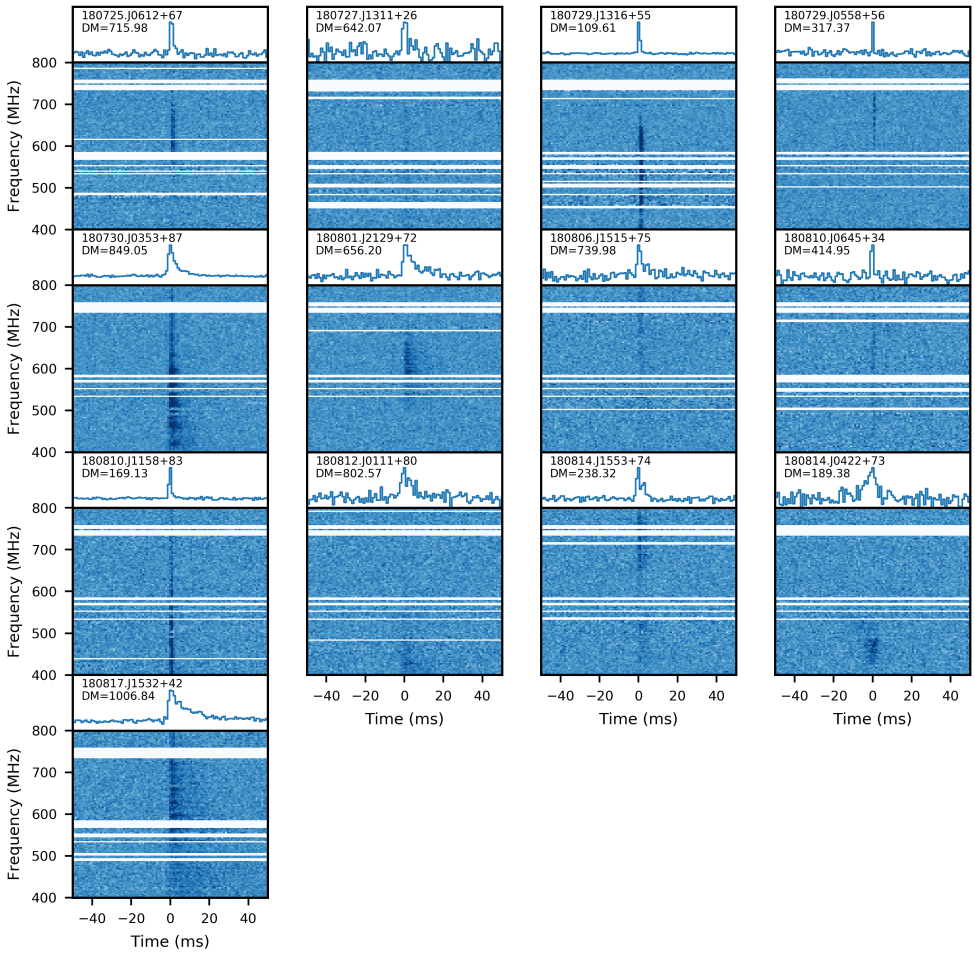

Figure 1.10: Waterfall plots for the first batch of (dedispersed) FRBs, discovered by the CHIME/FRB instrument [26]. White bands indicate RFI-contaminated channels that were masked.

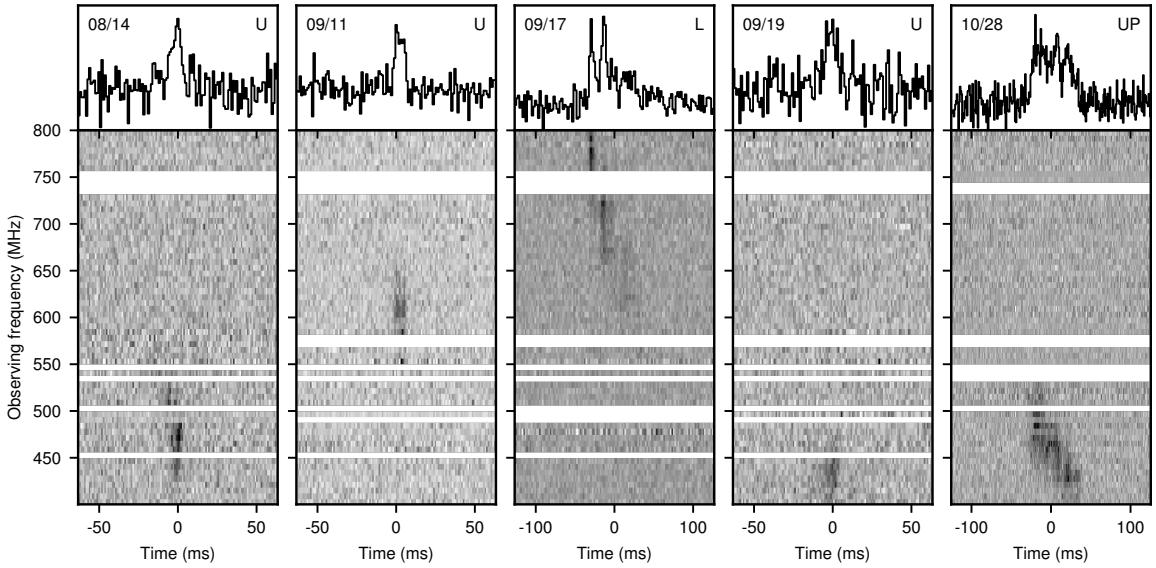

Figure 1.11: Waterfall plots for the second repeating FRB (R2) [24]. The rightmost panel is a detection by the CHIME/Pulsar backend. A DM of 189.4 $\mathrm{pc\,cm^{-3}}$ is assumed throughout. "U" and "L" refer to upper and lower (anti-podal) transits of the source with respect to the North Celestial Pole.

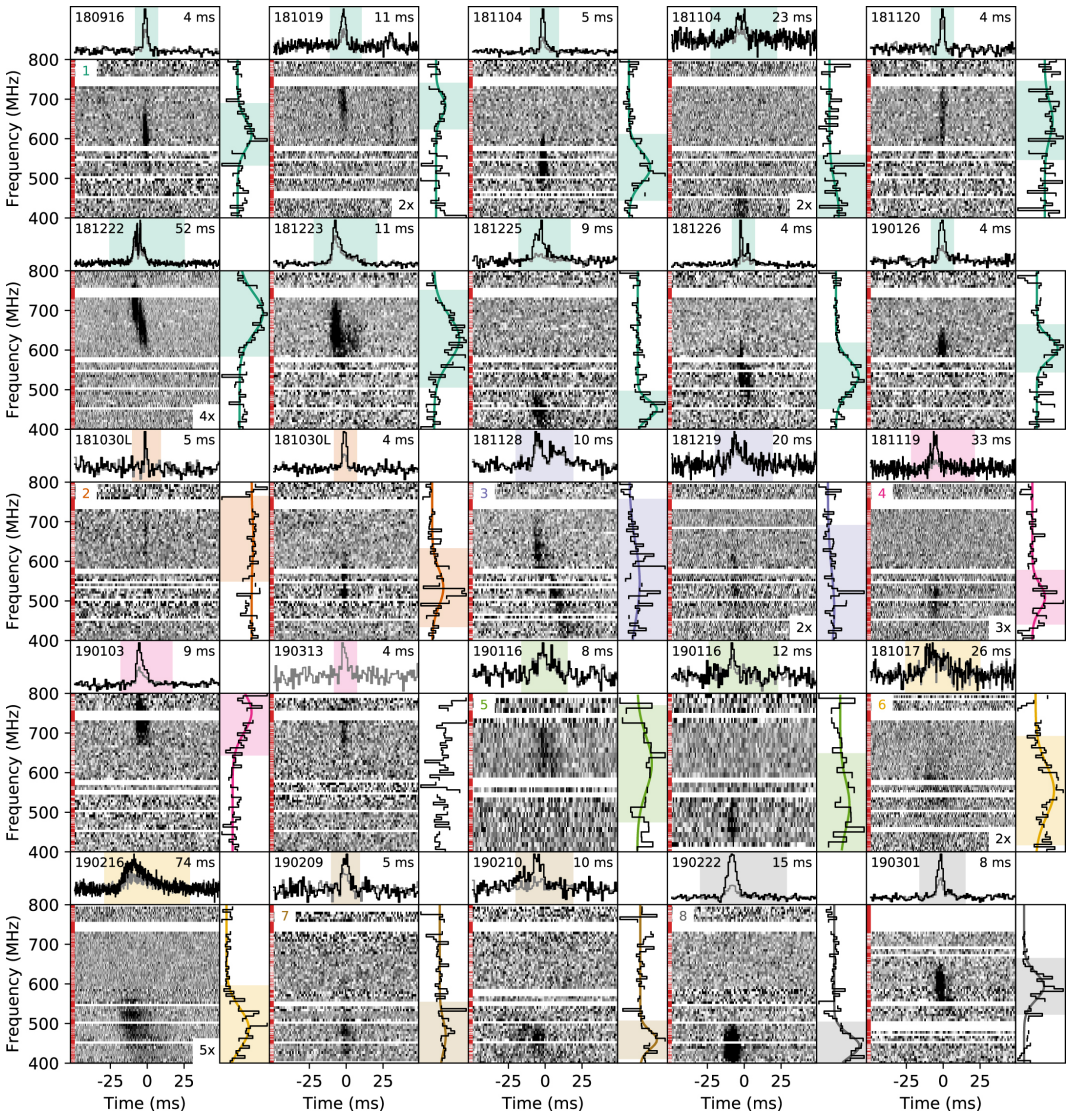

Figure 1.12: A sample of repeating FRBs with a wide range of morphologies [25]. Colors refer to unique FRB sources.

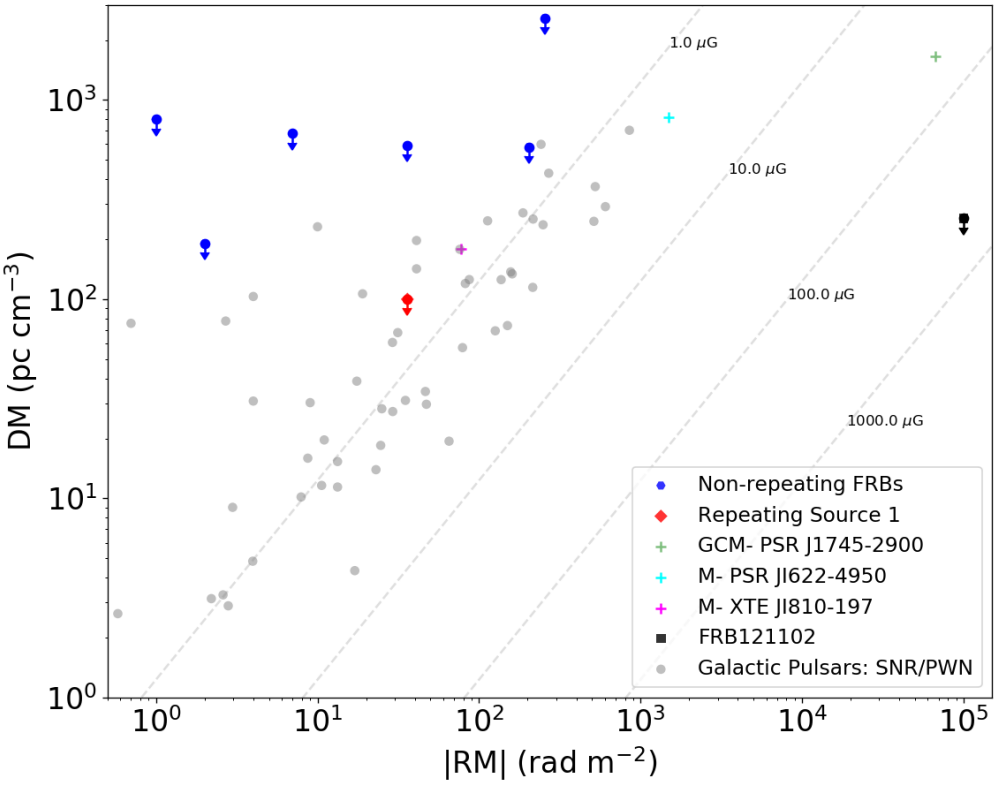

Figure 1.13: Dispersion measure versus rotation measure for a sample of astrophysical transients with polarization data. "Repeating source 1" refers to FRB180916 (R3), which shows a modest RM [25]. FRB121102 is R1 whose extreme RM hints at a host with highly magnetized plasma [80] (see Eq. 1.5).

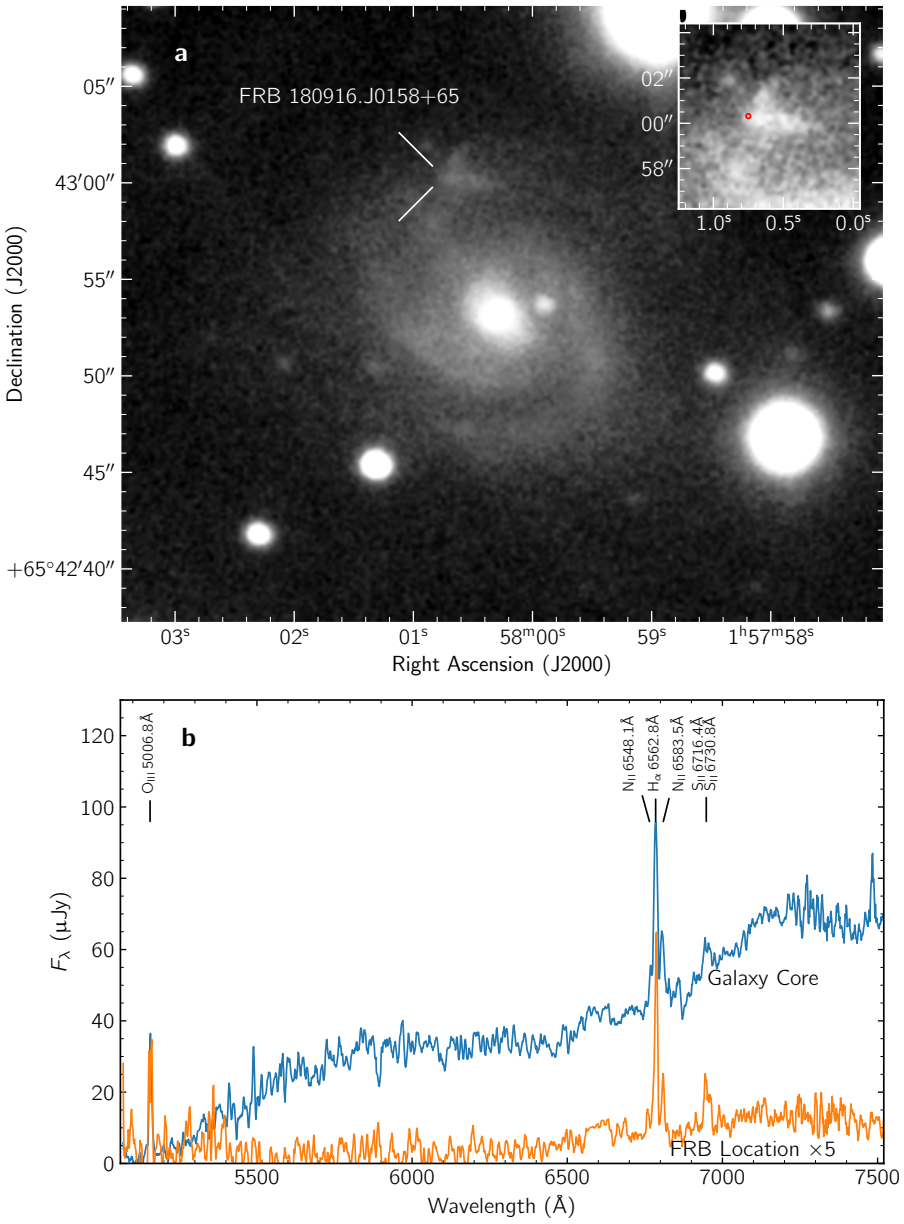

Figure 1.14: FRB180916 (R3) and its host galaxy [72]. *Top panel:* Gemini-North image in r' passband. *Bottom panel:* Star-forming signatures in spectral lines.

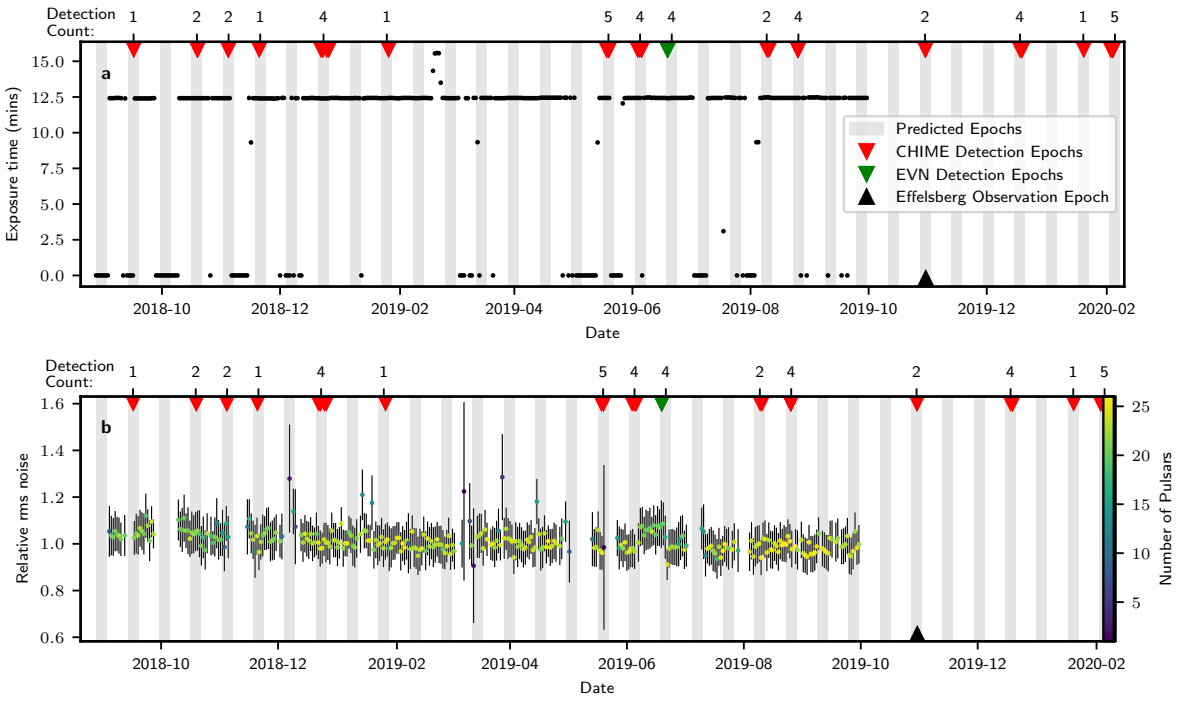

Figure 1.15: CHIME/FRB exposure timeline and periodic activity of FRB180916 (R3) in 5-day phase windows (gray shaded bands), with a period of 16.35 days [28].

- On 28 April 2020, an FRB lit up 93 of the 1024 CHIME/FRB beams! Using intensity data we localized the source to the vicinity of the Galactic magnetar SGR1935+2154 [27]. Figure 1.16 shows the burst, which was also detected by the Algonquin Radio Observatory (ARO) [27] in Ontario and the STARE2 instrument [19] in California. These detections were during an active period of X-ray emission from the magnetar [78, 62, 104]. The burst was ∼1000 times brighter than radio bursts from ∼30 known magnetars. If it were to originate from a nearby galaxy, then it would be indistinguishable from a typical FRB. This is the first Galactic FRB, which mostly fills the luminosity gap between FRBs and giant radio pulses from magnetars (see Figure 1.2).

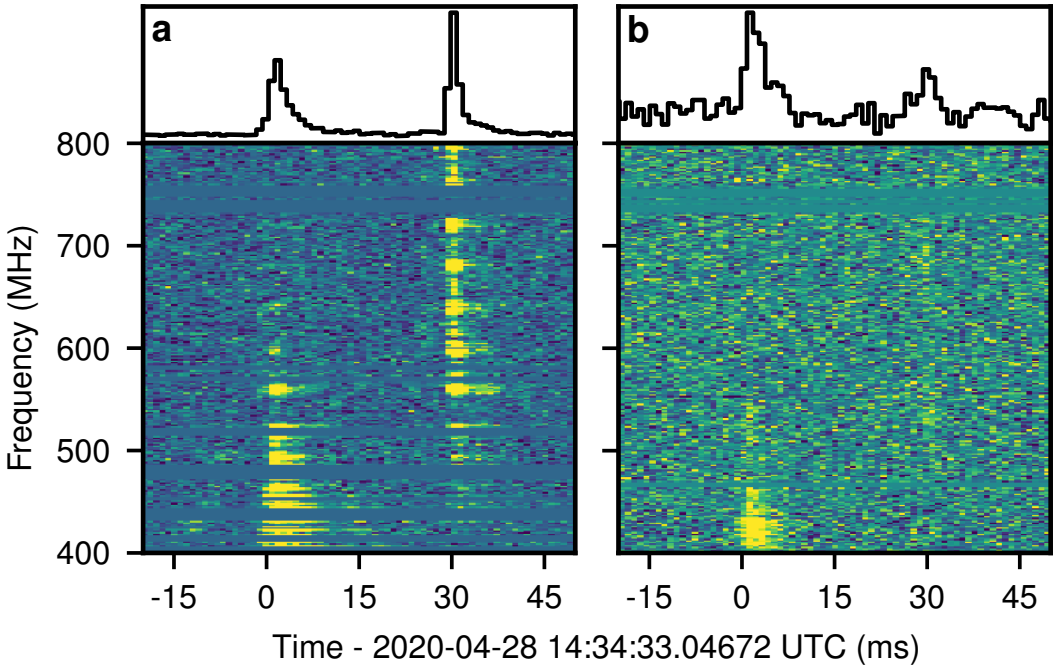

Figure 1.16: CHIME/FRB (left) and ARO (right) detections of a radio burst from the Galactic magnetar SGR1935+2154.

- Earlier in 2021, we localized the repeating FRB 20200120E with DM=87.82 pc cm^{-3} to a small area (≈ 14 sq. arcmin) in the M81 galaxy group [16] (see Figure 1.17). This discovery was made through CHIME/FRB baseband recording system, which

allows for localizing the host of bright FRBs with a low DM ($\lesssim 100$ pc cm^{-3}). FRB 20200120E was subsequently localized to an M81 globular cluster by the EVN [56].

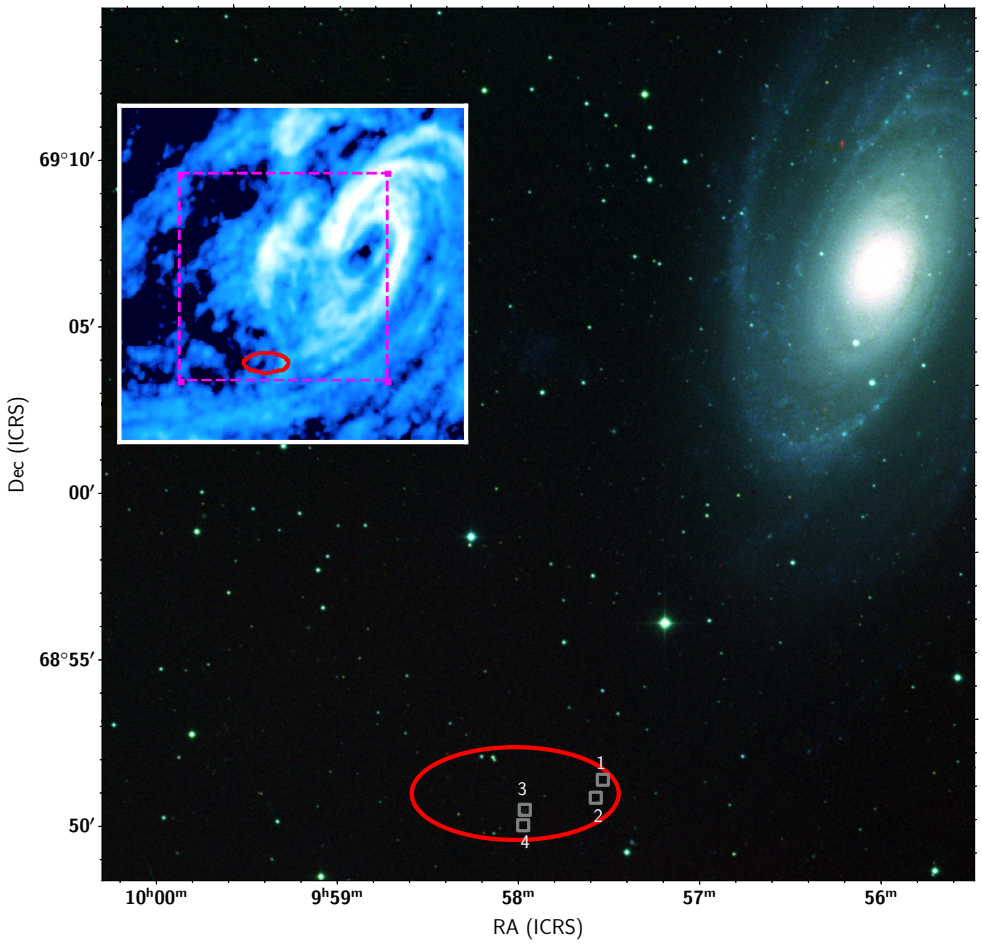

Figure 1.17: CHIME/FRB localization of FRB 20200120E (red ellipse) in the M81 galaxy group, imaged by the Digital Sky Survey (DSS). The inset shows the 21-cm line emission [31]. FRB host candidates are marked by gray squares. See [16] for more details.

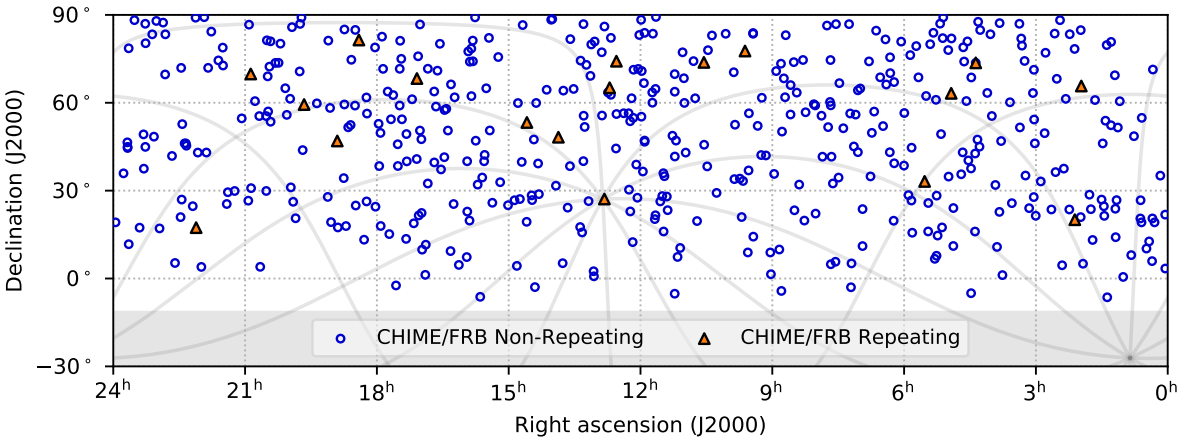

Figure 1.18: Mercator projection of CHIME/FRB Catalog 1 [29] sources in equatorial coordinates. The shaded region (declination $< -11°$) is outside the CHIME/FRB coverage.

- The CHIME/FRB instrument sees a few FRB candidates every day. In about a year of operation, between 25 July 2018 and 2 July 2019, we detected 535 verified FRBs, including 61 bursts from repeating sources. We published this large sample as a catalog in [29], increasing the number of known FRB sources by a factor ~ 4. Figure 1.18 shows the sky distribution of the first CHIME/FRB catalog (or Catalog 1).

- CHIME/FRB Catalog 1 enables FRB population studies, since the entire sample was subject to the same instrumental selection function. In [29], we characterized our selection function through real-time pulse injections. Figure 1.19 shows an example plot before and after the correction for selection effects: we find statistical evidence for a subpopulation of unobserved FRBs with a scattering time at 600 MHz in excess of 10 ms.

- In [50], we showed that the sky distribution of CHIME FRBs has no correlation with the Galactic plane after accounting for selection biases.

- In another companion paper [91], we compared the two populations of repeaters and non-repeaters, and found that repeaters generally have a larger width and narrower bandwidth. In addition, using intensity profiles as FRB fingerprints, we found that all CHIME/FRB sources can be divided into four morphologically distinct categories (see Figure 1.20), which could help constrain progenitor models and have implications for utilizing FRBs as standard probes.

- Finally, I led a project on FRB-LSS correlations, whose theory [95] and analysis [94] are presented in §2 and §3, respectively. In summary, we have detected a statistically significant correlation between CHIME FRBs and photometric redshift catalogs of galaxies. In addition, we have found statistical evidence for a population of FRBs with large host dispersion measure. These results will help answer key questions about the cosmological origin of fast radio bursts.

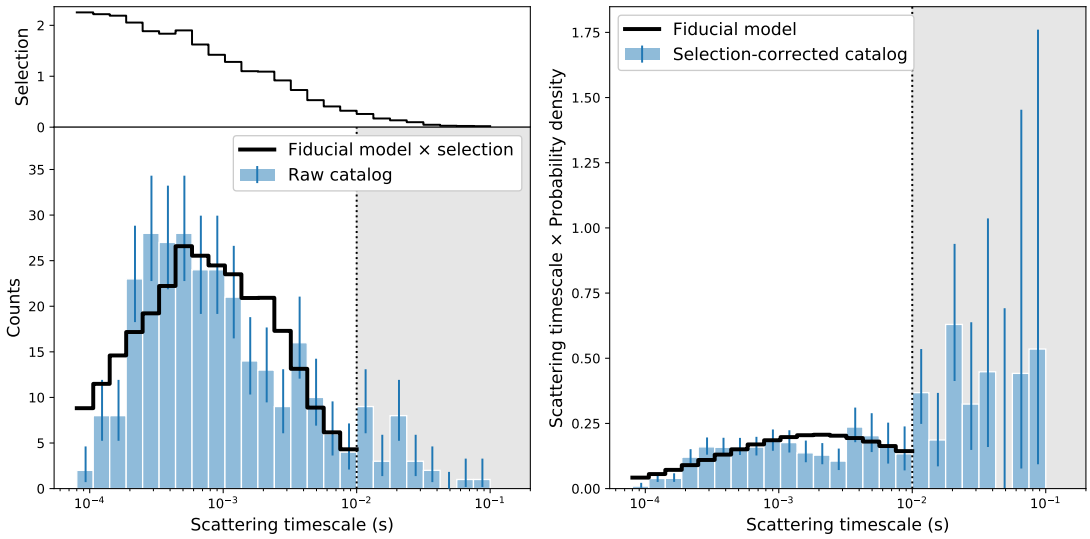

Figure 1.19: Scattering time (scaled to 600 MHz) distribution before (left panel) and after (right panel) correcting for CHIME/FRB instrumental biases [29]. The tail of selection-corrected distribution in the shaded region hints at a subpopulation of highly scattered ($\tau \sim 10$ ms) FRBs that have yet to be observed.

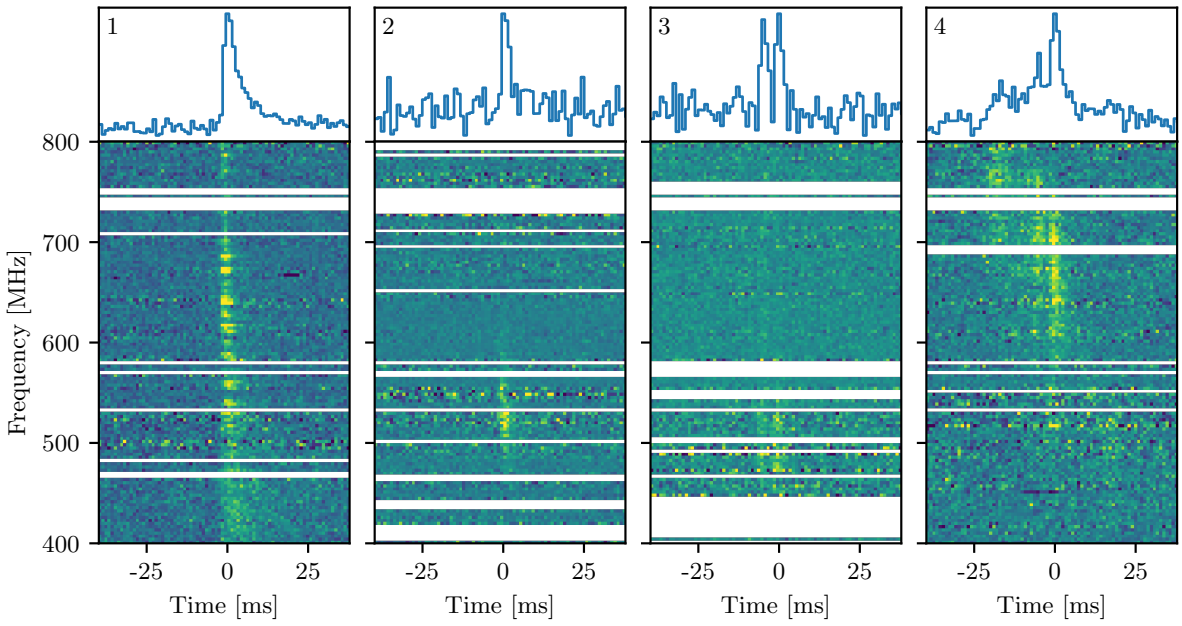

Figure 1.20: Various types of morphology in CHIME/FRB Catalog 1 [91]. From left to right: simple broadband, simple narrowband, temporally complex, and downward drifting.

Chapter 2

Characterizing fast radio bursts through statistical cross-correlations

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

2.1 Introduction

Fast radio bursts (FRBs) are an astrophysical transient whose origin is not yet understood. Since initial discovery in 2007 [66], interest in FRBs has grown, and explaining the FRB

phenomenon is now a central unsolved problem in astrophysics (see [54, 90, 87] for recent reviews).

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

$$\begin{aligned} (\text{Delay}) &= (\text{DM}) \left(\frac{e^2}{2\pi m_e c} \right) \nu^{-2} && (2.1) \\ &= (4.15 \text{ ms}) \left(\frac{\text{DM}}{1 \text{ pc cm}^{-3}} \right) \left(\frac{\nu}{1 \text{ GHz}} \right)^{-2} && (2.2) \end{aligned}$$

where

$$\text{DM} \equiv \int n_e(x)\, dx\,. \tag{2.3}$$

FRBs are a population of dispersed pulses whose observed DM significantly exceeds the maximum Galactic column density DM_{gal} (inferred from a model of the Galaxy [34, 126]). On most of the sky, DM_{gal} is ≤ 50 pc cm^{-3}, and FRBs are regularly observed with DM $\gtrsim$ 1000 pc cm^{-3}. From the outset, the large DM suggested that FRBs were extragalactic, although on its own the large DM could also be explained by a Galactic event with a large local free electron density. As more FRBs were observed, their sky distribution was found to be isotropic (i.e. not correlated with the Galactic plane), conclusively establishing an extragalactic origin.

At the time of this writing, 92 FRB discoveries have been published (according to FRBCAT [86], `frbcat.org`). Ten of these FRBs are "repeaters", meaning that multiple pulses have been observed from the same source [120, 109, 24, 25]. Nine of the repeaters were discovered by the CHIME/FRB instrument, and a much larger sample of non-repeating FRBs from CHIME/FRB is expected soon. (The authors are members of the CHIME/FRB collaboration, and forecasting the scientific reach of CHIME/FRB was the main motivation for this paper.)

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

FRBs do not have spectral lines, so FRB redshifts cannot be directly determined. When an FRB is observed, an upper bound on its redshift z can be inferred from its DM

[1] Jansky (Jy) is a non-SI unit for spectral flux density: $1\,\text{Jy} = 10^{-26}\,\text{W m}^{-2}\,\text{Hz}^{-1}$

as follows. We write the total DM of an FRB as the sum of contributions from our galaxy, the intergalactic medium (IGM), and the host galaxy:

$$\mathrm{DM} = \mathrm{DM}_{\mathrm{gal}} + \mathrm{DM}_{\mathrm{IGM}}(z) + \mathrm{DM}_{\mathrm{host}} \tag{2.4}$$

where the IGM contribution is related to the FRB redshift as:

$$\mathrm{DM}_{\mathrm{IGM}}(z) = n_{e,0} \int_0^z dz' \, \frac{1+z'}{H(z')} \tag{2.5}$$

where $n_{e,0}$ is the comoving electron number density and $H(z)$ is the Hubble expansion rate. If we assume that $\mathrm{DM}_{\mathrm{gal}}$ is known precisely and subtracted, then the inequality $\mathrm{DM}_{\mathrm{host}} \geq 0$ implies an upper bound on z. A $\mathrm{DM} = 1000\ \mathrm{pc\,cm^{-3}}$ FRB must satisfy $z \lesssim 0.95$, and a $\mathrm{DM} = 3000\ \mathrm{pc\,cm^{-3}}$ FRB satisfies $z \lesssim 3.08$. However, an alternative hypothesis is that FRBs are at much lower redshifts, and have large host DMs.

Three FRBs have been observed in long-baseline interferometers with sufficient angular resolution to uniquely identify a host galaxy, and thereby determine a redshift [21, 71, 122, 9, 98]. The inferred redshifts are $z = 0.19$, 0.32, and 0.66. These observations suggest that most of the DM is IGM-related, but with only three data points it cannot be concluded that this is true for the entire population.

Host galaxy associations are a powerful way to determine FRB redshifts, but require angular resolution around 1 arcsecond or better [39]. Unfortunately, most telescopes capable of finding large numbers of FRBs have angular resolution much worse than this. In particular, for most of the CHIME/FRB sources, the angular resolution is either $\approx 1'$ or $\approx 10'$, depending on whether baseband data are available for the event [73, 23, 25].

In this chapter, we study the following question. Given a catalog of FRBs whose resolution is insufficient for host galaxy associations on a per-object basis, is it possible to associate FRBs and galaxies on a *statistical* basis? To make this question precise, we model the angular cross power spectrum C_ℓ^{fg} between the FRB and galaxy catalogs and forecast its signal-to-noise ratio (SNR). The SNR turns out to be surprisingly large. For example, given a catalog of 1000 FRBs with $1'$ resolution, and the photometric galaxy catalog from SDSS-DR8 [4], we find an SNR of 25–100, depending on the FRB redshift distribution.

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

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

This chapter overlaps significantly with work in the galaxy clustering literature on "clustering redshifts" [76, 77, 96, 58, 85]. This term refers to the use of clustering statistics to determine the redshift distribution of a source population, by cross-correlating with a galaxy catalog.

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

Propagation effects arise from several distinct physical mechanisms: dispersion, scattering, and plasma lensing. In this chapter, we will analyze the dispersion case in detail, leaving the other cases to future work. The propagation effects which we will explore have some similarity with magnification bias in galaxy surveys (see e.g. [45] and references therein).

We also clarify which properties of the FRB population are observable via cross-correlations. It is well known that on large scales ("2-halo dominated" scales), the only observable is $(b_f dn_f/dz)$: the product of FRB redshift distribution dn_f/dz and the large-scale clustering bias $b_f(z)$. We find that there is an analogous observable $(\gamma_f dn_f/dz)$ which determines the FRB-galaxy correlation on smaller ("1-halo dominated") scales. The quantity $\gamma_f(z)$ measures the degree of similarity between the dark matter halos which contain FRBs and galaxies, and is defined and discussed in §2.4.

This chapter is complementary to previous works which have considered different FRB-related clustering statistics. In [74], the 3D clustering statistics of the FRB field were studied, using the DM as a radial coordinate. This is analogous to the way photometric galaxy surveys are analyzed in cosmology. Here we generalize to the cross-correlation between the FRB field and a galaxy survey. The FRB-galaxy cross-correlation has higher SNR than the FRB auto-correlation, since the number of galaxies is much larger than the

number of FRBs. Whereas [74] was entirely perturbative, we perform both perturbative calculations and non-linear simulations using a halo model. In addition we consider two propagation effects: DM shifting and completeness (to be defined below), whereas [74] considered only the former.

Another idea that has been considered is to cross-correlate a 2D map of FRB-derived dispersion measures with galaxy catalogs, to probe the distribution of electrons in dark matter halos [75, 113, 99, 81, 69]. The cross-correlation of DM vs galaxy density is related to the DM moment of the statistic $C_\ell^{fg}(z, \mathrm{DM})$ considered here. Therefore, our statistic contains a superset of the information in the statistic considered in these works.

In [63], a cross-correlation was observed between 2MPZ galaxies at $z \sim 0.01$, and a sample of 23 FRBs from ASKAP operating in "fly-eye" mode with $10'-60'$ angular resolution [10, 110]. This measurement is seemingly at odds with the three FRB host galaxy redshifts which imply a much more distant population. In the very near future, FRB catalogs will be available with much higher number density and better angular resolution, so it will be possible to measure the cross-correlation at higher SNR, and push the measurement to higher redshift. The machinery in this chapter will be essential for interpreting a high-SNR cross correlation, and separating the clustering signal from propagation effects.

This chapter is organized as follows. In §2.2, we define notation and our modeling assumptions. In §2.3, we define our primary observable, the FRB-galaxy cross power spectrum C_ℓ^{fg}. We explore and interpret clustering contributions to C_ℓ^{fg} in §2.4, and propagation effects in §2.5. We present signal-to-noise forecasts in §2.6, and in §2.7 we describe a Monte Carlo simulation pipeline which we use to validate our forecasts. We conclude in §2.8.

2.2 Preliminaries

2.2.1 Definitions and notation

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

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

distance to redshift z:

$$\chi(z) = \int_0^z \frac{dz'}{H(z')} \tag{2.6}$$

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

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

$$n_X^{3d}(z) = \frac{H(z)}{\chi(z)^2} \frac{dn_X^{2d}}{dz} \qquad n_X^{2d} = \int dz \, \frac{dn_X^{2d}}{dz} \tag{2.7}$$

2.2.2 Halo model

We model FRB and galaxy clustering using the halo model. For a review of the halo model, see [33]. In this section, we give a high-level summary of our halo modeling formalism. For details, see Appendix A.

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

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

$$n_h^{3d}(M, z)\Big(1 + b_h(M, z)\delta_{\rm lin}(x)\Big) d^3\mathbf{x}\, dM \tag{2.8}$$

where $n_h^{3d}(M, z)$ is the *halo mass function*, or number density of halos per unit comoving volume per unit halo mass, and $b_h(M, z)$ is the *halo bias*. We use the Sheth-Tormen mass function and bias (Eqs. A.4, A.6).

Third, we randomly assign FRBs and galaxies to halos. We always assume that the number counts (N_f, N_g) of FRBs and galaxies are independent from one halo to the next. That is, (N_f, N_g) is a bivariate random variable whose probability distribution (the *halo occupation distribution* or HOD) depends only on halo mass M and redshift z. Once the counts (N_f, N_g) have been simulated, we assign spatial locations to each FRB and galaxy

independently, by sampling from the NFW spatial profile (Eq. A.7). We assume that galaxy positions are measured with negligible uncertainty, but FRB positions have statistical errors (θ_x, θ_y) which are Gaussian with FWHM denoted θ_f. Unless stated otherwise, we take the FRB angular resolution to be $\theta_f = 1$ arcminute.

2.2.3 Fiducial FRB models

Throughout the chapter, we derive analytic results for an arbitrary HOD, but show numerical results for two specific FRB models: the fiducial "low-z" and "high-z" FRB models. Our two fiducial models are intended to bracket the range of possibilities for the FRB redshift distribution currently allowed by observations. The median FRB redshift in the low-z and high-z FRB models is $z \sim 0.022$ and $z \sim 0.76$ respectively. The *host* DM distributions in the two models have been chosen so that the distribution of *total* DMs is similar (Figure 2.1). In the high-z FRB model, observed DM is a fairly good indicator of the FRB redshift, whereas in the low-z FRB model, there is not much correlation between DM and redshift. The high-z FRB model was motivated by the FRB host galaxy associations at redshifts 0.19, 0.32, 0.66 reported in [21, 71, 122, 9, 98], and the low-z FRB model was motivated by the ASKAP-2MPZ cross-correlation at very low redshift reported in [63].

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

2.2.4 Galaxy surveys

We consider three galaxy surveys throughout the chapter. First, the SDSS-DR8 optical photometric survey over redshift range $0 \le z \le 1.1$, with redshift distribution taken from [111]. Second, the 2MPZ all-sky infrared photometric survey [17], Almost all ($\approx 98\%$) of the 2MPZ galaxies have photometric redshifts < 0.2. Finally, the upcoming DESI-ELG spectroscopic survey, whose redshift distribution is forecasted in [2] and covers the range $0.6 \le z \le 1.7$. For photometric surveys, we neglect photometric redshift uncertainties, since these will be small compared to the FRB redshift uncertainty arising from scatter in the FRB host DM.

The galaxy HOD is constructed so that halos above threshold mass $M_g(z)$ contain $(M/M_g(z))$ galaxies on average. The redshift-dependent threshold halo mass $M_g(z)$ is chosen to match the redshift distribution of the galaxy survey ("abundance matching"). Numerical values of $M_g(z)$ are shown in Figure 2.2.

For more details of the FRB and galaxy models, including precise specification of the FRB redshift and host DM distributions in the two fiducial models, see Appendices A.2, A.3.

2.3 The power spectrum C_ℓ^{fg}

2.3.1 Definition

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

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

$$\delta_f(\boldsymbol{\theta}) = \frac{1}{n_f^{2d}} \sum_{i=1}^{N_f} \delta^2(\boldsymbol{\theta} - \boldsymbol{\theta}_i^{(f)}) \tag{2.9}$$

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

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

$$\delta_f(\boldsymbol{\ell}) = \frac{1}{n_f^{2d}} \sum_{i=1}^{N_f} \exp\Big(- i\boldsymbol{\ell} \cdot \boldsymbol{\theta}_i^{(f)} \Big) \tag{2.10}$$

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

$$\langle \delta_f(\boldsymbol{\ell})^* \, \delta_g(\boldsymbol{\ell}') \rangle = C_\ell^{fg} (2\pi)^2 \delta^2(\boldsymbol{\ell} - \boldsymbol{\ell}') \tag{2.11}$$

where the delta function on the RHS is a consequence of translation invariance. This equation defines the power spectrum C_ℓ^{fg}.

The power spectrum C_ℓ^{fg} is one representation for the two-point correlation function between δ_f, δ_g. Other representations, such as the two-point correlation function as a

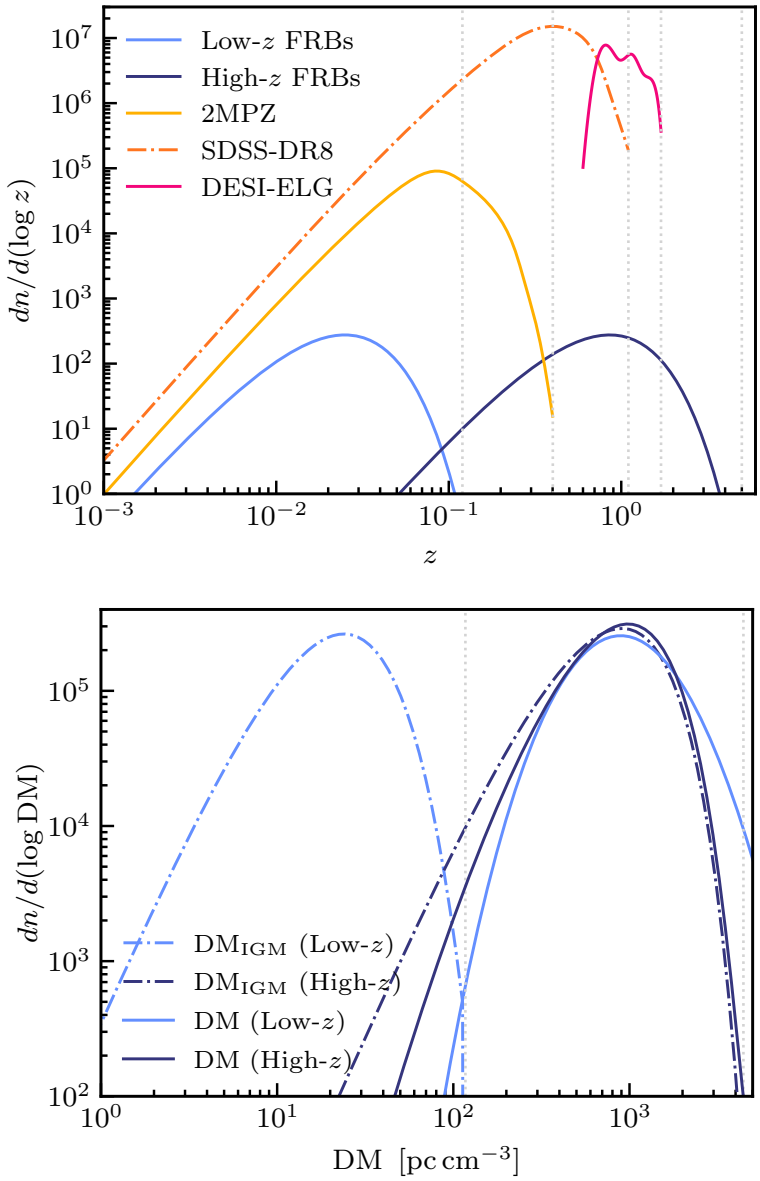

Figure 2.1: *Top panel:* FRB redshift distributions in our high-z and low-z fiducial FRB models (see §2.2), with galaxy redshift distributions shown for comparison. *Bottom panel:* FRB DM distributions in both fiducial models. We show total extragalactic DM (IGM+host, denoted "DM"), and the IGM contribution $\mathrm{DM}_{\mathrm{IGM}}(z)$. The total DM distribution is similar in the two fiducial models, but DMs are usually host-dominated in the low-z model, and IGM-dominated in the high-z model. Vertical dotted lines mark maximum redshift cutoffs.

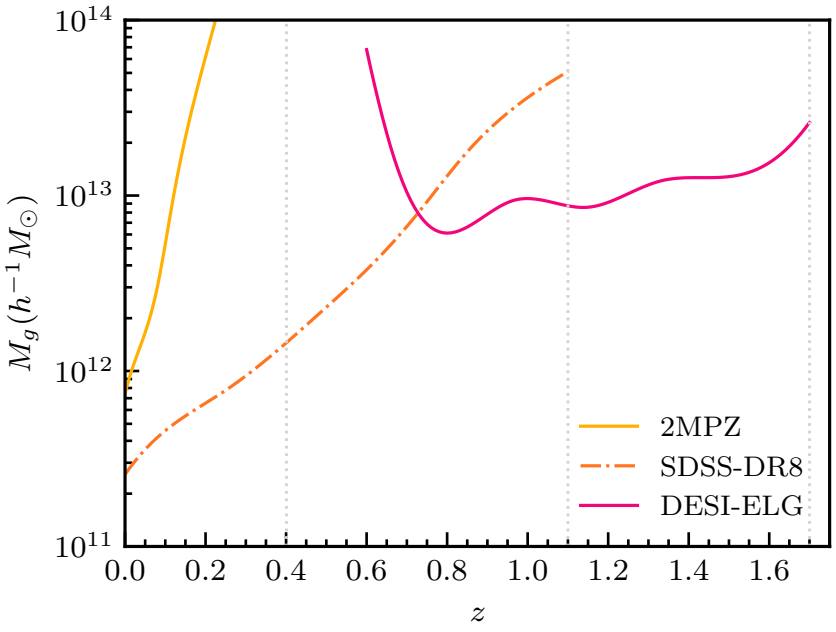

Figure 2.2: Threshold halo mass $M_g(z)$ for hosting a galaxy in the 2MPZ, SDSS-DR8 and DESI-ELG galaxy surveys, determined by abundance matching to the redshift distribution dn_g/dz as described in §2.2 and Appendix A.2. Vertical dotted lines mark maximum redshift cutoffs.

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

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

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

2.3.2 Two-halo and one-halo power spectra

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

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

$$C_\ell^{fg} = C_\ell^{fg(2h)} + C_\ell^{fg(1h)} \tag{2.12}$$

which correspond to correlations between FRBs and galaxies in different halos, or in the same halo. Some example 2-halo and 1-halo power spectra are shown in Figure 2.3.

The 2-halo term $C_\ell^{fg(2h)}$ is sourced by large-scale cosmological correlations, and is responsible for the large bump at low ℓ. For an arbitrary redshift z, the bump is at $\ell \sim k_{\mathrm{eq}}\chi(z)$, where $k_{\mathrm{eq}} \sim 0.02\ h\ \mathrm{Mpc}^{-1}$ is the scale of matter-radiation equality. The 2-halo term arises because FRBs and galaxies trace the same underlying large-scale cosmological density fluctuations. A complete expression for $C_\ell^{fg(2h)}$ is given by Eq. (A.39) in Appendix A. Here we give a simplified expression which applies on angular scales larger than the angular size of a halo ($\ell \lesssim \chi/R_{\mathrm{halo}}$, where R_{halo} is the comoving radius of a typical halo):

$$C_\ell^{fg(2h)} \rightarrow \frac{1}{n_f^{2d} n_g^{2d}} \int dz\, \frac{H(z)}{\chi(z)^2} \left(b_f(z) \frac{dn_f^{2d}}{dz} \right) \left(b_g(z) \frac{dn_g^{2d}}{dz} \right) P_{\mathrm{lin}}\left(\frac{\ell}{\chi(z)}, z \right) \tag{2.13}$$

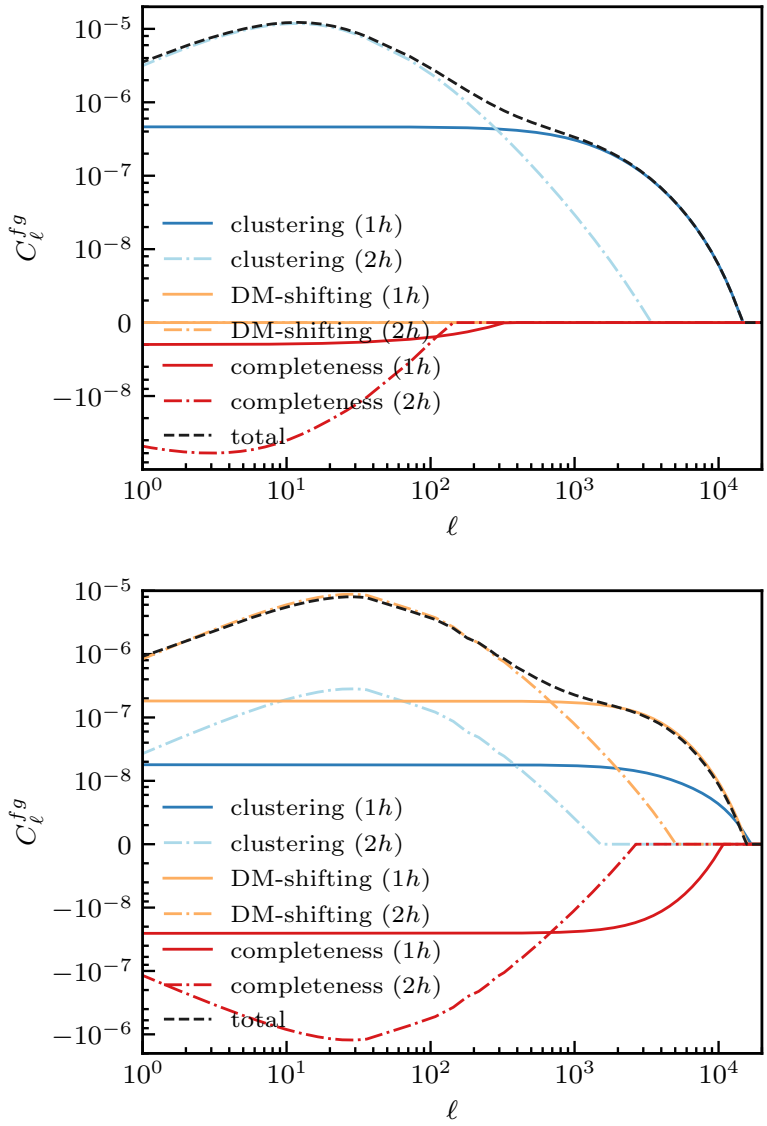

Figure 2.3: Angular cross power spectrum C_ℓ^{fg} as a function of ℓ for the fiducial high-z FRB model (see §2.2) and SDSS-DR8 galaxies. The total observed power spectrum is the sum of clustering and propagation contributions, and each contribution may be split into 1-halo and 2-halo terms, which we show separately here. Disentangling these terms is a challenge, and one of the main themes of this chapter. The clustering terms are described in §2.3.2, and the "DM-shifting" and "completeness" terms are propagation effects which will be described in §2.5. *Top panel:* unbinned FRB and galaxy fields. *Bottom panel:* FRB dispersion measure bin $1400 < \mathrm{DM} < 1500$ pc cm^{-3} and galaxy redshift bin $0.63 < z < 0.74$.

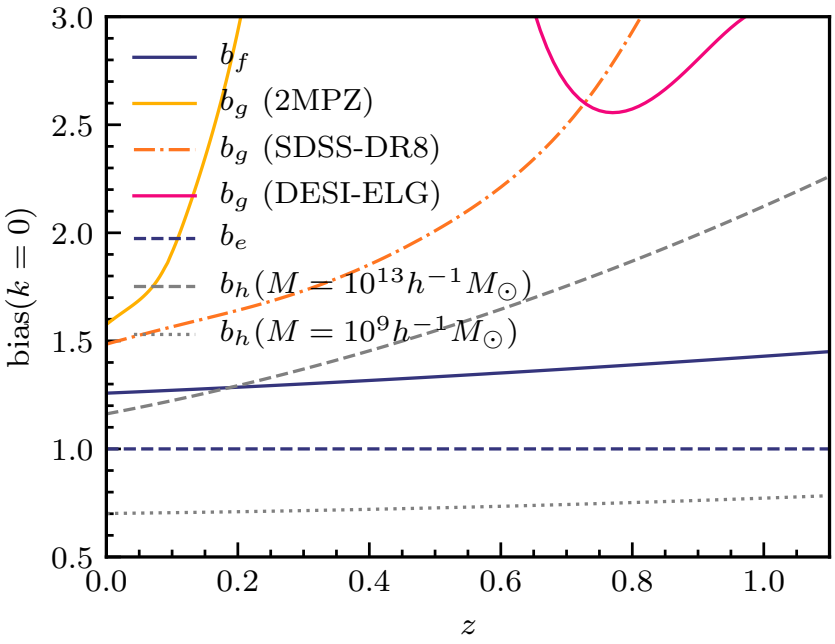

Figure 2.4: Large-scale bias parameters. The FRB bias $b_f(z)$ assumes minimum halo mass $M_f = 10^9\ h^{-1}\ M_\odot$. The galaxy bias $b_g(z)$ for the 2MPZ, SDSS-DR8 and DESI-ELG surveys assumes the minimum halo mass $M_g(z)$ from Figure 2.2. We take electron bias $b_e = 1$ throughout. We also show the halo bias $b_h(z)$ for two choices of halo mass.

Here, $b_f(z), b_g(z)$ are bias parameters which measure the coupling of FRBs and galaxies to the cosmological density field on large scales. The FRB bias b_f is defined by the statement that the FRB and matter overdensities are related by $\delta_f \approx b_f \delta_m$ on large scales, and likewise for b_g. An explicit formula for b_f, b_g is given in Eq. (A.40). and numerical values are shown in Figure 2.4. The 2-halo term mainly depends on the redshift overlap between the FRB and galaxy catalogs, via the factors $(b_f dn_f^{2d}/dz)(b_g dn_g^{2d}/dz)$ in Eq. (2.13).

As a technical comment, power spectra have been computed using the Limber approximation [64, 52, 53] throughout the chapter. We comment on the accuracy of the Limber approximation in Appendix B.

The 1-halo term $C_\ell^{fg(1h)}$ arises because FRBs and galaxies occupy the same dark matter halos. A complete expression for $C_\ell^{fg(1h)}$ is given by Eq. (A.39) in Appendix A. Here we give a simplified expression which applies on angular scales $\ell \lesssim \chi/R_{\rm halo}$ larger than the

angular size of a halo:

$$C_\ell^{fg(1h)} \to \frac{1}{n_f^{2d} n_g^{2d}} \int dz\, dM\, \frac{\chi(z)^2}{H(z)} n_h^{3d}(M,z) \left\langle N_f N_g \right\rangle_{M,z} \tag{2.14}$$

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

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

The 1-halo term $C_\ell^{fg(1h)}$ arises whenever FRBs and survey galaxies occupy the same halos. If FRBs actually inhabit the survey galaxies themselves, there will be an additional "Poisson" term $C_\ell^{fg(p)}$ which dominates on the smallest scales (high ℓ). We have neglected the Poisson term in our forecasts, since we are assuming that the FRB survey has insufficient resolution to associate FRBs and galaxies on a per-object basis, but this does make our forecasts slightly conservative. For more discussion of the Poisson term, see Eq. (A.41) in Appendix A.

2.4 The observables $b(dn/dz)$ and $\gamma(dn/dz)$

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

In this section, we will take a step in this direction, by showing how the ℓ-dependence can be absorbed into two observables, corresponding to the power spectrum amplitude in the 2-halo and 1-halo regimes. These observables, denoted $b(dn/dz)$ and $\gamma(dn/dz)$ for reasons to be explained shortly, will be functions of z and DM.

The basic idea is simple. For a narrow galaxy redshift bin $(z, z+\Delta z)$, the 2-halo and 1-halo power spectra in Eqs. (2.13), (2.14) have the following forms for $\ell \lesssim \chi/R_{\rm halo}$:

$$\begin{aligned} C_\ell^{fg(2h)} &\to \big(\text{Constant}\big)\, P_{\rm lin}\left(\frac{\ell}{\chi(z)}, z\right) \\ C_\ell^{fg(1h)} &\to \big(\text{Constant}\big) \end{aligned} \tag{2.15}$$

For $\ell \gtrsim \chi/R_{\rm halo}$, the power spectra acquire additional ℓ-dependence which gives information about halo profiles, but we will assume that this profile information is of secondary interest. Thus, the information in the ℓ-dependence of the power spectrum can be compressed into two numbers: the coefficients in Eq. (2.15). Given a measurement of the total power spectrum C_ℓ^{fg}, we can fit for both coefficients jointly, without much covariance between them.

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

Starting with the 2-halo power spectrum, we take Eq. (2.13) in the limit of a narrow redshift bin $(z, z+\Delta z)$, obtaining:

$$C_\ell^{fg(2h)} \to \frac{1}{n_f^{2d}} \frac{H(z)}{\chi(z)^2} \left(b_f(z) \frac{dn_f^{2d}}{dz} \right) b_g(z)\, P_{\rm lin}\left(\frac{\ell}{\chi(z)}, z \right) \tag{2.16}$$

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

The observable quantity $b_f(dn_f^{2d}/dz)$ is not as intuitive as the FRB redshift distribution (dn_f^{2d}/dz), but in practice the two are not very different. For example, in our fiducial model with threshold halo mass $M_f = 10^9\ h^{-1}\ M_\odot$, the FRB bias satisfies $1.2 \le b_f \le 1.5$ for $z \le 1$ (see Figure 2.4).

This interpretation of the 2-halo amplitude as a measurement of $b(dn/dz)$ is fairly standard and has been explored elsewhere [76, 77, 96, 58, 85]. The 1-halo amplitude is less straightforward to interpret, and does not seem to have a standard interpretation in the literature. In the rest of this section, we will define an analogous observable $\gamma(dn/dz)$ for the 1-halo amplitude. The definition is not specific to FRBs, and may be interesting in the context of other tracer populations.

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

We define the following 3D densities:

$$n^{3d}_{gg}(z) = \int dM\, n^{3d}_h(M,z)\, \langle N^2_g \rangle_{M,z} \tag{2.17}$$

$$n^{3d}_{fg}(z) = \int dM\, n^{3d}_h(M,z)\, \langle N_f N_g \rangle_{M,z} \tag{2.18}$$

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

$$\gamma_f(z) = \frac{n^{3d}_g(z)}{n^{3d}_f(z)} \frac{n^{3d}_{fg}(z)}{n^{3d}_{gg}(z)} \tag{2.19}$$

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

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

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

In Figure 2.5, we show $\gamma_f(z)$ for our fiducial HOD (Eqs. A.15, A.20) as a function of (M_f, M_g), the threshold halo masses for FRBs and galaxies. Consistent with the previous paragraph, if M_f and M_g are of the same order of magnitude, then γ_f is of order unity. In the regimes $M_f \ll M_g$ and $M_f \gg M_g$, the quantity γ_f will be $\lesssim 1$ and $\gtrsim 1$ respectively.

Now we show how the 1-halo amplitude can be interpreted as a measurement of $\gamma_f(z)(dn^{2d}_f/dz)$. We take Eq. (2.14) and specialize to a narrow redshift bin $(z, z + \Delta z)$, obtaining:

$$C_\ell^{fg(1h)} \to \frac{1}{n^{2d}_f} \frac{n^{3d}_{fg}(z)}{n^{3d}_g(z)} \tag{2.20}$$

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

$$C_\ell^{gg(1h)} \to \frac{1}{n^{2d}_g} \frac{n^{3d}_{gg}(z)}{n^{3d}_g(z)} \tag{2.21}$$

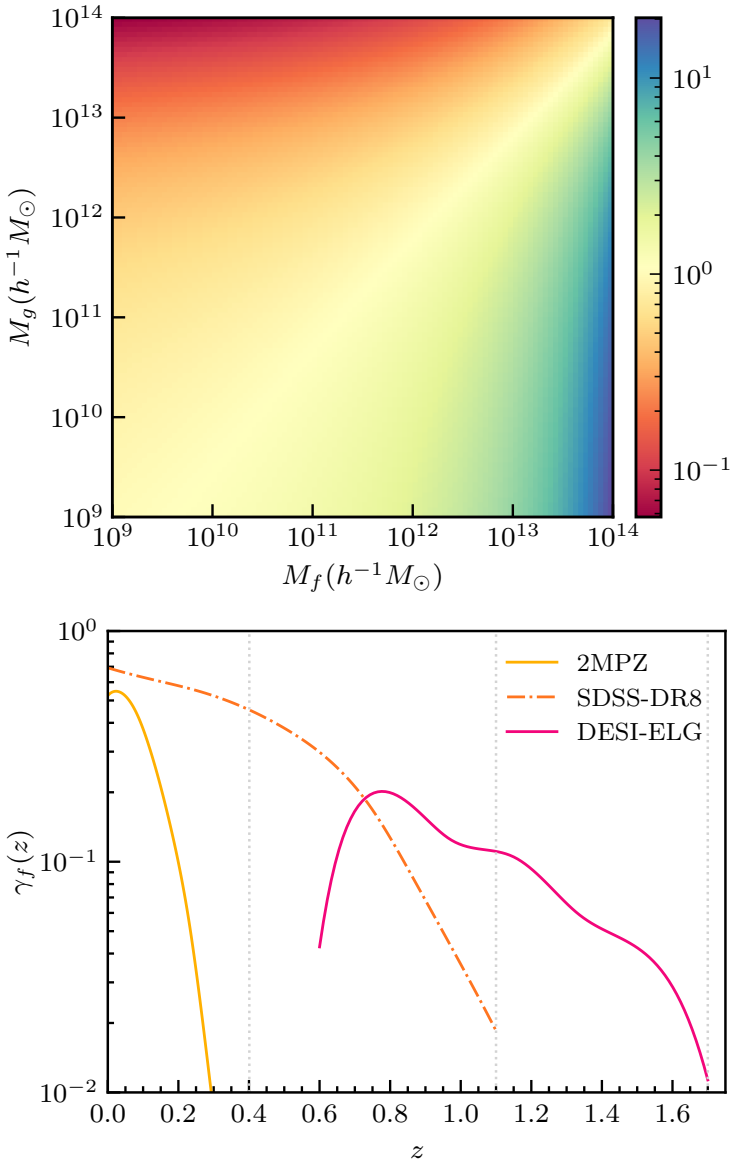

Figure 2.5: *Top panel:* Quantity $\gamma_f(z)$ defined in Eq. (2.19), as a function of threshold FRB halo mass M_f and threshold galaxy mass M_g, for Poisson HODs at redshift $z = 0.5$. If M_f and M_g are comparable, then γ_f is of order 1. *Bottom panel:* Quantity $\gamma_f(z)$ as a function of redshift, assuming FRB threshold halo mass $M_f = 10^9\ h^{-1}\ M_\odot$, and galaxy threshold halo mass $M_g(z)$ from Figure 2.2. At high redshifts, γ_f can be $\ll 1$ in our models, since galaxies are rare and our abundance-matching prescription gives a large value of M_g. Vertical dotted lines mark maximum redshift cutoffs.

by specializing Eq. (A.39) for $C_\ell^{gg(1h)}$ in Appendix A to low ℓ and a narrow redshift bin. Now we write $C_\ell^{fg(1h)}$ in the following form:

$$\begin{aligned} C_\ell^{fg(1h)} &\rightarrow \frac{n_g^{2d}}{n_f^{2d}} \gamma_f(z) \frac{n_f^{3d}(z)}{n_g^{3d}(z)} C_\ell^{gg(1h)} \\ &= \frac{\Delta z}{n_f^{2d}} \left(\gamma_f(z) \frac{dn_f^{2d}}{dz} \right) C_\ell^{gg(1h)} \end{aligned} \tag{2.22}$$

where the second line follows from the first by using Eq. (2.7). All factors on the RHS are known in advance except $\gamma_f(z) dn_f^{2d}/dz$, including the factor $C_\ell^{gg(1h)}$ which can be measured from the galaxy auto power spectrum. Therefore, the 1-halo amplitude can be interpreted as a measurement of the quantity $\gamma_f(z) dn_f^{2d}/dz$.

Summarizing, we have defined power spectrum observables $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$. By measuring the power spectrum C_ℓ^{fg} as a function of (ℓ, z), both observables may be constrained as functions of z. This extracts all information in C_ℓ^{fg}, except for suppression at high ℓ which contains information about halo profiles. The FRB catalog may be further binned in DM, to measure the observables $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$ as functions of (DM, z). In the top rows of Figures 2.6, 2.7, we show the observables as functions of (DM, z) in our fiducial model.

2.5 Propagation Effects

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

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

2.5.1 Generalities

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

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

$$\delta_f(\boldsymbol{\theta}) = \int dz\, W_f(z)\delta_e(\boldsymbol{\theta}, z) \tag{2.23}$$

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

Given the window function $W_f(z)$, the contribution to C_ℓ^{fg} due to propagation effects may be calculated from Eq. (2.23). In the Limber approximation, the result is:

$$C_\ell^{fg} = \frac{1}{n_g^{2d}} \int dz\, W_f(z) n_g^{3d}(z)\, P_{ge}\left(\frac{\ell}{\chi(z)}, z\right) \tag{2.24}$$

where $P_{ge}(k, z)$ is the 3D galaxy-electron power spectrum at comoving wavenumber k. We model P_{ge} using the halo model (Eq. A.42) in Appendix A). For a narrow galaxy redshift bin $(z, z + \Delta z)$, Eq. (2.24) becomes:

$$C_\ell^{fg} \to \frac{H(z)}{\chi(z)^2} W_f(z)\, P_{ge}\left(\frac{\ell}{\chi(z)}, z\right) \tag{2.25}$$

2.5.2 Dispersion-induced clustering

In this section we will compute the window function $W_f(z)$ defined by Eq. (2.23). There will be contributions to $W_f(z)$ from several propagation effects: dispersion, scattering, and lensing. In this chapter, we will describe the dispersion case in detail, deferring the other cases to future work.

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

$$\Delta(\boldsymbol{\theta}, z_f) = n_{e,0} \int_0^{z_f} dz\, \frac{1+z}{H(z)} \delta_e(\boldsymbol{\theta}, z) \tag{2.26}$$

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

$$n_f^{2d} = \int dz\, \frac{dn_f^{2d}}{dz} \tag{2.27}$$

We introduce the notation $(\partial/\partial\Delta)(dn_f^{2d}/dz)$ to denote the derivative of dn_f^{2d}/dz with respect to a foreground DM perturbation $\Delta(z)$ along the line of sight. Then, by differentiating Eq. (2.27), we can formally write the propagation-induced FRB anisotropy as:

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

Plugging in Eq. (2.26) for $\Delta(\boldsymbol{\theta}, z_f)$ and reversing the order of integration, we get:

$$\delta_f(\boldsymbol{\theta}) = \frac{n_{e,0}}{n_f^{2d}} \int dz \frac{1+z}{H(z)} \delta_e(\boldsymbol{\theta}, z) \int_z^\infty dz_f \left(\frac{\partial}{\partial \Delta} \frac{dn_f^{2d}}{dz_f} \right) \tag{2.28}$$

Comparing with the definition of W_f in Eq. (2.23) we read off the window function:

$$W_f(z) = \frac{n_{e,0}}{n_f^{2d}} \frac{1+z}{H(z)} \int_z^\infty dz' \left(\frac{\partial}{\partial \Delta} \frac{dn_f^{2d}}{dz'} \right) \tag{2.29}$$

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

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

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

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

$$\frac{dn_f^{2d}}{dz} = \int_{\mathrm{DM}_{\mathrm{min}}}^{\mathrm{DM}_{\mathrm{max}}} d(\mathrm{DM}) \, \frac{d^2 n_f^{2d}}{dz \, d(\mathrm{DM})} \tag{2.30}$$

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

Then the DM-shifting term is:

$$\left(\frac{\partial}{\partial\Delta}\frac{dn_f^{2d}}{dz}\right)_{\text{DM-shifting}} = \left(\frac{d^2n_f^{2d}}{dz\,d(\text{DM})}\right)_{\text{DM}_{\min}} - \left(\frac{d^2n_f^{2d}}{dz\,d(\text{DM})}\right)_{\text{DM}_{\max}} \tag{2.31}$$

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

1. The FRB catalog is constructed by selecting all objects above threshold signal-to-noise SNR_*.

2. All FRBs have the same intrinsic pulse width t_i.

3. In each redshift and DM bin, the FRB luminosity function is Euclidean: the number of FRBs above fluence F_* is proportional to $(F_*^{-3/2})$.[3]

4. SNR is related to fluence F by

$$\text{SNR} \propto \frac{F}{(t_i^2 + t_s^2 + t_d^2)^{1/4}} \tag{2.32}$$

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

$$t_d = 2\mu(\text{DM})\nu^{-3}(\Delta\nu) \tag{2.33}$$

   where ν is the observing frequency, $(\Delta\nu)$ is the channel bandwidth, and $\mu = 4.15$ ms GHz^2 is the coefficient in the FRB dispersion relation (delay) $= \mu(\text{DM})/\nu^2$ in Eq. (2.2).

Under these assumptions, we can calculate the derivative of $\log d^2n_f/(dz\,d(\text{DM}))$ with

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

respect to a foreground DM perturbation Δ, as follows:

$$\begin{aligned}
\frac{\partial}{\partial\Delta}\left(\log\frac{d^2n_f}{dz\,d(\mathrm{DM})}\right) &= -\frac{3}{2}\frac{\partial\log F_*}{\partial\Delta} \\
&= -\frac{3}{2}\frac{\partial\log(t_i^2+t_s^2+t_d^2)^{1/4}}{\partial\Delta} \\
&= -\frac{3t_d}{4(t_i^2+t_s^2+t_d^2)}\frac{\partial t_d}{\partial\Delta} \\
&= -\frac{3\mu(\Delta\nu)t_d}{2\nu^3(t_i^2+t_s^2+t_d^2)}
\end{aligned} \tag{2.34}$$

Here, the first line follows from toy model assumption 3, the second line follows from Eq. (2.32), the and the last line follows from differentiating Eq. (2.33) with respect to DM.

To get the completeness term in the derivative $(\partial/\partial\Delta)(dn_f^{2d}/dz)$, we integrate Eq. (2.34) over DM:

$$\begin{aligned}
&\left(\frac{\partial}{\partial\Delta}\frac{dn_f^{2d}}{dz}\right)_{\text{completeness}} \\
&\quad= \int d(\mathrm{DM})\left(\frac{\partial}{\partial\Delta}\frac{d^2n_f^{2d}}{dz\,d(\mathrm{DM})}\right) \\
&\quad= \int d(\mathrm{DM})\left(\frac{d^2n_f^{2d}}{dz\,d(\mathrm{DM})}\right)\left(-\frac{3\mu(\Delta\nu)t_d}{2\nu^3(t_i^2+t_s^2+t_d^2)}\right)
\end{aligned} \tag{2.35}$$

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

Summarizing, in this section we have calculated two contributions to C_ℓ^{fg} from propagation effects: a "DM-shifting" term and a "completeness" term. In both cases, the contribution to C_ℓ^{fg} is calculated as follows. We compute the intermediate quantity $(\partial/\partial\Delta)(dn_f^{2d}/dz)$ using Eq. (2.31) or Eq. (2.35), then the window function $W_f(z)$ using Eq. (2.29), and finally C_ℓ^{fg} using Eq. (2.24).

Finally, other studies have proposed to isolate these propagation effects to measure P_{ge} by cross-correlating galaxies with the 3D field $\bar{\Delta}(\boldsymbol{\theta})$ of DM averaged over all FRBs detected in a particular direction $\boldsymbol{\theta}$. Such statistics are related to the DM moment of C_ℓ^{fg}:

$$C_\ell^{\bar{\Delta}g} \propto \sum_i \mathrm{DM}_i \, n_{f_i}^{2d} \, C_\ell^{f_i g}, \tag{2.36}$$

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

2.5.3 Numerical results

In this section, we numerically compare contributions to C_ℓ^{fg} from spatial clustering, and two propagation effects: DM-shifting (Eq. 2.31) and completeness (Eq. 2.35). For the completeness effect, we have used FRB intrinsic width $t_i = 10^{-3}$ sec, and instrumental parameters matching CHIME/FRB: time sampling $t_s = 10^{-3}$ sec, channel bandwidth $\Delta\nu =$ 400 kHz, and central frequency $\nu = 600$ MHz.

To visualize contributions to C_ℓ^{fg}, we compress the power spectrum into two observables $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$, as described in §2.4. To compute these observables for propagation effects, we split the galaxy-electron power spectrum P_{ge} into 2-halo and 1-halo terms (see Eq. (A.42) in Appendix A). For $\ell \lesssim \chi/R_{\rm halo}$, these take the forms

$$P_{ge}^{2h}(k,z) \;\rightarrow\; b_g(z) b_e(z) P_{\rm lin}(k,z) \tag{2.37}$$

$$P_{ge}^{1h}(k,z) \;\rightarrow\; \frac{n_{ge}^{3d}(z)}{n_g^{3d}(z) n_e^{3d}(z)} \tag{2.38}$$

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

$$n_{ge}^{3d}(z) = \int dM \, n_h^{3d}(M,z) \left\langle N_g N_e \right\rangle_{M,z} \tag{2.39}$$

similar to the definition of $n_{fg}^{3d}(z)$ in Eq. (2.18). Now a calculation combining Eqs. (2.16), (2.22), (2.25), (2.37), (2.38) shows that the contribution to the power spectrum observables $(b_f dn_f^{2d}/dz)$ and $(\gamma_f dn_f^{2d}/dz)$ from propagation effects is:

$$\left(b_f \frac{dn_f^{2d}}{dz} \right)_{\rm prop} = W_f(z) \big(b_e(z) n_f^{2d} \big) \tag{2.40}$$

$$\left(\gamma_f \frac{dn_f^{2d}}{dz} \right)_{\rm prop} = W_f(z) \big(\gamma_e(z) n_f^{2d} \big) \tag{2.41}$$

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

$$\gamma_e(z) = \frac{n_g^{3d}(z)}{n_e^{3d}(z)} \frac{n_{ge}^{3d}(z)}{n_{gg}^{3d}(z)} \tag{2.42}$$

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

In Figures 2.6, 2.7, we show power spectrum observables $b_f(dn_f^{2d}/dz)$ and $\gamma_f(dn_f^{2d}/dz)$ from clustering and both propagation effects, in the (DM, z) plane. It is seen that propagation effects are comparable in size to the clustering signal! However, it is qualitatively clear from Figures 2.6, 2.7 that there is some scope for separating the two based on their dependence on redshift and DM.

2.5.4 Ideas for separating spatial clustering from propagation effects

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

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

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

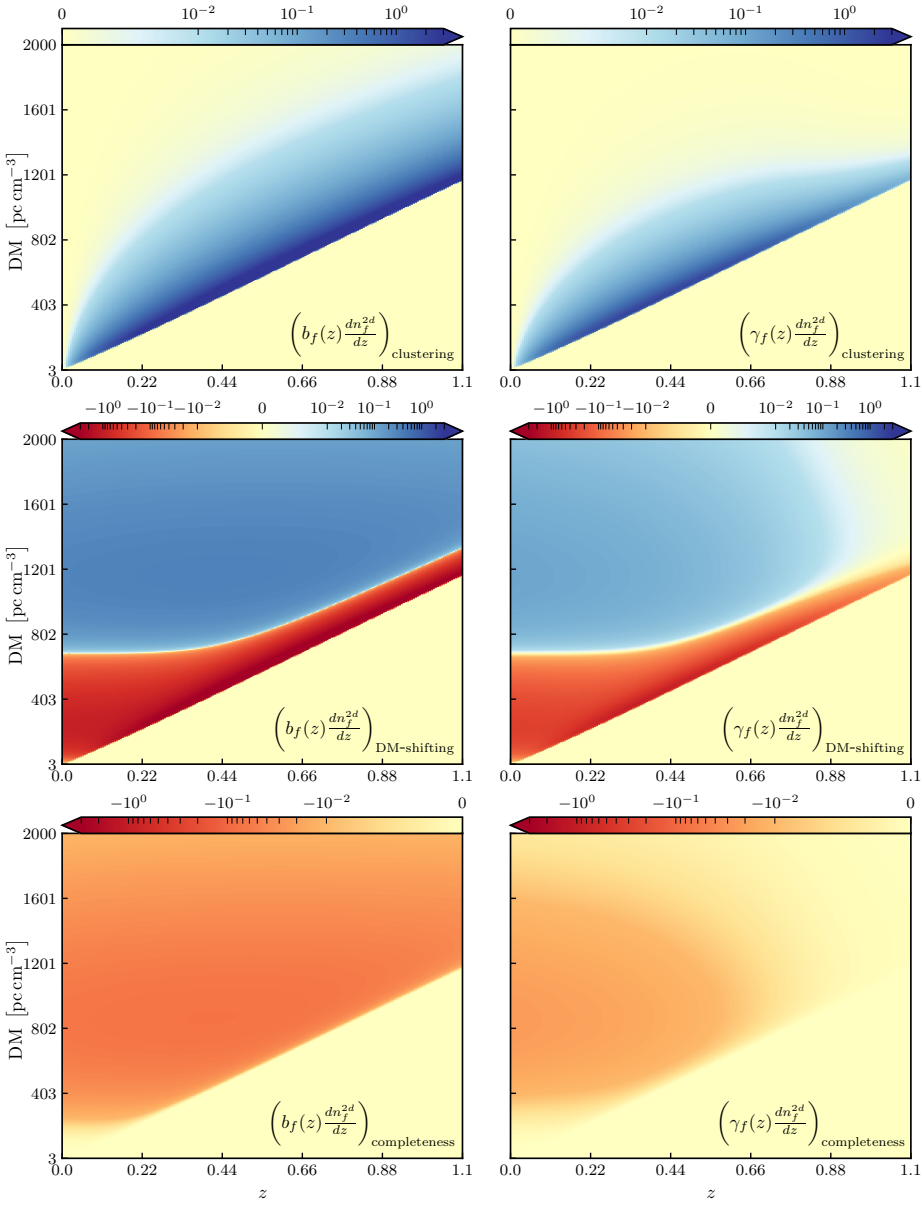

Figure 2.6: Visual comparison between clustering and propagation contributions to the clustering power spectrum C_ℓ^{fg}, for our fiducial high-z FRB model and SDSS-DR8. Each row corresponds to one such contribution: clustering (top), DM-shifting propagation effect (middle), and completeness propagation effect (bottom). Since C_ℓ^{fg} is a function of three variables (z, DM, ℓ), we compress the ℓ-dependence into two clustering observables $b_f dn_f/dz$ (left column) and $\gamma_f dn_f/dz$ (right column), as described in §2.4. Qualitatively, it is clear that clustering and propagation effects may be distinguished based on their (z, DM) dependence.

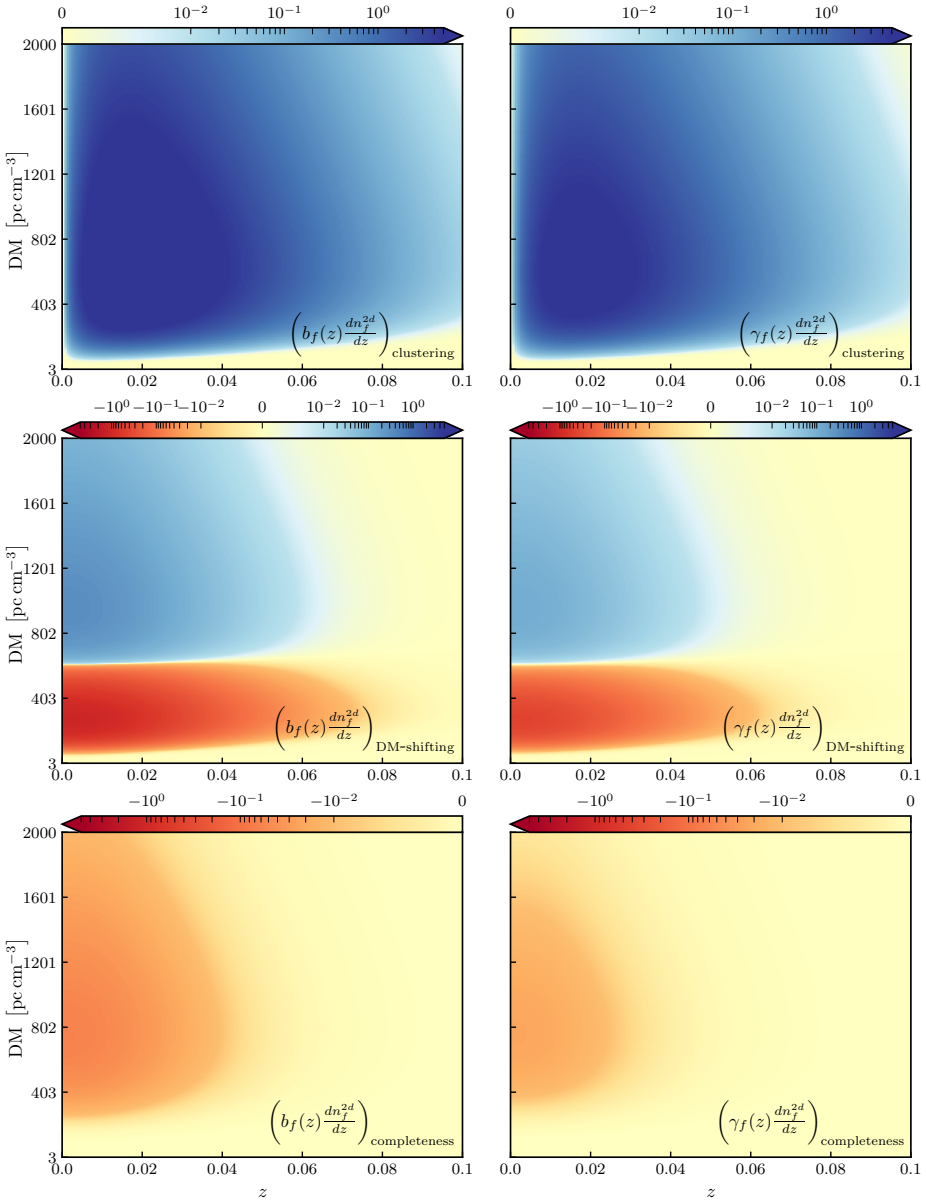

Figure 2.7: Same as Figure 2.6, but for the fiducial low-z FRB model.

Some propagation effects have a preferred sign, for example the completeness term in Eq. (2.35) is negative, since adding dispersion makes FRBs harder to detect.[4] Scattering is another example of a propagation effect with a negative sign, for the same reason.

Propagation effects appear in the C_ℓ^{fg} power spectrum via the product $W_f(z)P_{ge}(\ell/\chi, z)$ (Eq. 2.24). We will discuss separately how the window function $W_f(z)$ and galaxy-electron power spectrum $P_{ge}(k, z)$ might be modeled.

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

$$W_f(z) \;=\; n_{e,0}\frac{1+z}{H(z)}\int_z^\infty \frac{dz'}{n_f^{2d}}\left[\left(\frac{d^2n_f^{2d}}{dz'\,d(\mathrm{DM})}\right)_{\mathrm{DM}_{\mathrm{min}}} - \left(\frac{d^2n_f^{2d}}{dz'\,d(\mathrm{DM})}\right)_{\mathrm{DM}_{\mathrm{max}}}\right] \tag{2.43}$$

by combining Eqs. (2.29), (2.31). In the limit of low z this becomes:

$$\lim_{z\to 0} W_f(z) = \frac{n_{e,0}}{H_0}\frac{1}{n_f^{2d}}\left[\left(\frac{dn_f^{2d}}{d(\mathrm{DM})}\right)_{\mathrm{DM}_{\mathrm{min}}} - \left(\frac{dn_f^{2d}}{d(\mathrm{DM})}\right)_{\mathrm{DM}_{\mathrm{max}}}\right] \tag{2.44}$$

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

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

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

On 1-halo dominated scales, modeling $P_{ge}(k, z)$ is more difficult. One interesting near-future possibility is to measure $P_{ge}(k, z)$ through the kSZ (kinetic Sunyaev-Zeldovich) effect in the cosmic microwave background. Currently, the kSZ effect has been detected at a few

[4] As discussed near Eq. (2.35), this is true for our toy instrumental model, but not guaranteed to be true for a real pipeline.

sigma, but not constrained to high precision. However, measurements at the $\approx 10\sigma$ level are imminent, and future CMB experiments such as Simons Observatory and CMB-S4 will measure P_{ge} with percent-level accuracy [116, 1]. These measurements will be very informative for modeling FRB propagation effects.

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

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

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

2.6 Forecasts and signal-to-noise

2.6.1 Fisher matrix formalism

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

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

$$C_\ell^{f_i g_j} = \sum_\mu A_\mu C_\ell^{f_i g_j(\mu)} \tag{2.45}$$

where $\mu = 1, \cdots, P$. That is, the power spectrum is the sum of P terms whose ℓ, i, j dependence is fixed by a model, but whose coefficients A_μ are to be determined from data.

For example, we could take $\mu \in \{1h, 2h\}$ with $P = 2$, to forecast constraints on the overall amplitude of the 1-halo and 2-halo clustering terms. Propagation effects can similarly be included in the forecast.

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

$$F_{\mu\nu} = \Omega \sum_{ij} \int \frac{\ell \, d\ell}{2\pi} \frac{C_\ell^{f_i g_j(\mu)} C_\ell^{f_i g_j(\nu)}}{C_\ell^{f_i f_i} C_\ell^{g_j g_j}} \tag{2.46}$$

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

$$C_\ell^{f_i f_i} = \left(n_{f_i}^{2d}\right)^{-1} \tag{2.47}$$

but have written $C_\ell^{f_i f_i}$ in Eq. (2.46) for notational uniformity.

The Fisher matrix is the forecasted *inverse* covariance matrix of the amplitude parameters A_μ in Eq. (2.45). For example, if $P = 1$, then the 1-by-1 Fisher "matrix" F is the SNR^2, and the statistical error on the amplitude parameter is $\sigma(A) = 1/\sqrt{F}$.

A few technical comments. The form of the Fisher matrix in Eq. (2.46) assumes that FRB and galaxy fields are each uncorrelated, i.e.

$$C_\ell^{f_i f_j} = \delta_{ij} C_\ell^{f_i f_i} \qquad C_\ell^{g_i g_j} = \delta_{ij} C_\ell^{g_i g_i} \tag{2.48}$$

This assumption is satisfied for FRB fields, since we are assuming that auto spectra are Poisson noise dominated. The galaxy fields will also be uncorrelated if they are defined by a set of non-overlapping redshift bins. Eq. (2.46) also assumes that $C_\ell^{fg} \ll (C_\ell^{ff} C_\ell^{gg})^{1/2}$ in the fiducial model. This will be a good approximation if the FRB number density is not too large. Finally, in Eq. (2.46) we have written the Fisher matrix as a double sum over (redshift, DM) bins for maximum generality, but for numerical forecasts we take the limit of narrow bins, by replacing the sum by an appropriate double integral.

2.6.2 Numerical results

In Table 2.1, we show SNR forecasts for several FRB and galaxy surveys. We report SNR separately for six contributions to the power spectrum C_ℓ^{fg} as follows. First, we split the power spectrum into three terms from gravitational clustering, and the DM-shifting and completeness propagation effects described in §2.5. We then split each of these terms into 1-halo and 2-halo contributions, for a total of 6 terms. Each SNR entry in Table 2.1 is given by $\sqrt{F_{ii}}$, where F_{ii} is the appropriate diagonal element of the 6-by-6 Fisher matrix. This

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

Table 2.1: Forecasted SNR for FRB-galaxy cross-correlations. Each row corresponds to a choice of FRB model, galaxy survey, and FRB angular resolution θ_f. Each column corresponds to one contribution to the FRB-galaxy power spectrum. Each entry is total SNR after summing over angular wavenumber ℓ, and a narrow set of redshift and DM bins. We have assumed a catalog with $N_{\rm FRB} = 1000$ FRBs ($\mathrm{DM}_{\rm max} = 10^4$ pc cm^{-3}); in general each SNR value scales as $N_{\rm FRB}^{1/2}$.

corresponds to SNR of each contribution considered individually, without marginalizing the amplitude of the other terms in a joint fit.

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

One interesting feature of Table 2.1 is that if FRBs do extend to high redshift, the cross-correlation with a high-redshift galaxy sample is detectable (e.g. SNR=12 for the high-z FRB model, DESI-ELG, and $\theta_f = 1$ arcminute). Angular cross-correlations should be a powerful tool for probing the high-z end of the FRB redshift distribution, where galaxy surveys are far from complete, and FRB host galaxy associations are difficult.

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

$$\begin{pmatrix} 1.00 & 0.20 & -0.76 & -0.17 & -0.10 & -0.03 \\ 0.20 & 1.00 & -0.14 & -0.83 & -0.02 & -0.14 \\ -0.76 & -0.14 & 1.00 & 0.19 & -0.22 & -0.04 \\ -0.17 & -0.83 & 0.19 & 1.00 & -0.04 & -0.23 \\ -0.10 & -0.02 & -0.22 & -0.04 & 1.00 & 0.19 \\ -0.03 & -0.14 & -0.04 & -0.23 & 0.19 & 1.00 \end{pmatrix} \tag{2.49}$$

where the row ordering is the same as Table 2.1. We see that there is not much correlation between 1-halo and 2-halo signals, but the clustering signal is fairly anti-correlated to the DM-shifting signal. The correlation is not perfect, since there is some difference in the (redshift, DM) dependence, as can be seen directly by comparing the top and middle rows of Figure 2.6. The correlation matrix depends to some degree on model assumptions. For example, in the low-z FRB model, the correlation matrix is:

$$\begin{pmatrix} 1.00 & 0.17 & -0.02 & -0.00 & -0.78 & -0.16 \\ 0.17 & 1.00 & -0.00 & -0.02 & -0.13 & -0.86 \\ -0.02 & -0.00 & 1.00 & 0.19 & -0.19 & -0.04 \\ -0.00 & -0.02 & 0.19 & 1.00 & -0.04 & -0.20 \\ -0.78 & -0.13 & -0.19 & -0.04 & 1.00 & 0.19 \\ -0.16 & -0.86 & -0.04 & -0.20 & 0.19 & 1.00 \end{pmatrix} \tag{2.50}$$

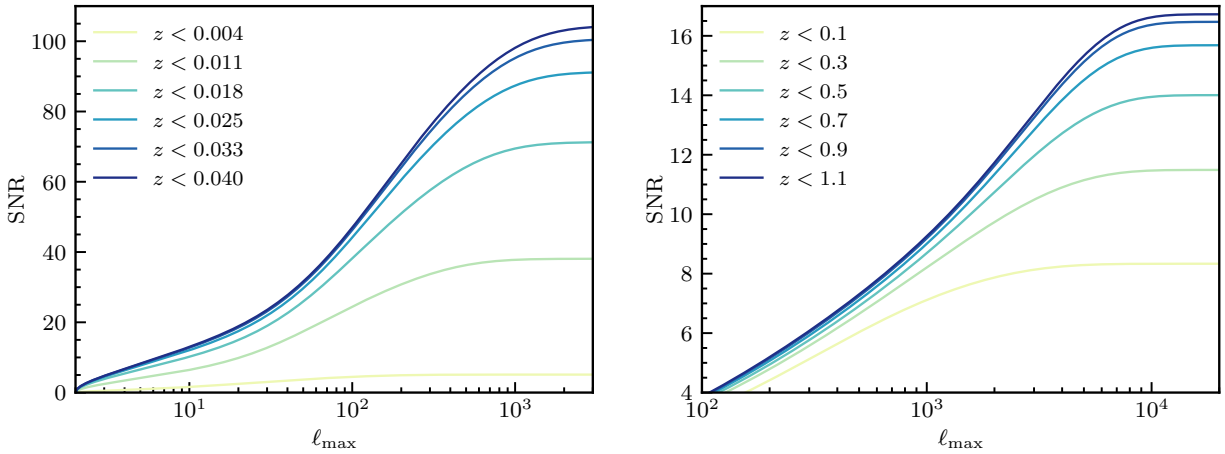

Figure 2.8: Forecasted SNR of FRB-galaxy cross power, for varying choices of maximum redshift $z_{\rm max}$ and maximum angular wavenumber $\ell_{\rm max}$, after summing over narrow (DM, z) bins. *Left panel:* Fiducial low-z FRB model and SDSS-DR8 galaxies. *Right panel:* Fiducial high-z FRB model and SDSS-DR8 galaxies.

where we have assumed the SDSS-DR8 galaxy survey. Here, there is a large correlation between clustering and completeness terms. (However, Table 2.1 shows that completeness terms are small in the low-z FRB model.)

Figure 2.8 shows the evolution of total SNR as a function of angular wavenumber and redshift. In the analysis of real data, large scales ($\ell \lesssim 20$) may be contaminated by Galactic systematic effects, such as dust extinction. Figure 2.8 shows that these scales make a small contribution to the total SNR, so our forecasts are robust against such systematics.

2.7 Simulations

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

$$\mathrm{Cov}(C_b^{fg}, C_{b'}^{fg}) \approx \frac{C_b^{ff} C_b^{gg}}{f_{\rm sky}(\ell_{\rm max}^{(b)2} - \ell_{\rm min}^{(b)2})} \delta_{bb'} \tag{2.51}$$

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

In reality, FRB and galaxy fields are non-Gaussian. The FRB catalog consists of a modest number of objects which obey Poisson (not Gaussian) statistics. Galaxy catalogs are larger, but Poisson statistics of the underlying halos may be important, since the number of halos is smaller than the number of galaxies. The purpose of this section is to determine whether the Gaussian covariance (2.51) is a good approximation, by carrying out Monte Carlo simulations of galaxies and FRBs.

2.7.1 Description of simulation pipeline

Our simulation pipeline is based on the halo model from §2.2 and Appendix A. We use the high-z FRB model and SDSS-DR8 galaxy survey. Because non-Gaussian effects are expected to be largest for the 1-halo term, our simulation pipeline only includes 1-halo clustering. In particular, we do not simulate the Gaussian linear density field $\delta_{\rm lin}$, because it is not needed to simulate 1-halo clustering.

We use a 10×10 deg^2 sky patch, in the flat-sky approximation with periodic boundary conditions. We sample Poisson random halos in 100 redshift bins, and 500 logarithmically-spaced mass bins between M_f and $M_{\rm max} = 10^{17} h^{-1} M_\odot$. For each halo, we assign an FRB and galaxy count by sampling a Poisson random variable whose expectation value is given by the HODs in Eqs. (A.15), (A.20). For each FRB and galaxy, we assign a 3D location within the halo using the NFW profile (Eq. A.7). Angular positions are computed by projecting 3D positions onto the sky patch. In the case of FRBs, we convolve sky locations by the beam (Eq. A.34). Finally, FRBs are assigned a random DM, which is the sum of the IGM contribution $\mathrm{DM}_{\rm IGM}(z)$ and a random host contribution $\mathrm{DM}_{\rm host}$ (see Eq. A.23).

Next, we grid the FRB and galaxy catalogs onto a real-space 2049×2049 pixelization with resolution ≈ 0.3 arcminutes, using the cloud-in-cell (CIC) weighting scheme. We take the Fourier transform to obtain Fourier-space fields $\delta_f(\boldsymbol{\ell})$, $\delta_g(\boldsymbol{\ell})$. Then, following Eq. (A.37), we estimate the angular cross power spectrum C_ℓ^{fg} by averaging the cross power $\langle \delta_f(\boldsymbol{\ell})^* \, \delta_g(\boldsymbol{\ell}) \rangle$ in a non-overlapping set of ℓ-bins.

2.7.2 Numerical results

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

in Eq. (2.51), we first estimate the covariance of the simulations as:

$$\mathrm{Cov}(C_b^{fg}, C_{b'}^{fg}) \;=\; \left(\frac{1}{n_{\mathrm{sim}}-1}\right)\sum_{i=1}^{n_{\mathrm{sim}}}(C_b^{fg,i}-\langle C_b^{fg}\rangle)(C_{b'}^{fg,i}-\langle C_{b'}^{fg}\rangle) \tag{2.52}$$

In Figure 2.9, we show the bandpower *correlation* matrix $r_{bb'}$, obtained from the Monte Carlo covariance matrix $C_{bb'}$ in Eq. (2.51) by

$$r_{bb'} = \frac{C_{bb'}}{(C_{bb}C_{b'b'})^{1/2}} \tag{2.53}$$

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

In Figure 2.10, we compare the total SNR of the FRB-galaxy cross-correlation obtained from simulations to the Gaussian approximation. The total SNR was computed as:

$$\mathrm{SNR}^2 = \sum_{b,b'}(C_b^{fg})\,\mathrm{Cov}(C_b^{fg}, C_{b'}^{fg})^{-1}(C_{b'}^{fg}) \tag{2.54}$$

where $\mathrm{Cov}(C_b^{fg}, C_{b'}^{fg})$ is either the Monte Carlo covariance matrix in Eq. (2.52) or the Gaussian approximation in Eq. (2.51). From Figure 2.10, the total SNR in the simulations agrees almost perfectly with the Gaussian forecast. This indicates that our forecasts in previous sections, which assume Gaussian statistics, are good approximations to the true non-Gaussian statistics of the FRB and galaxy fields.

2.8 Discussion

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

One complication is that the FRB redshift distribution (dn_f/dz) is not quite directly measurable. In §2.4 we studied this issue and showed that there are two clustering observables ($b_f dn_f/dz$) and ($\gamma_f dn_f/dz$) in the 2-halo and 1-halo regimes respectively. Here, b_f is

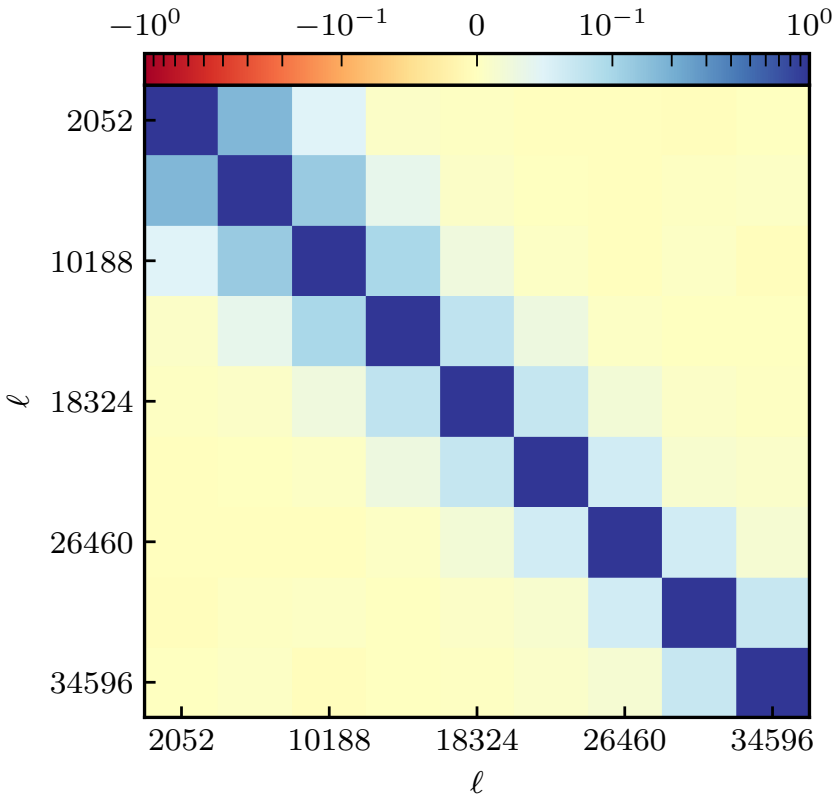

Figure 2.9: Bandpower correlation matrix $r_{bb'}$ of the FRB-galaxy cross power spectrum $C_\ell^{fg(1h)}$, estimated from simulations (see Eq. 2.53). We have used the fiducial high-z FRB model, SDSS-DR8 galaxies, FRB angular resolution $\theta_f = 1'$, and maximum dispersion measure $\mathrm{DM_{max}} = 10^4$ pc cm^{-3} . Correlations between bandpowers are $\approx 20\%$ for adjacent ℓ-bins, and decay rapidly after that. This is one way of quantifying the importance of non-Gaussian statistics, since off-diagonal correlations would be zero if the FRB and galaxy fields were Gaussian.

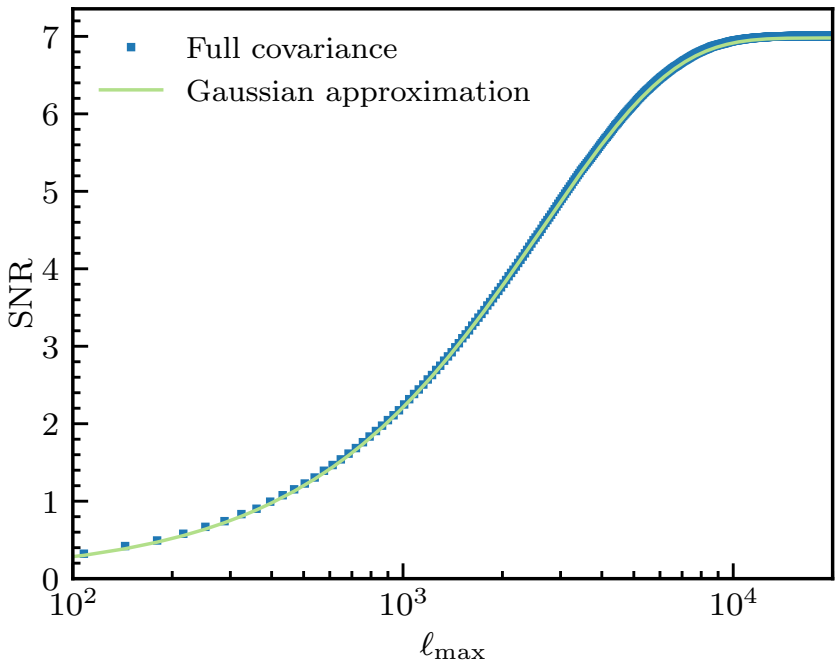

Figure 2.10: Cumulative SNR for the FRB-galaxy cross power spectrum $C_\ell^{fg(1h)}$, using the Monte Carlo bandpower covariance (Eq. 2.52), with the Gaussian approximation shown for comparison (Eq. 2.51). The two agree almost perfectly, justifying the Gaussian forecasts used throughout the chapter. We have used the fiducial high-z FRB model, SDSS-DR8 galaxies, FRB angular resolution $\theta_f = 1'$, and maximum dispersion measure $\mathrm{DM}_{\rm max} = 10^4$ pc cm^{-3} .

the usual large-scale bias parameter, and the quantity γ_f (defined in Eq. 2.19) depends on details of HODs.

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

Propagation effects can be separated from clustering based on their dependence as functions of (z, DM, ℓ). This is shown qualitatively in Figures 2.6 and 2.7, where clustering and propagation signals have very different (z, DM) dependence (after compressing the ℓ dependence into the two clustering observables $(b_f dn_f/dz)$ and $(\gamma_f dn_f/dz)$). For a longer, more systematic discussion, see §2.5.4.

Propagation effects are both a potential contaminant of the clustering signal, and a potential source of information about ionized electrons in the universe. Indeed, the "DM-shifting" propagation effect identified in §2.5 can be used to probe the distribution of electrons in the CGM, along the lines of [75, 74, 113, 99, 81, 69].

We now interpret our forecasts in relation to the 3σ correlation between ASKAP-discovered FRBs and 2MPZ galaxies measured in [63]. Scaling to a sample of 21 galaxies, and noting the weak dependence on angular resolution, our low-z FRB model predicts an intrinsic clustering correlation SNR of roughly 12, a factor of 4 higher than that observed. While it is not straightforward to interpret SNR units—the difference could be one of either signal amplitude, estimator optimality, or modeling—this would nonetheless seem to disfavor a completely nearby population. However, the measured SNR is far greater than what our high-z FRB model predicts and cannot be explained by DM-shifting (the measurement was unbinned in DM) or completeness as modeled (wrong sign and too small of an amplitude). As such, we suggest that the true FRB population may be somewhere between these two models, which could still be consistent with the 3 direct localizations (at high redshifts $z = 0.19$, 0.32, 0.66).

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

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

The answer depends on the angular scale ℓ where the power spectrum C_ℓ^{fg} is measured. On angular scales which are 2-halo dominated, FRBs and galaxies correlate via the observable $(b_f b_g dn_f/dz)$, so the observation just means that the galaxy bias b_g is larger for star-forming galaxies. On 1-halo dominated scales, the observation would imply that FRBs preferentially inhabit *halos* which contain star-forming galaxies, but this does not necessarily imply that FRBs inhabit the star-forming galaxies themselves. Finally, at very high ℓ where C_ℓ^{fg} is dominated by the Poisson term (a regime which we have mostly ignored in this chapter, but see discussion in §2.3), the observation would imply that FRBs do preferentially inhabit star-forming galaxies.

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

Chapter 3

Cross-correlating CHIME/FRB sources with galaxies

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×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 ($\sim 400\ \mathrm{pc\,cm^{-3}}$), 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.

3.1 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. [35, 87, 90]). The Canadian Hydrogen Intensity Mapping Experiment / Fast Radio Burst Project [23] has recently released its first catalog of fast radio bursts (FRBs) containing

492 unique sources [29], increasing the number of known FRBs by a factor $\sim$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 [16]. 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 §3.3. By measuring the correlation as a function of galaxy redshift and FRB dispersion measure (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 [75, 74, 113, 69, 95, 103, 102, 6], and applied to the ASKAP and 2MPZ/HIPASS catalogs by [63]. In this chapter, we will use machinery developed in §2 [95] 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 [33].

Before summarizing the main results presented here, we recall the definition of FRB dispersion measure (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:

$$\mathrm{DM} \equiv \int n_e(x)\, dx\,. \tag{3.1}$$

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 [68]. We write the total DM as the sum of contributions from our Galaxy and halo ($\mathrm{DM_{gal}}$), the IGM ($\mathrm{DM_{IGM}}$), and the FRB host galaxy and halo ($\mathrm{DM_{host}}$):

$$\mathrm{DM} = \mathrm{DM_{gal}} + \mathrm{DM_{IGM}}(z) + \mathrm{DM_{host}}\,. \tag{3.2}$$

[1] For a complete list of known FRBs, see https://www.herta-experiment.org/frbstats [118] or the TNS [89].

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

$$\mathrm{DM}_{\mathrm{IGM}}(z) = n_{e,0} \int_0^z dz'\, f_d(z') \frac{1+z'}{H(z')}\,, \tag{3.3}$$

where $f_d(z)$ is the mean electron ionization fraction at redshift z, $n_{e,0} = 2.13 \times 10^{-7}\,\mathrm{cm}^{-3}$ is the comoving electron density, and $H(z)$ is the Hubble expansion rate. If f_d is assumed independent of redshift, then Eq. (3.3) has the following useful approximation:

$$\mathrm{DM}_{\mathrm{IGM}}(z) \approx \left(1000\ \mathrm{pc\,cm}^{-3}\right) f_d z \tag{3.4}$$

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 $\mathrm{DM}_{\mathrm{IGM}}(z) \approx 900z\ \mathrm{pc\,cm}^{-3}$.

We briefly summarize the main results of the chapter. We find a statistically significant correlation between CHIME FRBs and galaxies in the redshift range $0.3 \lesssim z \lesssim 0.5$. The correlation is seen in three photometric galaxy surveys: WISE×SCOS, DESI-BGS, and DESI-LRG (described in §3.2.2). 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 $\geq 785\ \mathrm{pc\,cm}^{-3}$) FRBs and galaxies at $z \sim 0.4$. This implies the existence of an FRB subpopulation with host DM $\gtrsim 400\ \mathrm{pc\,cm}^{-3}$. 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 [27, 19]) have been localized to host galaxies, all of which have $\mathrm{DM}_{\mathrm{host}} \lesssim 200\ \mathrm{pc\,cm}^{-3}$. In §3.4.2, we explain why these observations are not in conflict. We also show that host DMs $\gtrsim 400\ \mathrm{pc\,cm}^{-3}$ can arise from ionized gas in large ($M \gtrsim 10^{14}\,M_\odot$) dark matter halos, if FRBs are located near the halo center ($r \lesssim 100$ kpc).

This chapter is structured as follows. In §3.2, we describe the observations and data reduction. Clustering results are presented in §3.3 and interpreted in §3.4. We conclude in §3.5. Throughout, we adopt a flat ΛCDM 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 $\sum_\nu m_\nu = 0.06$ eV, and CMB temperature $T_{\rm CMB} = 2.726$ K. These parameters are consistent with Planck results [3].

3.2 Data

3.2.1 FRB catalog

The first CHIME/FRB catalog is described in [29]. 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 non-repeating 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 non-repeaters have different sky distributions [29]. 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 chapter, *all DM values are extragalactic.* That is, before further processing of the CHIME FRBs, we subtract the Galactic contribution $\mathrm{DM_{gal}}$ from the observed DM. The value of $\mathrm{DM_{gal}}$ is estimated using the YMW16 [126] model. In §3.4.2, we show that using the NE2001 [34] model does not affect results qualitatively. The CHIME/FRB extragalactic DM distribution is shown in Figure 3.1.

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 $10 \lesssim \mathrm{DM_{halo}} \lesssim 100$ pc cm^{-3}, and the (dipole-dominated) anisotropy is expected to be small [93, 55]. The results of this chapter are qualitatively unaffected by the value of $\mathrm{DM_{halo}}$.

The CHIME/FRB pipeline assigns a nominal sky location to each FRB based on the observed 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 multi-beam detections, the nominal location is roughly a weighted average of the beam centers [25, 29]. Statistical errors on CHIME/FRB locations are difficult to model, since they depend on both the details of the CHIME telescope, and selection biases which depend on the underlying FRB population. We discuss this further in §3.3.1 and Appendix C.

Survey	$f_{\rm sky}$	$[z_{\rm min}, z_{\rm max}]$	$z_{\rm med}$	$N_{\rm gal}$	$N_{\rm FRB}$
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 3.1: Galaxy survey parameters: sky fraction $f_{\rm sky}$ (not accounting for CHIME/FRB coverage), redshift range $[z_{\rm min}, z_{\rm max}]$, median redshift $z_{\rm med}$, total number of unmasked galaxies $N_{\rm gal}$, and number of FRBs $N_{\rm FRB}$ overlapping the survey. The "BGS+LRG" catalog is used only in §3.3.5, and consists of all unique objects from the DESI-BGS and DESI-LRG catalogs.

3.2.2 Galaxy catalogs

On the galaxy side, we have chosen five photometric redshift catalogs: 2MPZ, WISE×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 3.1 summarizes key properties of our reduced samples for the cross-correlation analysis, and the redshift distributions are shown in Figure 3.1.

The 2MASS Photometric Redshift (2MPZ) catalog [17] contains ∼1 million galaxies with $z \lesssim 0.3$ (redshift error $\sigma_z \sim 0.02$), enabling the construction of a 3D view of large-scale structure at low redshifts (see, e.g. [5, 8]). In this work, we use the mask made by [5] for the 2MPZ catalog. Following [17], 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, [18]) 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 [59], which includes probabilities $(p_{\rm gal}, p_{\rm star}, p_{\rm qso})$ for each object to be a galaxy, star, or quasar, respectively. We use objects with $p_{\rm gal} \geq 0.9$, which is consistent with the weighted mean purity of identified galaxies across the $W1$ band [59]. We use a standard mask[2] to remove the Galactic foreground, Magellanic Clouds and bright stars. Additionally, we mask out

[2] http://ssa.roe.ac.uk/WISExSCOS.html

regions that are contaminated visually due to their proximity to the Galactic plane:

$$\begin{aligned}
&(|b| \le 20^\circ) \quad \text{and} \quad ((0^\circ \le l \le 30^\circ) \text{ or } (330^\circ \le l \le 360^\circ)) \,, \\
&(|b| \le 18^\circ) \quad \text{and} \quad ((30^\circ \le l \le 60^\circ) \text{ or } (300^\circ \le l \le 330^\circ)) \,, \\
&(|b| \le 17^\circ) \quad \text{and} \quad (0^\circ \le l \le 360^\circ) \,.
\end{aligned} \tag{3.5}$$

The Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys [36] were designed to identify galaxies for spectroscopic follow-up. We use the catalogs from the DR8 release, with photometric redshifts from [130]. 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 $0.05 \le z \le 0.4$, $0.3 \le z \le 1$, and $0.6 \le z \le 1.4$ respectively (Figure 3.1).

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 (Dec $> 32.^\circ375$, $b \ge +17^\circ$), which contains $\sim$2 times as many CHIME/FRB sources as the Southern part. Note that the Nothern and Southern DESI surveys are obtained from different telescopes and may have different systematics. For the DESI-ELG sample, we impose the additional constraint $b \ge +45^\circ$ 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 [130]. 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 [106], [129], and [97] respectively. For BGS, we include both "faint" ($19.5 < r < 20$) and "bright" ($r < 19.5$) galaxies (terminology from [106]). 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.

3.3 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, then computing the spherical harmonic transform and the angular power spectrum. Error bars are assigned using mock FRB catalogs.

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

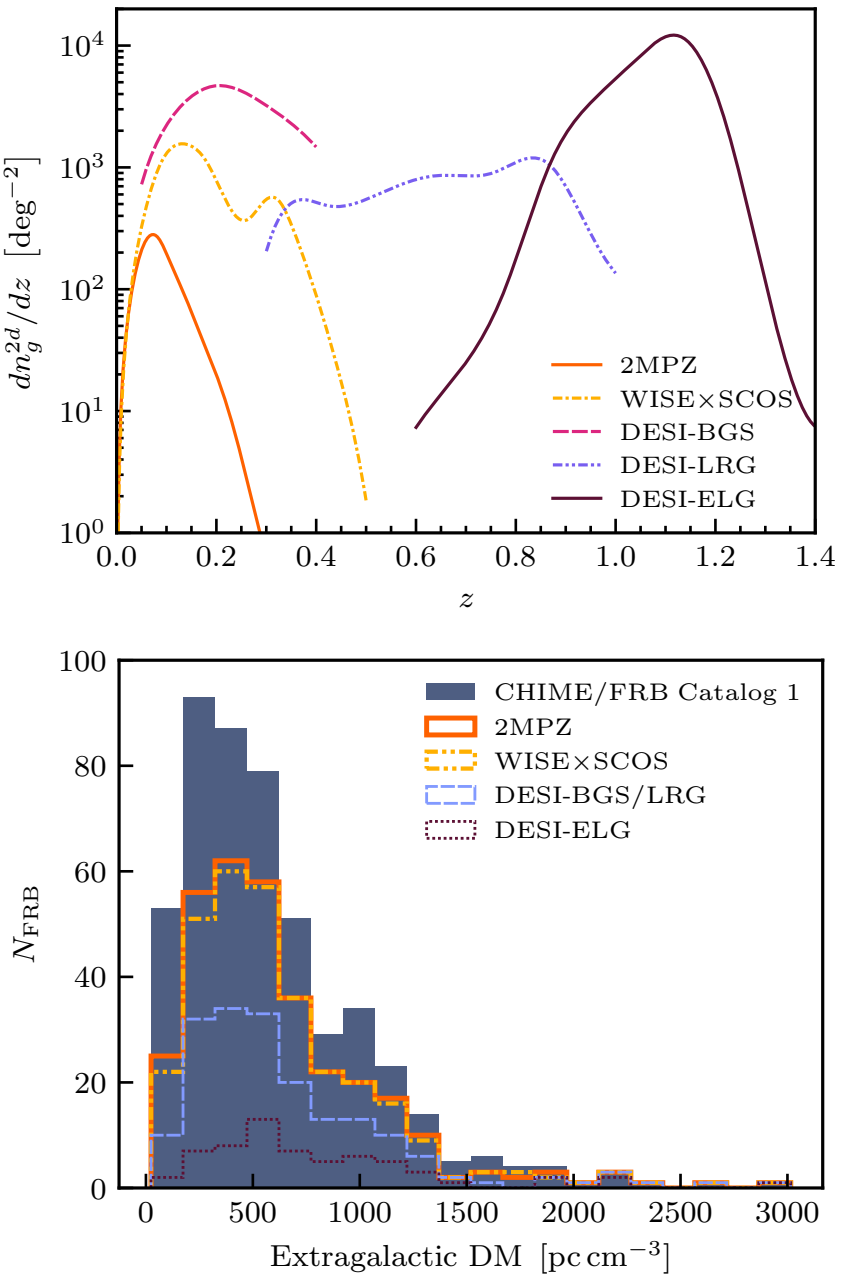

Figure 3.1: *Top panel:* Redshift distributions for the five galaxy samples in this chapter (§3.2.2). *Bottom panel:* FRB extragalactic dispersion measure distributions for the CHIME/FRB catalog (solid), and for the subset of the CHIME/FRB catalog which overlaps spatially with each galaxy survey.

3.3.1 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:

$$\left\langle a^f_{\ell m} a^{g*}_{\ell' m'} \right\rangle = C_\ell^{fg} \delta_{\ell\ell'} \delta_{mm'} \,, \tag{3.6}$$

where $a^Y_{\ell m}$ 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} 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 non-overlapping bins. This method was used by [63] 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 non-overlapping ℓ-bins are nearly uncorrelated, making it straightforward to infer statistical significance from plots.

Throughout the chapter, it will be useful to have a model FRB-galaxy power spectrum C_ℓ^{fg} in mind. In Figure 3.2, we show C_ℓ^{fg} for a galaxy population at $z \sim 0.4$, calculated using the "high-z" FRB model from §2 [95], 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).

- 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 3.2) 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.

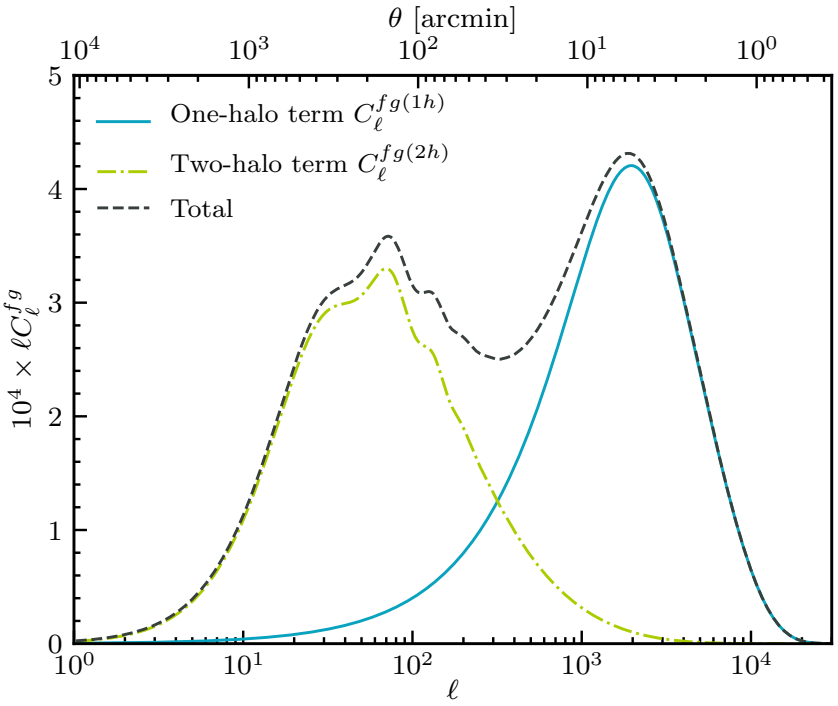

Figure 3.2: Model FRB-galaxy power spectrum C_ℓ^{fg} from §3.3.1, for a galaxy population near $z \sim 0.37$ and FRB angular resolution $1'$. Note that we have plotted (ℓC_ℓ^{fg}), for consistency with later plots in the chapter. In this and later plots in the chapter, the angular scale on the top axis is $\theta = \pi/\ell$, and is intended to provide an intuitive mapping between angular multipole ℓ and an angular scale.

Although the one-halo and two-halo terms look comparable in Figure 3.2, the SNR of the one-halo term is a few times larger. In this chapter, we do not detect the two-halo term with statistical significance (see Figures 3.7–3.9). Therefore, throughout the chapter 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 3.2, we have plotted ℓC_ℓ^{fg}, for consistency with later figures in the chapter.) 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 chapter, both effects can

be modeled as Gaussian, i.e. high-ℓ suppression of the form $e^{-\ell^2/L^2}$:

$$C_\ell^{fg(1h)} = \alpha e^{-\ell^2/L^2}\,, \tag{3.7}$$

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. (3.7) 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 C, we explore modeling issues in detail, and show that a plausible (but conservative, i.e. wide) range of L-values is $315 \le L \le 1396$.

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. (3.7). We treat the amplitude α as a free parameter, and vary the template scale L over the range $315 \le L \le 1396$, to evaluate the correlation amplitude as a function of scale.

3.3.2 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:

$$\delta_Y(\mathbf{x}) = \frac{1}{n_Y^{2d}\,\Omega_{\rm pix}}\Big(N_{Y\in\mathbf{x}} - \overline{N_{Y\in\mathbf{x}}}\Big) \tag{3.8}$$

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 $\overline{N_{Y\in\mathbf{x}}}$ denotes the expected number of catalog objects in pixel $\mathbf{x}$ due to the survey geometry. The prefactor $1/(n_Y^{2d}\Omega_{\rm pix})$ is conventional, where n_Y^{2d} is the 2D number density and $\Omega_{\rm pix}$ is the pixel area. For CHIME/FRB, the expected number density $\overline{N}_{f\in\mathbf{x}}$ depends on declination. The definition (3.8) 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. (3.8), 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 $n_g^{2d} = 5000$ 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 [42] mask, and assume uniform galaxy density outside the mask:

$$\bar{N}_{g\in\mathbf{x}} = \begin{cases} n_g^{2d}\,\Omega_{\rm pix} & \text{if } \mathbf{x} \text{ is unmasked} \\ 0 & \text{if } \mathbf{x} \text{ is masked} \end{cases} \tag{3.9}$$

The mask geometries for 2MPZ and WISE×SCOS were described previously in §3.2.2.

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 (Dec), because CHIME is a cylindrical drift-scan telescope oriented North-South [29]. 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. 3.8).

In Figure 3.3, 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 [50], who found no evidence for Galactic latitude dependence in the CHIME/FRB number density after correcting for selection effects. As described in §3.2.2, 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.

3.3.3 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:

$$\hat{C}_\ell^{fg} = \frac{1}{f_{\rm sky}^{fg}} \sum_{m=-\ell}^{\ell} \frac{1}{2\ell+1} a_{\ell m}^{f*} a_{\ell m}^{g} \tag{3.10}$$

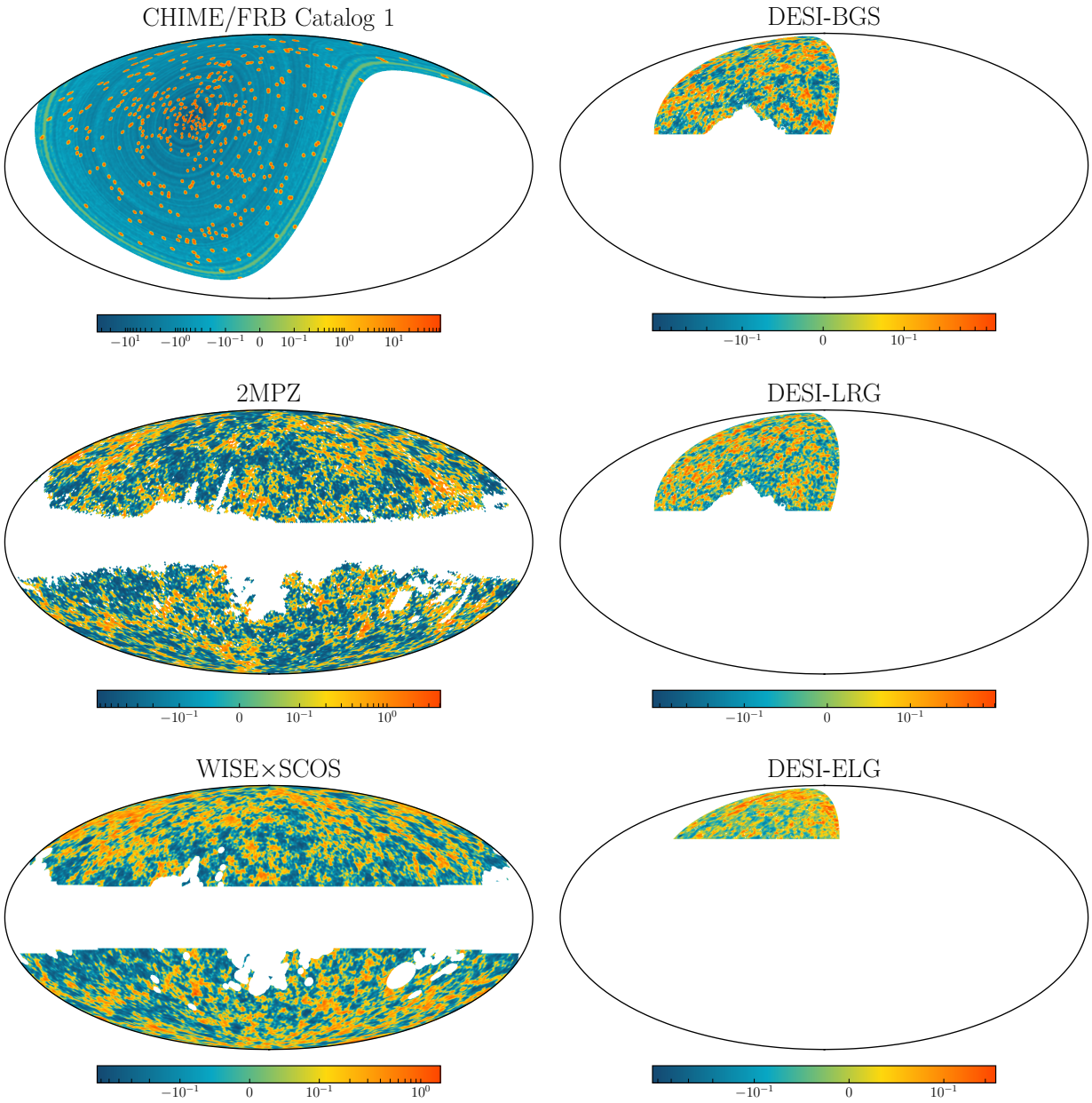

Figure 3.3: CHIME/FRB overdensity map $\delta_f(\mathbf{x})$, and galaxy overdensity maps $\delta_g(\mathbf{x})$ for each galaxy survey. Maps are shown in Mollweide projection, centered on $l = 180°$ in the Galactic coordinate system, after applying the angular masks used in the analysis pipeline. To interpret the color scale, note that by Eq. (3.8), each object in a pixel contributes $1/(n_Y^{2d}\Omega_{\rm pix})$ to the overdensity δ_Y.

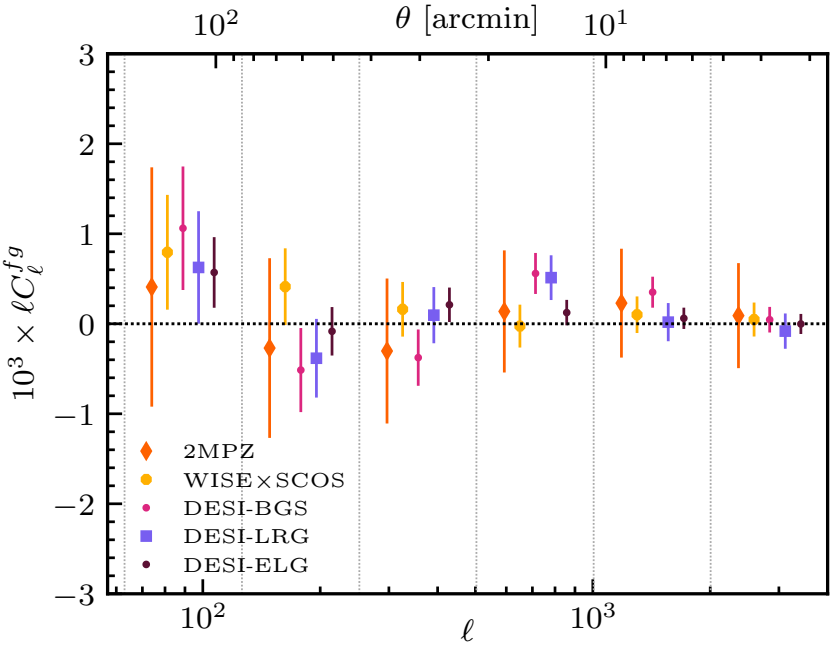

Figure 3.4: FRB-galaxy cross power spectrum C_ℓ^{fg} in a set of non-overlapping ℓ bins delimited by vertical lines, with 1σ error bars. Data points are shifted slightly from the center of corresponding ℓ bins for visual clarity. Here, we have used all galaxies in the catalogs; if we restrict the redshift ranges, then the correlation is more significant (Figures 3.7–3.9).

where $f_{\rm sky}^{fg}$ is the fractional sky area subtended by the intersection of the FRB and galaxy surveys. The $f_{\rm sky}^{fg}$ 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.′7 resolution ($N_{\rm side} = 2048$), and estimate the power spectrum to a maximum multipole of $\ell_{\rm max} = 2000$, corresponding to angular scale $\theta = \pi/\ell_{\rm max} = 5.'4$.

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 §3.3.2. 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

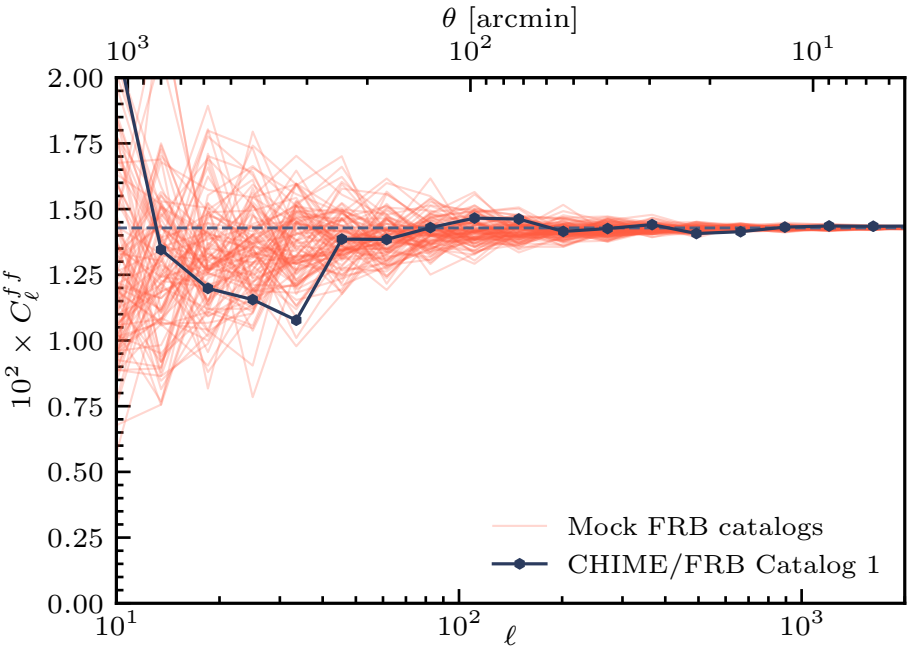

Figure 3.5: Angular auto power spectrum C_ℓ^{ff} for the CHIME/FRB catalog. Transparent lines represent 100 mock FRB catalogs that spatially model the real data. Throughout the analysis, we assume that the power spectrum C_ℓ^{ff} approaches a constant (dashed line) on small scales (high ℓ). Specifically, $C_\ell^{ff} \approx 1/n_f^{2d}$ for $315 \leq \ell \leq 1396$.

represent the survey geometry. For FRBs, spatial clustering is small compared to Poisson noise ($C_\ell^{ff} \approx 1/n_f^{2d}$, see Figure 3.5), so we can make the approximation that clustering is negligible.

3.3.4 Statistical significance and look-elsewhere effect

In Figure 3.4, we show the angular power spectrum C_ℓ^{fg} for a set of non-overlapping 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 §3.3.1, 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:

$$\hat{\alpha}_L = \frac{1}{\mathcal{N}_L} \sum_{\ell \geq \ell_{\rm min}} (2\ell+1) \frac{e^{-\ell^2/L^2}}{C_\ell^{gg}} \hat{C}_\ell^{fg} \,, \tag{3.11}$$

where $\hat{C}_\ell^{fg}$ was defined in Eq. (3.10), and the normalization $\mathcal{N}_L$ is defined by:

$$\mathcal{N}_L = \sum_{\ell \geq \ell_{\rm min}} (2\ell+1) \frac{e^{-2\ell^2/L^2}}{C_\ell^{gg}} \,. \tag{3.12}$$

We have included a cutoff at $\ell_{\rm min} = 50$ to mitigate possible large-scale systematics. This is a conservative choice, since Figure 3.5 does not show evidence for systematic power in the auto power spectrum C_ℓ^{ff} for $\ell \gtrsim 15$. Eq. (3.11) is derived by noting that:

$$\begin{aligned} {\rm Var}(\hat{C}_\ell^{fg}) &\propto \frac{C_\ell^{ff} C_\ell^{gg} + (C_\ell^{fg})^2}{2\ell+1} \\ &\propto \frac{C_\ell^{gg}}{2\ell+1} \,, \end{aligned} \tag{3.13}$$

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:

$${\rm SNR}_L = \frac{\hat{\alpha}_L}{{\rm Var}(\hat{\alpha}_L)^{1/2}} \,, \tag{3.14}$$

which is the statistical significance of the C_ℓ^{fg} detection in "sigmas", for a fixed choice of L. In Figure 3.6, we show the quantity ${\rm SNR}_L$, as a function of scale L.

We pause for a notational comment: throughout the chapter, ℓ denotes a multipole (as in C_ℓ^{fg}), and L denotes the template scale defined in Eq (3.7). The value of ${\rm SNR}_L$ (or $\hat{\alpha}_L$) is obtained by summing C_ℓ^{fg} over $\ell \lesssim L$, as in Eq. (3.11). When C_ℓ^{fg} is computed as a function of ℓ (Figure 3.4), neighboring ℓ bins are nearly uncorrelated, whereas when ${\rm SNR}_L$ is computed as a function of L (Figure 3.6), nearby L-values are highly correlated.

In Figure 3.6, it is seen that ${\rm SNR}_L$ 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.

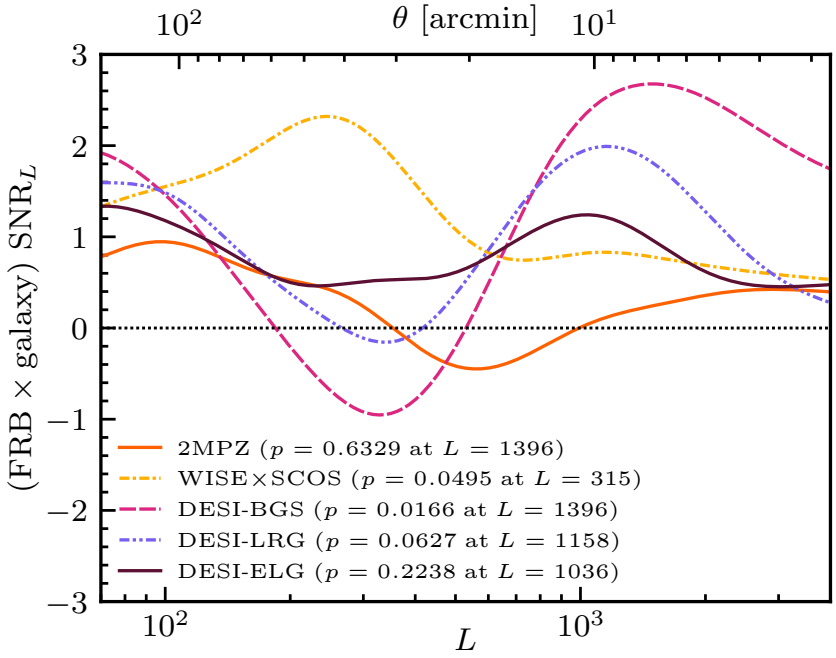

Figure 3.6: Quantity SNR_L, defined in Eq. (3.14), as a function of template scale L. As explained in §3.3.4, SNR_L is the statistical significance of the FRB-galaxy correlation in "sigmas", for a fixed choice of L. The p-values in the legend are bottom-line detection significances after accounting for 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 (Figures 3.7–3.9).

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

$$\mathrm{SNR}_{\mathrm{max}} = \max_{315 \leq L \leq 1396} \mathrm{SNR}_L \tag{3.15}$$

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

In more detail, we compare the "data" value of $\mathrm{SNR}_{\mathrm{max}}$ (e.g. $\mathrm{SNR}_{\mathrm{max}} = 2.67$ for DESI-BGS) to an ensemble of Monte Carlo simulations, obtained by cross-correlating mock FRB catalogs with the real galaxy catalog as in §3.3.3. We assign a p-value by computing the fraction of mocks with $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{mock})} \geq \mathrm{SNR}_{\mathrm{max}}^{(\mathrm{data})}$. 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 3.6.

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 3.4 and 3.6 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.

3.3.5 Redshift dependence

To illustrate our method for studying redshift dependence, we will use the WISE×SCOS galaxy catalog as a running example. Suppose we cross-correlate CHIME FRBs with WISE×SCOS galaxies above some minimum redshift $z_{\rm min}$, where $z_{\rm min}$ is a free parameter that will be varied. For each $z_{\rm min}$, we repeat the analysis of the previous subsection. The power spectrum $C_\ell^{fg}(z_{\rm min})$ and quantity $\mathrm{SNR}_L(z_{\rm min})$ (defined in Eq. 3.14) are now functions of two parameters: $z_{\rm min}$ and template scale L.

In Figure 3.7, we show the power spectrum $C_\ell^{fg}(z_{\rm min})$ for the fixed choice of redshift $z_{\rm min} = 0.3125$, and $\mathrm{SNR}_L(z_{\rm min})$ as a function of L and $z_{\rm min}$. For specific parameter choices, we see a large FRB-galaxy correlation, e.g. $\mathrm{SNR}_L(z_{\rm min}) = 4.88$ at $L = 543$ and $z_{\rm min} = 0.3125$. As in the previous subsection, this would imply a 4.88σ cross-correlation for these cherry-picked values of $(L, z_{\rm min})$, 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 scan over two parameters $(L, z_{\rm min})$ rather than one (L). Formally, we define

$$\mathrm{SNR}_{\rm max} = \max_{0 \le z_{\rm min} \le 0.5} \; \max_{315 \le L \le 1396} \mathrm{SNR}_L(z_{\rm min}) \tag{3.16}$$

analogously to Eq. (3.15) from the previous subsection.

To assign bottom-line statistical significance, we would like to rank the "data" value $\mathrm{SNR}_{\rm max} = 4.88$ within a histogram of $\mathrm{SNR}_{\rm max}$ 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 $\mathrm{SNR}_{\rm max} = 4.88$, so we fit the tail of the $\mathrm{SNR}_{\rm max}$ distribution to an analytic distribution (a truncated Gaussian), and compute the p-value analytically. For details of the tail-fitting procedure, see Appendix E. We obtain detection significance $p = 2.7 \times 10^{-5}$ for WISE×SCOS with $z_{\rm min} = 0.3125$. This analysis "scans"

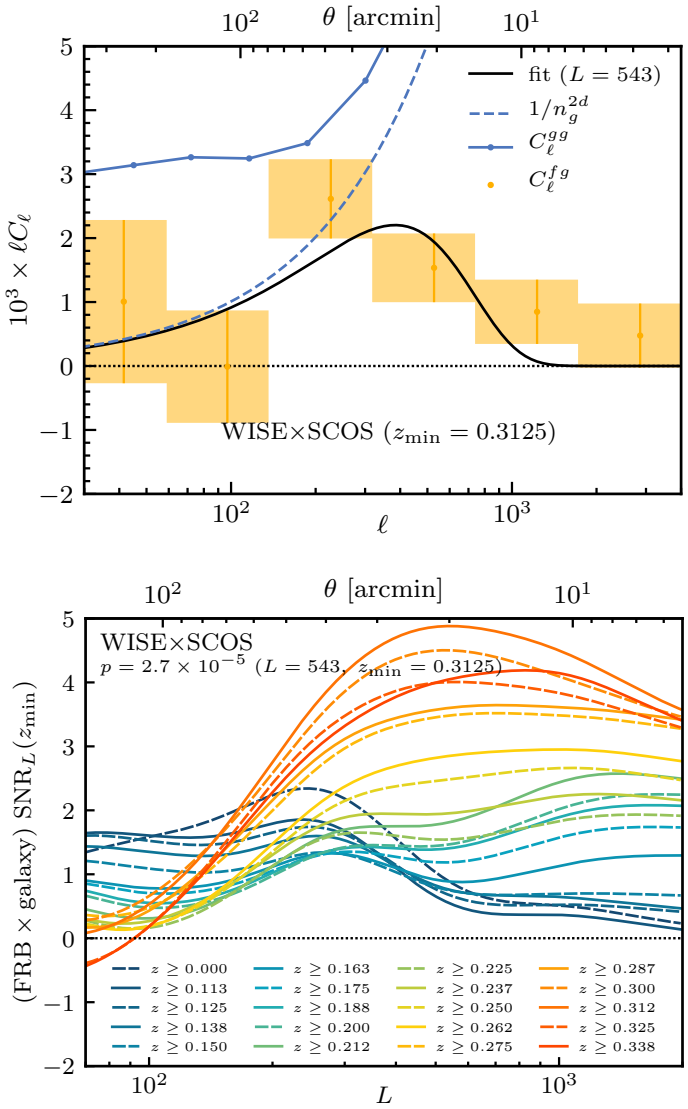

Figure 3.7: FRB-galaxy correlation analysis with two parameters: template scale L (defined in Eq. 3.7), and the redshift endpoint $z_{\rm min}$ for WISE×SCOS. *Top panel:* Angular cross power spectrum C_ℓ^{fg} and auto power spectrum C_ℓ^{gg}, for the fixed choice of redshift endpoint that maximizes FRB-galaxy correlation. The cross power "fit" is a best-fit template of the form $C_\ell^{fg} = \alpha e^{-\ell^2/L^2}$. *Bottom panel:* Quantity SNR_L, defined in Eq. (3.14), as a function of L and redshift endpoint. As explained in §3.3.4, 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 §3.3.5).

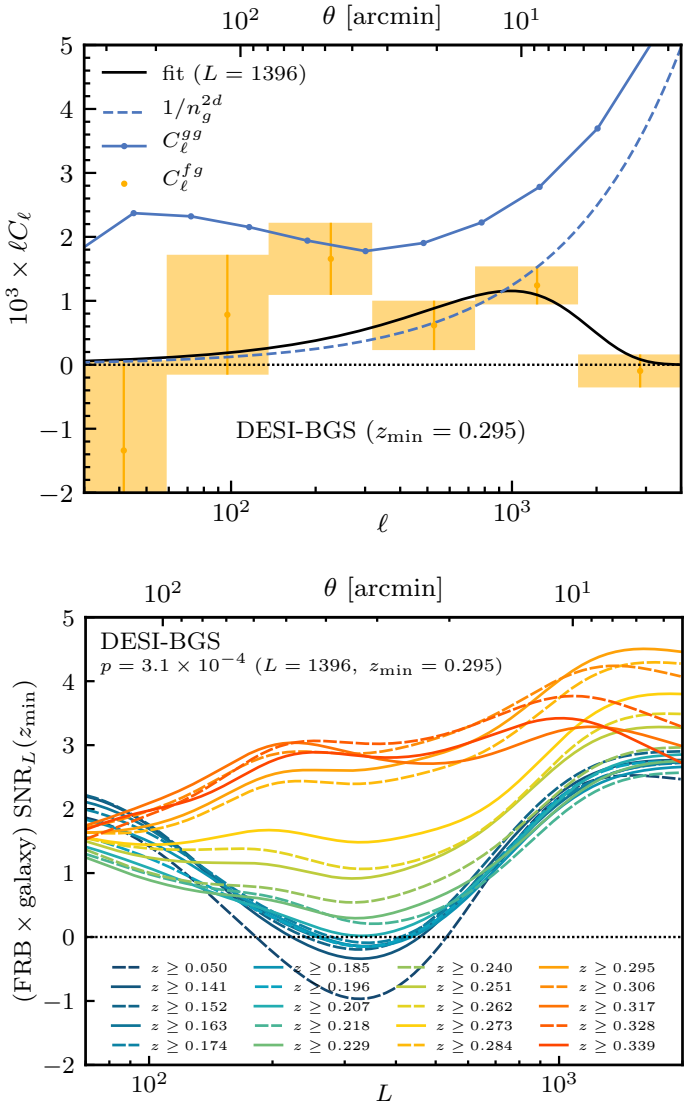

Figure 3.8: Same as Figure 3.7, but for DESI-BGS.

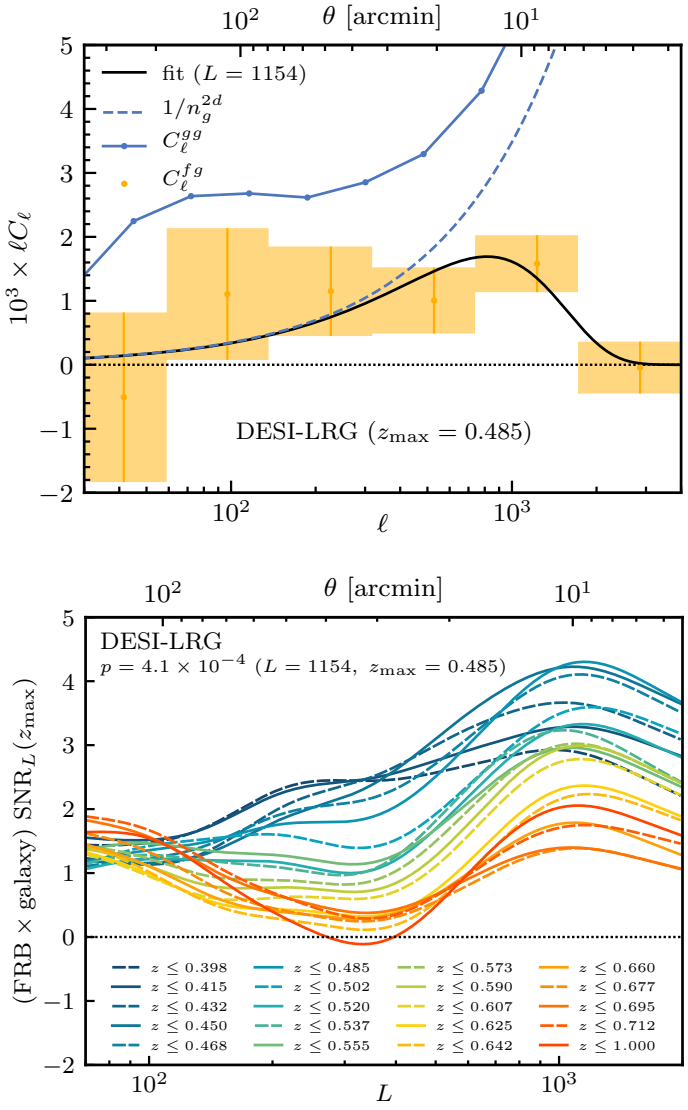

Figure 3.9: Same as Figure 3.7, but with the redshift endpoint $z_{\max}$ for DESI-LRG.

over minimum redshift $z_{\rm min}$ 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 $z_{\rm min} = 0.295$, scanning over $z_{\rm min}$ and L. For DESI-LRG, we use a *maximum* redshift $z_{\rm max}$ instead of a minimum redshift $z_{\rm min}$, since DESI-LRG is at higher redshift than WISE×SCOS or DESI-BGS (Figure 3.1). Scanning over $z_{\rm max}$ and scale L, we obtain $p = 4.1 \times 10^{-4}$ with $z_{\rm max} = 0.485$ for DESI-LRG. These results are shown in Figures 3.7–3.9.

Finally, we find borderline evidence $p = 0.0421$ ($L = 1396, z_{\rm max} = 0.86$) for a cross-correlation between DESI-ELG galaxies (varying $z_{\rm max}$) and CHIME FRBs with DM ≥ 500 pc cm^{-3}, where the choice of minimum DM is fixed. To justify this choice of DM$_{\rm min}$, note that since host DMs must be positive, we do not expect a correlation between DESI-ELG galaxies ($z_{\rm min} = 0.6$) and CHIME FRBs with DM < 500 pc cm^{-3} (allowing for statistical fluctuations in DM$_{\rm IGM}$ 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 $z_{\rm min} \sim 0.3$. For DESI-LRG, the minimum survey redshift is 0.3, and we find a strong detection when we impose a maximum redshift $z_{\rm max} \sim 0.5$. The borderline detection in DESI-ELG and non-detection 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 3.10 we cross-correlate the FRB catalog with the combined BGS+LRG catalog (Table 3.1, bottom row) in non-overlapping 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 ten bins.)

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

3.4 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.

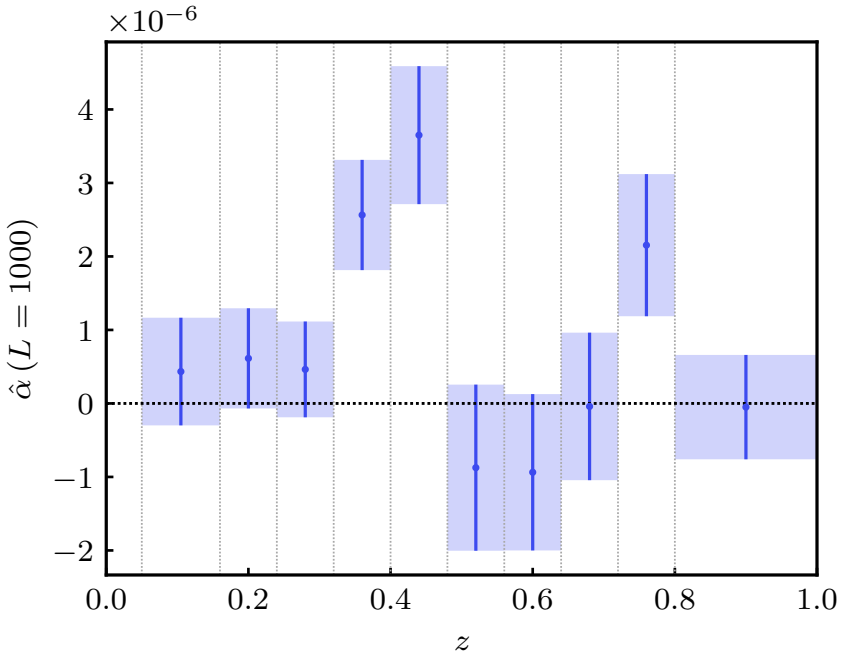

Figure 3.10: Redshift dependence of the FRB-galaxy correlation. We divide the BGS+LRG catalog into non-overlapping redshift bins (dotted lines), and cross-correlate with CHIME FRBs. The quantity $(\hat{\alpha}_L)_{L=1000}$ on the y-axis is a measure of the level of cross-correlation, defined in Eq. (3.11).

In this section, we will interpret the FRB-galaxy correlation, and explore implications for FRBs.

As explained in §3.3.1, the output of our pipeline is a constraint on the coefficient α in the template fit:

$$C_\ell^{fg} \approx C_\ell^{fg(1h)} \approx \alpha e^{-\ell^2/L^2} \tag{3.17}$$

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 intepret the low-ℓ limit of the model as a prediction for the coefficient α above. Formally, we define

$$\alpha \equiv \lim_{\ell \to 0} C_\ell^{fg(1h)} \tag{3.18}$$

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. (3.11). 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 Figures 3.7–3.9).

3.4.1 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:

$$\eta_f = \text{number of survey galaxies linked to FRB } f. \tag{3.19}$$

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

$$\eta = \langle \eta_f \rangle \tag{3.20}$$

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:

$$\alpha = \lim_{\ell \to 0} C_\ell^{fg(1h)} = \frac{\eta}{n_g^{2d}} \tag{3.21}$$

where the first equality is Eq. (3.18), and the second equality follows from a short halo model calculation (§2). 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 3.2, we specify a choice of galaxy catalog and redshift range. The redshift ranges have been chosen to maximize C_ℓ^{fg}, as in §3.3.4. 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. (3.21).

Taken together, the η measurements in Table 3.2 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 which 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 which are in the catalog redshift range.

Survey	$[z_{\rm min}, z_{\rm max}]$	$(\alpha_L)_{L=1000}$	n_g^{2d} [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 3.2: Clustering analysis in §3.4.1. The FRB-galaxy clustering statistic α_L (Eq. 3.11) 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}.

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 Figures 3.7–3.9). By a link counting argument similar to Eq. (3.21), 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 3.2), 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 ($\ell \sim 100$) (§2), which will help break the degeneracy and measure (dn_f^{2d}/dz) and $\tilde{\eta}$ separately.

3.4.2 DM dependence

In Figure 3.11, we divide the FRB catalog in extragalactic DM bins, and explore DM dependence of the FRB-galaxy cross-correlation.

A striking feature in Figure 3.11 is the nonzero correlation in the three highest-DM bins, corresponding to extragalactic DM ≥ 785 pc cm^{-3}.[4] For reference, the last 3 bins

[4] A technical comment here: for some DM bins in Figure 3.11, the large values of C_ℓ^{fg} lead to link counts η which are a few times larger than the link counts reported in Table 3.2 for the whole catalog, although

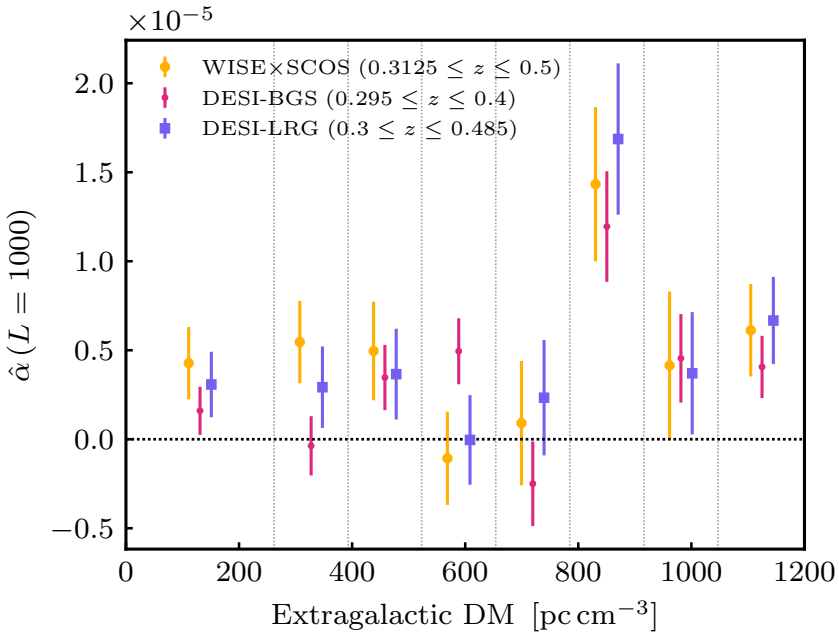

Figure 3.11: DM dependence of the FRB-galaxy correlation. We divide the CHIME/FRB catalog into DM bins (delimited by vertical lines) after subtracting the YMW16 estimate of the Galactic DM, and cross-correlate each DM bin with the galaxy catalogs. The last DM bin extends to DM = 3020 pc cm^{-3}. For each galaxy survey, we use the same redshift range (see legend) as in the top panel of Figures 3.7–3.9. The quantity $(\hat{\alpha}_L)_{L=1000}$ on the y-axis is defined in Eq. (3.11), 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 2.1

.

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 $\mathrm{DM_{IGM}}(z) \sim 360$ pc cm^{-3}. Therefore, the observed FRB-galaxy correlation at DM ≥ 785 pc cm^{-3} is evidence for a subpopulation of FRBs with host DMs of order $\mathrm{DM_{host}} \sim 400$ 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 $\mathrm{DM_{host}} \lesssim 200$

statistical errors are large. However, the correlation coefficient between the FRB and galaxy fields is never larger than one. 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

pc cm^{-3} [120, 21, 122, 57, 11, 92, 98, 9, 28, 30, 15, 14, 70, 72, 68, 114, 60, 43, 46, 16]. The rest of this section is devoted to interpreting this result further.

In Figure 3.11, the DM bin at $785 < \mathrm{DM} < 916$ 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 3.11 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 3.11 using the NE2001 [34] model for Galactic DM, instead of the YMW16 model. The effect of this change is small compared to the statistical errors in Figure 3.11.

We also performed the following visual check. The outlier bin with $785 < \mathrm{DM} < 916$ pc cm^{-3} in Figure 3.11 only contains 12 FRBs in the DESI footprint. In Figure 3.12 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 3.11 gives a statistically significant cross-correlation on its own, but the total FRB-galaxy correlation is significant at the 3–4σ level. (We caution the reader that the galaxy counts in Figure 3.12 do not obey Poisson statistics, since the galaxies are clustered.) There are no visual red flags in Figure 3.12, 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 3.11 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 ≥ 785 pc cm^{-3} has an order-one probability of having redshift $z \sim 0.4$ (implying $\mathrm{DM_{host}} \gtrsim 400$ pc cm^{-3}).

2. A random FRB at redshift $z \sim 0.4$ has an order-one probability of having extragalactic DM ≥ 785 pc cm^{-3}.

The high-DM signal in Figure 3.11 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 Figures 3.7–3.9. The factor $\sqrt{8}$ was derived by matching the variance $(\Theta^2/2)$ of a radius-Θ tophat to the variance $(4/L^2)$ of a Gaussian beam $e^{-\theta^2 L^2/4}$.

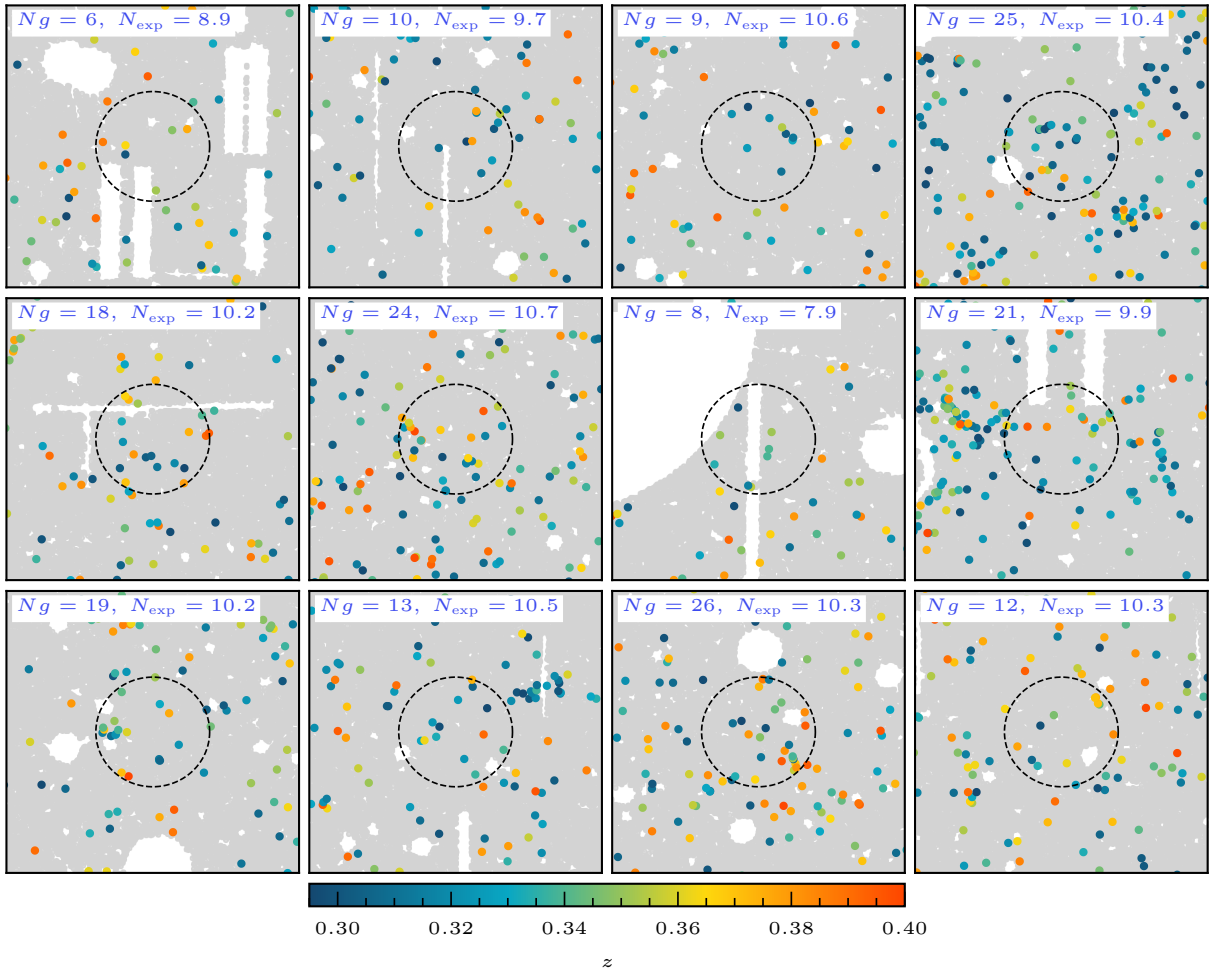

Figure 3.12: Visual representation of the cross-correlation between FRBs with $785 < \mathrm{DM} < 916\ \mathrm{pc\,cm^{-3}}$, and DESI-BGS galaxies. There are 12 FRBs in this DM range in the DESI footprint. For each such FRB, we plot the DESI-BGS galaxies in the redshift range $0.295 < z < 0.4$ in the 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 §3.4.2). The value of N_g in the upper left is the observed number of galaxies in the circle. The value of $N_{\rm exp}$ is the expected number of galaxies in the circle, inferred from randoms. The FRB-galaxy correlation appears as a statistical preference for $N_g > N_{\rm exp}$.

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 ≥ 785 pc cm^{-3}: an FRB with YMW16-subtracted DM 850 pc cm^{-3} at $z = 0.6$ [60]. 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 3.11, and direct FRB host associations to date. The number of direct associations is rapidly growing, and we predict that FRBs with extragalactic DM ≥ 785 pc cm^{-3} at $z \sim 0.4$ will be found in direct associations soon (see §3.5 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 which are sensitive to fainter sources may detect larger numbers of high-redshift FRBs. In this scenario, it is possible that FRBs with extragalactic DM ≥ 785 pc cm^{-3} will mostly come from $z \sim 0.8$, as expected from the Macquart relation.

3.4.3 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 sightlines 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 [125]. 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$_{\rm hh}$, since the term "halo DM" is often used to refer to the contribution from the Milky Way halo.

Can host halo DMs plausibly be of order DM$_{\rm hh} \gtrsim 400$ pc cm^{-3}? To answer this question, in Figure 3.13, we show DM$_{\rm hh}$ histograms for simulated FRBs in a halo of mass $M = 10^{14} M_\odot$. The halo gas profile is the "ICM" model from [93], based on X-ray observations from [124][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 DM$_{\rm hh} \gtrsim 400$ pc cm^{-3}.

[7] To calculate the host halo DM DM$_{\rm hh} = \int dr\, n_e(r)$, 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.

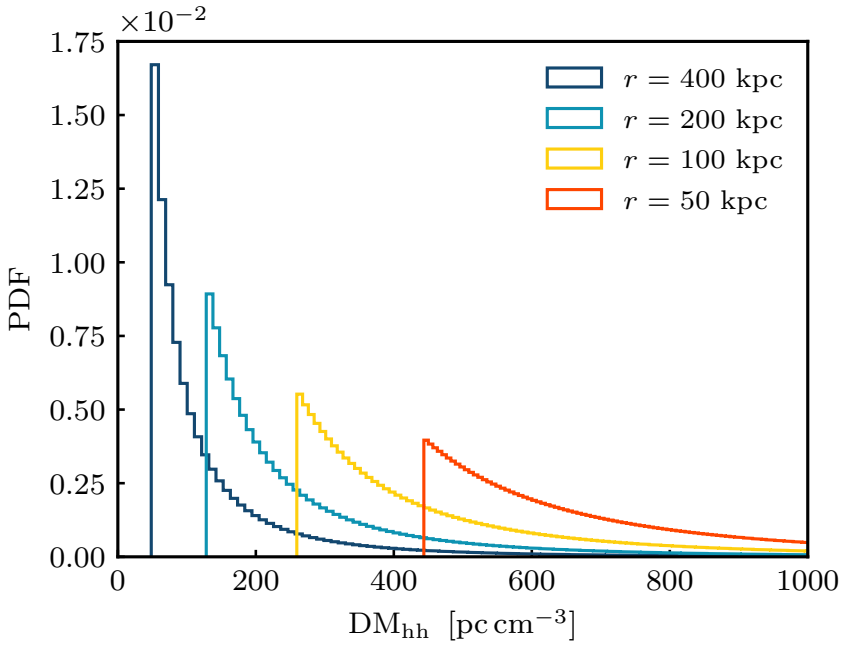

Figure 3.13: Host halo DM distributions for FRBs in a halo of mass $M = 10^{14} M_\odot$. The host halo DM is determined by two parameters: the distance r between the FRB and halo center, and viewing angle θ. Each histogram corresponds to one choice of r, with 10^5 values of θ. The halo gas profile is the "ICM" model from [93].

Thus, the high-DM signal in Figure 3.11 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 $\mathrm{DM_{host}} \gtrsim 400\ \mathrm{pc\,cm^{-3}}$, and strongly correlate with galaxies, since bright galaxies are often in high-mass halos.

This mechanism is a proof-of-concept to show that $\mathrm{DM_{hh}} \gtrsim 400\ \mathrm{pc\,cm^{-3}}$ is plausible in some halo gas models. Other mechanisms may also be possible, such as augmentation by intervening foreground galaxies [47]. We emphasize that the statistical evidence for a population of FRBs with $\mathrm{DM_{host}} \gtrsim 400\ \mathrm{pc\,cm^{-3}}$, presented in previous subsection, does not depend on the assumption of a particular model or mechanism.

3.4.4 Propagation effects

So far, we have assumed that the observed FRB-galaxy correlation is due 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 $\mathrm{DM_{host}} \sim 70\ \mathrm{pc\,cm^{-3}}$), 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 anti-correlation 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 §2. We will use a fiducial model in which host DMs are small ($\mathrm{DM_{host}} \sim 70\ \mathrm{pc\,cm^{-3}}$), 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 3.11. We describe the fiducial model in the next few paragraphs.

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

$$\frac{dn_f^{2d}}{dz} \propto z^2 e^{-\gamma z} \tag{3.22}$$

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

$$p(\mathrm{DM_{host}}) \propto \frac{1}{\mathrm{DM_{host}}} \exp\left(-\frac{(\log \mathrm{DM_{host}} - \mu_{\log})^2}{2\sigma_{\log}^2}\right) . \tag{3.23}$$

In Eqs. (3.22), (3.23), we choose parameters:

$$\gamma = 6.7 \qquad \mu_{\log} = 4.2 \qquad \sigma_{\log} = 5 \tag{3.24}$$

The total DM is $\mathrm{DM} = \mathrm{DM_{IGM}}(z) + \mathrm{DM_{host}}$. These parameters have been chosen so that the median FRB redshift is 0.4, the median host DM is $67\ \mathrm{pc\,cm^{-3}}$, and the distribution of total DMs is similar to the observed DM distribution in Figure 3.14.

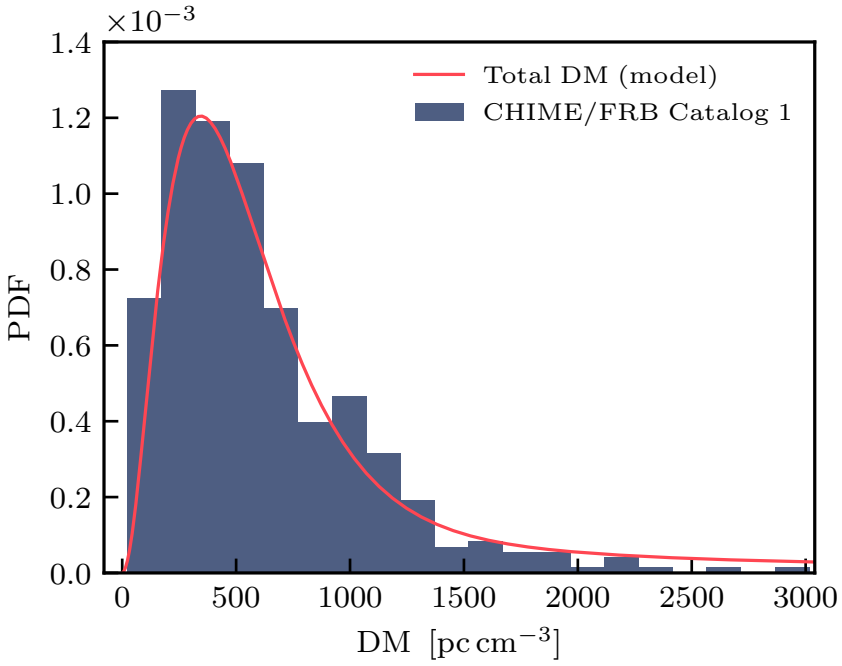

Figure 3.14: DM distribution (solid curve) for the fiducial FRB model used to study propagation effects in §3.4.4, with the CHIME/FRB DM distribution shown for comparison (histogram). In this model, host DMs are small, to explore the hypothesis that the correlation between $z \sim 0.4$ galaxies and high-DM FRBs is due to propagation effects, rather than large host DMs. The host DM distribution (not shown) is sharply peaked at $\mathrm{DM_{host}} \sim 70$ pc cm^{-3}.

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 $k \to 0$, which is (§2):

$$\lim_{k\to 0} P_{ge}^{1h}(k,z) = \frac{\langle N_e^{\rm ion}\rangle}{n_{e,0}}\,, \tag{3.25}$$

where $\langle N_e^{\rm ion}\rangle$ 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 $\langle N_e^{\rm ion}\rangle$, we assume that survey galaxies are contained in dark matter halos whose mass M_h is lognormal-distributed, with parameters:

$$\big\langle \lambda \big\rangle = 13.4 \qquad \sigma\big(\lambda\big) = 0.35 \tag{3.26}$$

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 [107] 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 $S(\mathrm{DM})$ 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 [29]. Here, we will use the following rough visual fit:

$$\log S(\mathrm{DM}) = 0.1 - 0.14\left[\log\left(\frac{\mathrm{DM}}{1000}\right)\right]^2\,. \tag{3.27}$$

The selection function $S(\mathrm{DM})$ 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 (RFI) removal. (Scattering biases will be discussed later in this section.) This combination of biases results in the selection function (Eq. 3.27) with a local maximum at DM ~ 1000 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 $dS/d(\mathrm{DM})$ 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 §2, the contribution to C_ℓ^{fg} from DM-completeness is calculated:

$$C_\ell^{fg} = \frac{1}{n_g^{2d}} \int dz \, \frac{H(z)}{\chi(z)^2} \frac{dn_g^{2d}}{dz} W_f(z) \, P_{ge}\left(\frac{\ell}{\chi(z)}, z\right), \tag{3.28}$$

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

$$\begin{aligned} W_f(z) &= \frac{n_{e,0}}{n_f^{2d}} \frac{1+z}{H(z)} \int_z^\infty dz' \\ &\quad \times \int_{\mathrm{DM}_{\mathrm{min}}}^{\mathrm{DM}_{\mathrm{max}}} d(\mathrm{DM}) \frac{d^2 n_f^{2d}}{dz' \, d(\mathrm{DM})} \frac{d \log S}{d(\mathrm{DM})} \end{aligned} \tag{3.29}$$

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:

$$\begin{aligned} \alpha &= \lim_{\ell \to 0} C_\ell^{fg(1h)} \\ &= \frac{1}{n_g^{2d}} \int dz \, \frac{H(z)}{\chi(z)^2} \frac{dn_g^{2d}}{dz} W_f(z) \frac{\langle N_e^{\mathrm{ion}}(z) \rangle}{n_{e,0}}, \end{aligned} \tag{3.30}$$

where we have used Eq. (3.18) in the first line, and Eqs. (3.25), (3.28) 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 3.11. 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 §2, the DM-shifting bias to α_L is given by the previous expression (3.30), but with the following expression for the DM-shifting weight function:

$$W_f(z) = -\frac{n_{e,0}}{n_f^{2d}} \frac{1+z}{H(z)} \int_z^\infty dz' \left[\frac{d^2 n_f^{2d}}{dz' \, d(\mathrm{DM})} \right]_{\mathrm{DM}_{\mathrm{min}}}^{\mathrm{DM}_{\mathrm{max}}} \tag{3.31}$$

In Figure 3.15, we show α_L-biases from the DM-completeness and DM-shifting propagation effects in our fiducial model, computed using Eqs. (3.29)–(3.31). For simplicity, we have

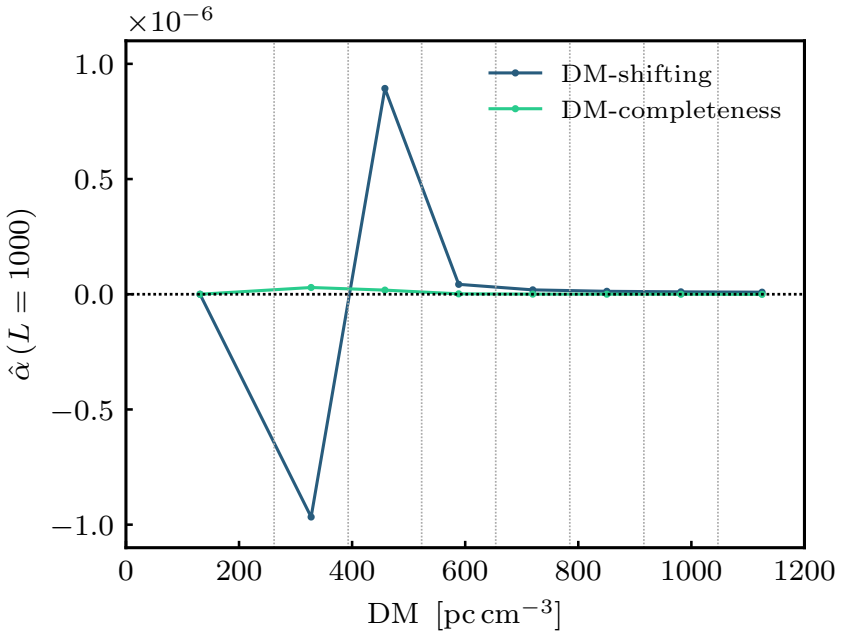

Figure 3.15: Predicted contribution to the FRB-galaxy correlation α_L (Eq. 3.11) from propagation effects, in the fiducial model from §3.4.4. The DM binning is the same as Figure 3.11. Comparing to the error bars in Figure 3.11, the DM-shifting contribution is $\sim 0.5\sigma$ in the second and third DM bins ($262 < \mathrm{DM} < 393$ and $393 < \mathrm{DM} < 523$ pc cm^{-3}), and $\lesssim 0.1\sigma$ in the other bins. The DM-completeness contribution is very small.

approximated the precise z-dependence of the redshift-binned galaxy surveys in Figure 3.11 by assuming $dn_g^{2d}/dz = \text{const}$ for $0.3 \le z \le 0.4$. (The results are not very sensitive to the galaxy redshift distribution.)

Comparing to the FRB-galaxy correlation shown previously in Figure 3.11, we see that the total bias is $\sim 0.5\sigma$ in the second DM bin ($262 < \mathrm{DM} < 393$ pc cm^{-3}), and $\lesssim 0.1\sigma$ 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 3.11.

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

3.5 Summary and conclusions

In this chapter, 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×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 ($\ell \sim 100$) (§2), which will help break this degeneracy.

[8] Formally, the selection function for scattering is a decreasing function of scattering width. This can be seen directly in Figure 15 of [29].

We find statistical evidence for a population of FRBs with large host DMs, on the order of $\mathrm{DM_{host}} \sim 400$ pc cm^{-3}. More precisely, we detect a nonzero correlation between FRBs with DM ≥ 785 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 $\mathrm{DM_{IGM}}(z) \sim 360$ 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 $\mathrm{DM_{host}} \lesssim 200$ pc cm^{-3}. However, FRBs with DM ≥ 785 pc cm^{-3} are presently 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 presently only one direct association with YMW16-subtracted DM ≥ 785 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 $\gtrsim 400$ 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 ($r \lesssim 100$ kpc) of large ($M \sim 10^{14} M_\odot$) halos, the host halo DM can be $\gtrsim 400$ pc cm^{-3} (Figure 3.13), at least in one widely-used ICM model [93]. 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 (§3.4.4).

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 dispersion measure, 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 [39]. Very high-z FRBs, if present, can be used to constrain cosmic reionization history [128, 20, 65]. 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 chapter 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 [79]. The FRB-galaxy correlation presented here should have much higher statistical significance in future CHIME/FRB catalogs, and will be exciting to explore.

References

[1] Peter Ade et al. The Simons Observatory: Science goals and forecasts. *JCAP*, 1902:056, 2019.

[2] Amir Aghamousa et al. The DESI Experiment Part I: Science,Targeting, and Survey Design. 2016.

[3] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, and et al. Planck 2018 results. *Astronomy & Astrophysics*, 641:A6, Sep 2020.

[4] Hiroaki Aihara et al. The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III. *Astrophys. J. Suppl.*, 193:29, 2011.

[5] D. Alonso, A. I. Salvador, F. J. Sánchez, M. Bilicki, J. García-Bellido, and E. Sánchez. Homogeneity and isotropy in the Two Micron All Sky Survey Photometric Redshift catalogue. *Monthly Notices of the Royal Astronomical Society*, 449(1):670–684, 12/9/2020 2015.

[6] David Alonso. Linear anisotropies in dispersion-measure-based cosmological observables. *arXiv e-prints*, page arXiv:2103.14016, 03 2021.

[7] Stefania Amodeo et al. The Atacama Cosmology Telescope: Modelling the Gas Thermodynamics in BOSS CMASS galaxies from Kinematic and Thermal Sunyaev-Zel'dovich Measurements. 9 2020.

[8] A. Balaguera-Antolínez, M. Bilicki, E. Branchini, and A. Postiglione. Extracting cosmological information from the angular power spectrum of the 2MASS Photometric Redshift catalogue. *Monthly Notices of the Royal Astronomical Society*, 476:1050–1070, 05 2018.

[9] K. W. Bannister, A. T. Deller, C. Phillips, J. P. Macquart, J. X. Prochaska, N. Tejos, S. D. Ryder, E. M. Sadler, R. M. Shannon, S. Simha, C. K. Day, M. McQuinn, F. O. North-Hickey, S. Bhandari, W. R. Arcus, V. N. Bennert, J. Burchett, M. Bouwhuis, R. Dodson, R. D. Ekers, W. Farah, C. Flynn, C. W. James, M. Kerr, E. Lenc, E. K. Mahony, J. O'Meara, S. Osłowski, H. Qiu, T. Treu, V. U, T. J. Bateman, D. C. J. Bock, R. J. Bolton, A. Brown, J. D. Bunton, A. P. Chippendale, F. R. Cooray, T. Cornwell, N. Gupta, D. B. Hayman, M. Kesteven, B. S. Koribalski, A. MacLeod, N. M. McClure-Griffiths, S. Neuhold, R. P. Norris, M. A. Pilawa, R. Y. Qiao, J. Reynolds, D. N. Roxby, T. W. Shimwell, M. A. Voronkov, and C. D. Wilson. A single fast radio burst localized to a massive galaxy at cosmological distance. *Science*, 365(6453):565, 08 2019.

[10] Keith Bannister et al. The detection of an extremely bright fast radio burst in a phased array feed survey. *Astrophys. J.*, 841:L12, 2017.

[11] C. G. Bassa, S. P. Tendulkar, E. A. K. Adams, N. Maddox, S. Bogdanov, G. C. Bower, S. Burke-Spolaor, B. J. Butler, S. Chatterjee, J. M. Cordes, J. W. T. Hessels, V. M. Kaspi, C. J. Law, B. Marcote, Z. Paragi, S. M. Ransom, P. Scholz, L. G. Spitler, and H. J. van Langevelde. FRB 121102 Is Coincident with a Star-forming Region in Its Host Galaxy. *The Astrophysical Journal*, 843(1):L8, 2017.

[12] Nicholas Battaglia. The Tau of Galaxy Clusters. *JCAP*, 08:058, 2016.

[13] Adam J. Batten, Alan R. Duffy, Nastasha Wijers, Vivek Gupta, Chris Flynn, Joop Schaye, and Emma Ryan-Weber. The Cosmic Dispersion Measure in the EAGLE Simulations. 11 2020.

[14] Shivani Bhandari, Keith W. Bannister, Emil Lenc, Hyerin Cho, Ron Ekers, Cherie K. Day, Adam T. Deller, Chris Flynn, Clancy W. James, Jean-Pierre Macquart, Elizabeth K. Mahony, Lachlan Marnoch, Vanessa A. Moss, Chris Phillips, J. Xavier Prochaska, Hao Qiu, Stuart D. Ryder, Ryan M. Shannon, Nicolas Tejos, and O. Ivy Wong. Limits on Precursor and Afterglow Radio Emission from a Fast Radio Burst in a Star-forming Galaxy. *The Astrophysical Journal*, 901:L20, 10 2020.

[15] Shivani Bhandari, Elaine M. Sadler, J. Xavier Prochaska, Sunil Simha, Stuart D. Ryder, Lachlan Marnoch, Keith W. Bannister, Jean-Pierre Macquart, Chris Flynn, Ryan M. Shannon, Nicolas Tejos, Felipe Corro-Guerra, Cherie K. Day, Adam T. Deller, Ron Ekers, Sebastian Lopez, Elizabeth K. Mahony, Consuelo Nuñez, and Chris Phillips. The Host Galaxies and Progenitors of Fast Radio Bursts Localized

with the Australian Square Kilometre Array Pathfinder. *The Astrophysical Journal*, 895(2):L37, 2020.

[16] M. Bhardwaj, B. M. Gaensler, V. M. Kaspi, T. L. Landecker, R. Mckinven, D. Michilli, Z. Pleunis, S. P. Tendulkar, B. C. Andersen, P. J. Boyle, T. Cassanelli, P. Chawla, A. Cook, M. Dobbs, E. Fonseca, J. Kaczmarek, C. Leung, K. Masui, M. Mnchmeyer, C. Ng, M. Rafiei-Ravandi, P. Scholz, K. Shin, K. M. Smith, I. H. Stairs, and A. V. Zwaniga. A Nearby Repeating Fast Radio Burst in the Direction of M81. *The Astrophysical Journal*, 910:L18, 04 2021.

[17] Maciej Bilicki, Thomas H. Jarrett, John A. Peacock, Michelle E. Cluver, and Louise Steward. TWO MICRON ALL SKY SURVEY PHOTOMETRIC REDSHIFT CATALOG: A COMPREHENSIVE THREE-DIMENSIONAL CENSUS OF THE WHOLE SKY. *The Astrophysical Journal Supplement Series*, 210(1):9, 2013.

[18] Maciej Bilicki, John A. Peacock, Thomas H. Jarrett, Michelle E. Cluver, Natasha Maddox, Michael J. I. Brown, Edward N. Taylor, Nigel C. Hambly, Aleksandra Solarz, Benne W. Holwerda, Ivan Baldry, Jon Loveday, Amanda Moffett, Andrew M. Hopkins, Simon P. Driver, Mehmet Alpaslan, and Joss Bland-Hawthorn. WISE×SuperCOSMOS PHOTOMETRIC REDSHIFT CATALOG: 20 MILLION GALAXIES OVER 3π STERADIANS. *The Astrophysical Journal Supplement Series*, 225(1):5, 2016.

[19] C. D. Bochenek, V. Ravi, K. V. Belov, G. Hallinan, J. Kocz, S. R. Kulkarni, and D. L. McKenna. A fast radio burst associated with a Galactic magnetar. *Nature*, 587:59–62, 11 2020.

[20] M. Caleb, C. Flynn, and B. W. Stappers. Constraining the era of helium reionization using fast radio bursts. *Monthly Notices of the Royal Astronomical Society*, 485(2):2281–2286, 3/30/2021 2019.

[21] S. Chatterjee, C. J. Law, R. S. Wharton, S. Burke-Spolaor, J. W. T. Hessels, G. C. Bower, J. M. Cordes, S. P. Tendulkar, C. G. Bassa, P. Demorest, B. J. Butler, A. Seymour, P. Scholz, M. W. Abruzzo, S. Bogdanov, V. M. Kaspi, A. Keimpema, T. J. W. Lazio, B. Marcote, M. A. McLaughlin, Z. Paragi, S. M. Ransom, M. Rupen, L. G. Spitler, and H. J. van Langevelde. A direct localization of a fast radio burst and its host. *Nature*, 541(7635):58–61, 2017.

[22] CHIME/FRB Collaboration. First detection of fast radio bursts between 400 and 800 MHz by CHIME/FRB. *The Astronomer's Telegram*, 11901:1, 08 2018.

[23] CHIME/FRB Collaboration. The CHIME Fast Radio Burst Project: System Overview. *The Astrophysical Journal*, 863:48, 08 2018.

[24] CHIME/FRB Collaboration. A Second Source of Repeating Fast Radio Bursts. *Nature*, 566(7743):235–238, 2019.

[25] CHIME/FRB Collaboration. CHIME/FRB Discovery of Eight New Repeating Fast Radio Burst Sources. *The Astrophysical Journal*, 885:L24, 11 2019.

[26] CHIME/FRB Collaboration. Observations of Fast Radio Bursts at Frequencies down to 400 Megahertz. *Nature*, 566(7743):230–234, 2019.

[27] CHIME/FRB Collaboration. A bright millisecond-duration radio burst from a Galactic magnetar. *Nature*, 587(7832):54–58, 2020.

[28] CHIME/FRB Collaboration. Periodic activity from a fast radio burst source. *Nature*, 582:351–355, 06 2020.

[29] CHIME/FRB Collaboration. The First CHIME/FRB Fast Radio Burst Catalog. *arXiv e-prints*, page arXiv:2106.04352, 06 2021.

[30] Jay S. Chittidi, Sunil Simha, Alexandra Mannings, J. Xavier Prochaska, Marc Rafelski, Marcel Neeleman, Jean-Pierre Macquart, Nicolas Tejos, Regina A. Jorgenson, Stuart D. Ryder, Cherie K. Day, Lachlan Marnoch, Shivani Bhandari, Adam T. Deller, Hao Qiu, Keith W. Bannister, Ryan M. Shannon, and Kasper E. Heintz. Dissecting the Local Environment of FRB 190608 in the Spiral Arm of its Host Galaxy. *arXiv e-prints*, page arXiv:2005.13158, 05 2020.

[31] Katie M. Chynoweth, Glen I. Langston, Min S. Yun, Felix J. Lockman, K. H. R. Rubin, and Sarah A. Scoles. NEUTRAL HYDROGEN CLOUDS IN THE M81/M82 GROUP. 135(6):1983–1992, 2008.

[32] James J. Condon and Scott M. Ransom. *Essential Radio Astronomy*. Princeton University Press, 01 2016.

[33] Asantha Cooray and Ravi K. Sheth. Halo models of large scale structure. *Phys. Rept.*, 372:1–129, 2002.

[34] J. M. Cordes and T. J. W. Lazio. NE2001.I. A New Model for the Galactic Distribution of Free Electrons and its Fluctuations. *arXiv e-prints*, pages astro–ph/0207156, 07 2002.

[35] James M. Cordes and Shami Chatterjee. Fast Radio Bursts: An Extragalactic Enigma. *Annual Review of Astronomy and Astrophysics*, 57:417–465, 08 2019.

[36] Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R. Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, Martin Landriau, Michael Levi, Ian McGreer, Aaron Meisner, Adam D. Myers, John Moustakas, Peter Nugent, Anna Patej, Edward F. Schlafly, Alistair R. Walker, Francisco Valdes, Benjamin A. Weaver, Christophe Yèche, Hu Zou, Xu Zhou, Behzad Abareshi, T. M. C. Abbott, Bela Abolfathi, C. Aguilera, Shadab Alam, Lori Allen, A. Alvarez, James Annis, Behzad Ansarinejad, Marie Aubert, Jacqueline Beechert, Eric F. Bell, Segev Y. BenZvi, Florian Beutler, Richard M. Bielby, Adam S. Bolton, César Briceño, Elizabeth J. Buckley-Geer, Karen Butler, Annalisa Calamida, Raymond G. Carlberg, Paul Carter, Ricard Casas, Francisco J. Castander, Yumi Choi, Johan Comparat, Elena Cukanovaite, Timothée Delubac, Kaitlin DeVries, Sharmila Dey, Govinda Dhungana, Mark Dickinson, Zhejie Ding, John B. Donaldson, Yutong Duan, Christopher J. Duckworth, Sarah Eftekharzadeh, Daniel J. Eisenstein, Thomas Etourneau, Parker A. Fagrelius, Jay Farihi, Mike Fitzpatrick, Andreu Font-Ribera, Leah Fulmer, Boris T. Gänsicke, Enrique Gaztanaga, Koshy George, David W. Gerdes, Satya Gontcho A Gontcho, Claudio Gorgoni, Gregory Green, Julien Guy, Diane Harmer, M. Hernandez, Klaus Honscheid, Lijuan (Wendy) Huang, David J. James, Buell T. Jannuzi, Linhua Jiang, Richard Joyce, Armin Karcher, Sonia Karkar, Robert Kehoe, Kneib Jean-Paul, Andrea Kueter-Young, Ting-Wen Lan, Tod R. Lauer, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Michael Lesser, Laurence Perreault Levasseur, Ting S. Li, Justin L. Mann, Robert Marshall, C. E. Martínez-Vázquez, Paul Martini, Hélion du Mas des Bourboux, Sean McManus, Tobias Gabriel Meier, Brice Ménard, Nigel Metcalfe, Andrea Muñoz-Gutiérrez, Joan Najita, Kevin Napier, Gautham Narayan, Jeffrey A. Newman, Jundan Nie, Brian Nord, Dara J. Norman, Knut A. G. Olsen, Anthony Paat, Nathalie Palanque-Delabrouille, Xiyan Peng, Claire L. Poppett, Megan R. Poremba, Abhishek Prakash, David Rabinowitz, Anand Raichoor, Mehdi Rezaie, A. N. Robertson, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Gregory Rudnick, Sasha Safonova, Abhijit Saha, F. Javier Sánchez, Elodie Savary, Heidi Schweiker, Adam Scott, Hee-Jong Seo, Huanyuan Shan, David R. Silva, Zachary Slepian, Christian Soto, David Sprayberry, Ryan Staten, Coley M. Stillman, Robert J. Stupak, David L. Summers, Suk Sien Tie, H. Tirado, Mariana Vargas-Magaña, A. Katherina Vivas, Risa H. Wechsler, Doug Williams, Jinyi Yang, Qian Yang, Tolga Yapici, Dennis Zaritsky, A. Zenteno, Kai Zhang, Tianmeng Zhang, Rongpu Zhou, and Zhimin Zhou. Overview of the DESI Legacy Imaging Surveys. *The Astronomical Journal*, 157(5):168, apr 2019.

[37] Alan R. Duffy, Joop Schaye, Scott T. Kay, and Claudio Dalla Vecchia. Dark matter halo concentrations in the Wilkinson Microwave Anisotropy Probe year 5 cosmology. *Mon. Not. Roy. Astron. Soc.*, 390:L64, 2008.

[38] Aaron A. Dutton and Andrea V. Macciò. Cold dark matter haloes in the Planck era: evolution of structural parameters for Einasto and NFW profiles. *Mon. Not. Roy. Astron. Soc.*, 441(4):3359–3374, 2014.

[39] T. Eftekhari and E. Berger. Associating Fast Radio Bursts with Their Host Galaxies. *The Astrophysical Journal*, 849:162, 11 2017.

[40] Vincent R. Eke, Julio F. Navarro, and Carlos S. Frenk. The Evolution of x-ray clusters in low density universes. *Astrophys. J.*, 503:569, 1998.

[41] E. Fonseca, B. C. Andersen, M. Bhardwaj, P. Chawla, D. C. Good, A. Josephy, V. M. Kaspi, K. W. Masui, R. Mckinven, D. Michilli, Z. Pleunis, K. Shin, S. P. Tendulkar, K. M. Bandura, P. J. Boyle, C. Brar, T. Cassanelli, D. Cubranic, M. Dobbs, F. Q. Dong, B. M. Gaensler, G. Hinshaw, T. L. Landecker, C. Leung, D. Z. Li, H. H. Lin, J. Mena-Parra, M. Merryfield, A. Naidu, C. Ng, C. Patel, U. Pen, M. Rafiei-Ravandi, M. Rahman, S. M. Ransom, P. Scholz, K. M. Smith, I. H. Stairs, K. Vanderlinde, P. Yadav, and A. V. Zwaniga. Nine New Repeating Fast Radio Burst Sources from CHIME/FRB. 891(1):L6, 2020.

[42] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622(2):759–771, 2005.

[43] Kasper E. Heintz, J. Xavier Prochaska, Sunil Simha, Emma Platts, Wen-fai Fong, Nicolas Tejos, Stuart D. Ryder, Kshitij Aggerwal, Shivani Bhandari, Cherie K. Day, Adam T. Deller, Charles D. Kilpatrick, Casey J. Law, Jean-Pierre Macquart, Alexandra Mannings, Lachlan J. Marnoch, Elaine M. Sadler, and Ryan M. Shannon. Host Galaxy Properties and Offset Distributions of Fast Radio Bursts: Implications for Their Progenitors. *The Astrophysical Journal*, 903:152, 11 2020.

[44] J. L. Hodges. The significance probability of the smirnov two-sample test. *Ark. Mat.*, 3(5):469–486, 1958.

[45] Lam Hui, Enrique Gaztanaga, and Marilena LoVerde. Anisotropic Magnification Distortion of the 3D Galaxy Correlation. 1. Real Space. *Phys. Rev.*, D76:103502, 2007.

[46] C. W. James, J. X. Prochaska, J. P. Macquart, F. North-Hickey, K. W. Bannister, and A. Dunning. The fast radio burst population evolves with the star-formation rate. *arXiv e-prints*, page arXiv:2101.07998, 01 2021.

[47] C. W. James, J. X. Prochaska, J. P. Macquart, F. North-Hickey, K. W. Bannister, and A. Dunning. The z–DM distribution of fast radio bursts. *arXiv e-prints*, page arXiv:2101.08005, 01 2021.

[48] K. G. Jansky. Directional Studies of Atmospherics at High Frequencies. *Proceedings of the Institute of Radio Engineers*, 20(12):1920–1932, 1932.

[49] Karl G. Jansky. Radio Waves from Outside the Solar System. *Nature*, 132(3323):66–66, 1933.

[50] A. Josephy, P. Chawla, A. P. Curtin, V. M. Kaspi, M. Bhardwaj, P. J. Boyle, C. Brar, T. Cassanelli, E. Fonseca, B. M. Gaensler, C. Leung, H. H. Lin, K. W. Masui, R. McKinven, J. Mena-Parra, D. Michilli, C. Ng, Z. Pleunis, M. Rafiei-Ravandi, M. Rahman, P. Sanghavi, P. Scholz, K. M. Smith, I. H. Stairs, S. P. Tendulkar, and A. V. Zwaniga. No Evidence for Galactic Latitude Dependence of the Fast Radio Burst Sky Distribution. *arXiv e-prints*, page arXiv:2106.04353, 06 2021.

[51] A. Josephy, P. Chawla, E. Fonseca, C. Ng, C. Patel, Z. Pleunis, P. Scholz, B. C. Andersen, K. Bandura, M. Bhardwaj, M. M. Boyce, P. J. Boyle, C. Brar, D. Cubranic, M. Dobbs, B. M. Gaensler, A. Gill, U. Giri, D. C. Good, M. Halpern, G. Hinshaw, V. M. Kaspi, T. L. Landecker, D. A. Lang, H. H. Lin, K. W. Masui, R. Mckinven, J. Mena-Parra, M. Merryfield, D. Michilli, N. Milutinovic, A. Naidu, U. Pen, M. Rafiei-Ravandi, M. Rahman, S. M. Ransom, A. Renard, S. R. Siegel, K. M. Smith, I. H. Stairs, S. P. Tendulkar, K. Vanderlinde, P. Yadav, and A. V. Zwaniga. CHIME/FRB Detection of the Original Repeating Fast Radio Burst Source FRB 121102. 882(2):L18, 2019.

[52] Nick Kaiser. Weak Gravitational Lensing of Distant Galaxies. *Astrophys. J.*, 388:272, April 1992.

[53] Nick Kaiser. Weak Lensing and Cosmology. *Astrophys. J.*, 498(1):26–42, May 1998.

[54] J. I. Katz. Fast radio bursts — A brief review: Some questions, fewer answers. *Mod. Phys. Lett.*, A31(14):1630013, 2016.

[55] Laura C. Keating and Ue-Li Pen. Exploring the dispersion measure of the Milky Way halo. *Monthly Notices of the Royal Astronomical Society: Letters*, 496(1):L106–L110, 3/2/2021 2020.

[56] F. Kirsten, B. Marcote, K. Nimmo, J. W. T. Hessels, M. Bhardwaj, S. P. Tendulkar, A. Keimpema, J. Yang, M. P. Snelders, P. Scholz, A. B. Pearlman, C. J. Law, W. M. Peters, M. Giroletti, D. M. Hewitt, U. Bach, V. Bezukovs, M. Burgay, S. T. Buttaccio, J. E. Conway, A. Corongiu, R. Feiler, O. Forssén, M. P. Gawroński, R. Karuppusamy, M. A. Kharinov, M. Lindqvist, G. Maccaferri, A. Melnikov, O. S. Ould-Boukattine, Z. Paragi, A. Possenti, G. Surcis, N. Wang, J. Yuan, K. Aggarwal, R. Anna-Thomas, G. C. Bower, R. Blaauw, S. Burke-Spolaor, T. Cassanelli, T. E. Clarke, E. Fonseca, B. M. Gaensler, A. Gopinath, V. M. Kaspi, N. Kassim, T. J. W. Lazio, C. Leung, D. Li, H. H. Lin, K. W. Masui, R. Mckinven, D. Michilli, A. Mikhailov, C. Ng, A. Orbidans, U. L. Pen, E. Petroff, M. Rahman, S. M. Ransom, K. Shin, K. M. Smith, I. H. Stairs, and W. Vlemmings. A repeating fast radio burst source in a globular cluster. *arXiv e-prints*, page arXiv:2105.11445, 05 2021.

[57] Mitsuru Kokubo, Kazuma Mitsuda, Hajime Sugai, Shinobu Ozaki, Yosuke Minowa, Takashi Hattori, Yutaka Hayano, Kazuya Matsubayashi, Atsushi Shimono, Shigeyuki Sako, and Mamoru Doi. Hα Intensity Map of the Repeating Fast Radio Burst FRB 121102 Host Galaxy from Subaru/Kyoto 3DII AO-assisted Optical Integral-field Spectroscopy. *The Astrophysical Journal*, 844:95, 08 2017.

[58] Ely D. Kovetz, Alvise Raccanelli, and Mubdi Rahman. Cosmological Constraints with Clustering-Based Redshifts. *Mon. Not. Roy. Astron. Soc.*, 468(3):3650–3656, 2017.

[59] T. Krakowski, K. Małek, M. Bilicki, A. Pollo, A. Kurcz, and M. Krupa. Machine-learning identification of galaxies in the WISE×SuperCOSMOS all-sky catalogue. *A&A*, 596, 2016.

[60] Casey J. Law, Bryan J. Butler, J. Xavier Prochaska, Barak Zackay, Sarah Burke-Spolaor, Alexandra Mannings, Nicolas Tejos, Alexander Josephy, Bridget Andersen, Pragya Chawla, Kasper E. Heintz, Kshitij Aggarwal, Geoffrey C. Bower, Paul B. Demorest, Charles D. Kilpatrick, T. Joseph W. Lazio, Justin Linford, Ryan Mckinven, Shriharsh Tendulkar, and Sunil Simha. A Distant Fast Radio Burst Associated with Its Host Galaxy by the Very Large Array. *The Astrophysical Journal*, 899:161, 08 2020.

[61] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, 538:473–476, 2000.

[62] C. K. Li, L. Lin, S. L. Xiong, M. Y. Ge, X. B. Li, T. P. Li, F. J. Lu, S. N. Zhang, Y. L. Tuo, Y. Nang, B. Zhang, S. Xiao, Y. Chen, L. M. Song, Y. P. Xu, C. Z. Liu, S. M. Jia, X. L. Cao, J. L. Qu, S. Zhang, Y. D. Gu, J. Y. Liao, X. F. Zhao, Y. Tan, J. Y. Nie, H. S. Zhao, S. J. Zheng, Y. G. Zheng, Q. Luo, C. Cai, B. Li, W. C. Xue, Q. C. Bu, Z. Chang, G. Chen, L. Chen, T. X. Chen, Y. B. Chen, Y. P. Chen, W. Cui, W. W. Cui, J. K. Deng, Y. W. Dong, Y. Y. Du, M. X. Fu, G. H. Gao, H. Gao, M. Gao, Y. D. Gu, J. Guan, C. C. Guo, D. W. Han, Y. Huang, J. Huo, L. H. Jiang, W. C. Jiang, J. Jin, Y. J. Jin, L. D. Kong, G. Li, M. S. Li, W. Li, X. Li, X. F. Li, Y. G. Li, Z. W. Li, X. H. Liang, B. S. Liu, G. Q. Liu, H. W. Liu, X. J. Liu, Y. N. Liu, B. Lu, X. F. Lu, T. Luo, X. Ma, B. Meng, G. Ou, N. Sai, R. C. Shang, X. Y. Song, L. Sun, L. Tao, C. Wang, G. F. Wang, J. Wang, W. S. Wang, Y. S. Wang, X. Y. Wen, B. B. Wu, B. Y. Wu, M. Wu, G. C. Xiao, H. Xu, J. W. Yang, S. Yang, Y. J. Yang, Yi-Jung Yang, Q. B. Yi, Q. Q. Yin, Y. You, A. M. Zhang, C. M. Zhang, F. Zhang, H. M. Zhang, J. Zhang, T. Zhang, W. Zhang, W. C. Zhang, W. Z. Zhang, Y. Zhang, Yue Zhang, Y. F. Zhang, Y. J. Zhang, Z. Zhang, Zhi Zhang, Z. L. Zhang, D. K. Zhou, J. F. Zhou, Y. Zhu, Y. X. Zhu, and R. L. Zhuang. HXMT identification of a non-thermal X-ray burst from SGR J1935+2154 and with FRB 200428. *Nature Astronomy*, 02 2021.

[63] Dongzi Li, Almog Yalinewich, and Patrick C. Breysse. Statistical inference of the distance to ASKAP FRBs. 2 2019.

[64] D. Nelson Limber. The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *Astrophys. J.*, 117:134, January 1953.

[65] Eric V. Linder. Detecting helium reionization with fast radio bursts. *Physical Review D*, 101(10):103019–, 05 2020.

[66] D. R. Lorimer, M. Bailes, M. A. McLaughlin, D. J. Narkevic, and F. Crawford. A Bright Millisecond Radio Burst of Extragalactic Origin. *Science*, 318:777, 11 2007.

[67] Marilena Loverde and Niayesh Afshordi. Extended Limber approximation. *Phys. Rev.*, 78(12):123506, December 2008.

[68] J. P. Macquart, J. X. Prochaska, M. McQuinn, K. W. Bannister, S. Bhandari, C. K. Day, A. T. Deller, R. D. Ekers, C. W. James, L. Marnoch, S. Osłowski, C. Phillips,

S. D. Ryder, D. R. Scott, R. M. Shannon, and N. Tejos. A census of baryons in the Universe from localized fast radio bursts. *Nature*, 581:391–395, 05 2020.

[69] Mathew S. Madhavacheril, Nicholas Battaglia, Kendrick M. Smith, and Jonathan L. Sievers. Cosmology with the kinematic Sunyaev-Zeldovich effect: Breaking the optical depth degeneracy with fast radio bursts. *Physical Review D*, 100(10):103532–, 11 2019.

[70] Alexandra G. Mannings, Wen-fai Fong, Sunil Simha, J. Xavier Prochaska, Marc Rafelski, Charles D. Kilpatrick, Nicolas Tejos, Kasper E. Heintz, Shivani Bhandari, Cherie K. Day, Adam T. Deller, Stuart D. Ryder, Ryan M. Shannon, and Shriharsh P. Tendulkar. A High-Resolution View of Fast Radio Burst Host Environments. *arXiv e-prints*, page arXiv:2012.11617, 12 2020.

[71] B. Marcote et al. The Repeating Fast Radio Burst FRB 121102 as Seen on Milliarcsecond Angular Scales. *Astrophys. J.*, 834(2):L8, 2017.

[72] B. Marcote, K. Nimmo, J. W. T. Hessels, S. P. Tendulkar, C. G. Bassa, Z. Paragi, A. Keimpema, M. Bhardwaj, R. Karuppusamy, V. M. Kaspi, C. J. Law, D. Michilli, K. Aggarwal, B. Andersen, A. M. Archibald, K. Bandura, G. C. Bower, P. J. Boyle, C. Brar, S. Burke-Spolaor, B. J. Butler, T. Cassanelli, P. Chawla, P. Demorest, M. Dobbs, E. Fonseca, U. Giri, D. C. Good, K. Gourdji, A. Josephy, A. Yu. Kirichenko, F. Kirsten, T. L. Landecker, D. Lang, T. J. W. Lazio, D. Z. Li, H. H. Lin, J. D. Linford, K. Masui, J. Mena-Parra, A. Naidu, C. Ng, C. Patel, U. L. Pen, Z. Pleunis, M. Rafiei-Ravandi, M. Rahman, A. Renard, P. Scholz, S. R. Siegel, K. M. Smith, I. H. Stairs, K. Vanderlinde, and A. V. Zwaniga. A repeating fast radio burst source localized to a nearby spiral galaxy. *Nature*, 577:190–194, 01 2020.

[73] Kiyoshi W. Masui, J. Richard Shaw, Cherry Ng, Kendrick M. Smith, Keith Vanderlinde, and Adiv Paradise. Algorithms for fft beamforming radio interferometers. *arXiv e-prints*, page arXiv:1710.08591, 10 2017.

[74] Kiyoshi Wesley Masui and Kris Sigurdson. Dispersion Distance and the Matter Distribution of the Universe in Dispersion Space. *Physical Review Letters*, 115:121301, 09 2015.

[75] Matthew McQuinn. Locating the "missing" baryons with extragalactic dispersion measure estimates. *The Astrophysical Journal Letters*, 780(2):L33, 2014.

[76] Matthew McQuinn and Martin White. On using angular cross-correlations to determine source redshift distributions. *Mon. Not. Roy. Astron. Soc.*, 433:2857–2883, 2013.

[77] Brice Ménard, Ryan Scranton, Samuel Schmidt, Chris Morrison, Donghui Jeong, Tamas Budavari, and Mubdi Rahman. Clustering-based redshift estimation: method and application to data. 2013.

[78] S. Mereghetti, V. Savchenko, C. Ferrigno, D. Götz, M. Rigoselli, A. Tiengo, A. Bazzano, E. Bozzo, A. Coleiro, T. J. L. Courvoisier, M. Doyle, A. Goldwurm, L. Hanlon, E. Jourdain, A. von Kienlin, A. Lutovinov, A. Martin-Carrillo, S. Molkov, L. Natalucci, F. Onori, F. Panessa, J. Rodi, J. Rodriguez, C. Sánchez-Fernández, R. Sunyaev, and P. Ubertini. INTEGRAL Discovery of a Burst with Associated Radio Emission from the Magnetar SGR 1935+2154. *The Astrophysical Journal*, 898:L29, 08 2020.

[79] D. Michilli, K. W. Masui, R. Mckinven, D. Cubranic, M. Bruneault, C. Brar, C. Patel, P. J. Boyle, I. H. Stairs, A. Renard, K. Bandura, S. Berger, D. Breitman, T. Cassanelli, M. Dobbs, V. M. Kaspi, C. Leung, J. Mena-Parra, Z. Pleunis, L. Russell, P. Scholz, S. R. Siegel, S. P. Tendulkar, and K. Vanderlinde. An Analysis Pipeline for CHIME/FRB Full-array Baseband Data. *The Astrophysical Journal*, 910:147, 04 2021.

[80] D. Michilli, A. Seymour, J. W. T. Hessels, L. G. Spitler, V. Gajjar, A. M. Archibald, G. C. Bower, S. Chatterjee, J. M. Cordes, K. Gourdji, G. H. Heald, V. M. Kaspi, C. J. Law, C. Sobey, E. A. K. Adams, C. G. Bassa, S. Bogdanov, C. Brinkman, P. Demorest, F. Fernandez, G. Hellbourg, T. J. W. Lazio, R. S. Lynch, N. Maddox, B. Marcote, M. A. McLaughlin, Z. Paragi, S. M. Ransom, P. Scholz, A. P. V. Siemion, S. P. Tendulkar, P. Van Rooy, R. S. Wharton, and D. Whitlow. An extreme magneto-ionic environment associated with the fast radio burst source FRB 121102. *Nature*, 553(7687):182–185, 2018.

[81] Julian B. Muñoz and Abraham Loeb. Finding the Missing Baryons with Fast Radio Bursts and Sunyaev-Zeldovich Maps. *Phys. Rev.*, D98(10):103518, 2018.

[82] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. A Universal density profile from hierarchical clustering. *Astrophys. J.*, 490:493–508, 1997.

[83] L. B. Newburgh et al. HIRAX: A Probe of Dark Energy and Radio Transients. *Proc. SPIE Int. Soc. Opt. Eng.*, 9906:99065X, 2016.

[84] Chen-Hui Niu, Di Li, Rui Luo, Wei-Yang Wang, Jumei Yao, Bing Zhang, Wei-Wei Zhu, Pei Wang, Haoyang Ye, Yong-Kun Zhang, Jia-rui Niu, Ning-yu Tang, Ran Duan, Marko Krco, Shi Dai, Yi Feng, Chenchen Miao, Zhichen Pan, Lei Qian, Mengyao Xue, Mao Yuan, Youling Yue, Lei Zhang, and Xinxin Zhang. CRAFTS for Fast Radio Bursts Extending the dispersion-fluence relation with new FRBs detected by FAST. *arXiv e-prints*, page arXiv:2102.10546, 02 2021.

[85] Samuel Passaglia, Alessandro Manzotti, and Scott Dodelson. Cross-correlating 2D and 3D galaxy surveys. *Phys. Rev.*, D95(12):123508, 2017.

[86] E. Petroff, E. D. Barr, A. Jameson, E. F. Keane, M. Bailes, M. Kramer, V. Morello, D. Tabbara, and W. van Straten. FRBCAT: The Fast Radio Burst Catalogue. *Publ. Astron. Soc. Austral.*, 33:e045, 2016.

[87] E. Petroff, J. W. T. Hessels, and D. R. Lorimer. Fast radio bursts. *Astronomy and Astrophysics Review*, 27:4, 05 2019.

[88] E. Petroff, E. F. Keane, E. D. Barr, J. E. Reynolds, J. Sarkissian, P. G. Edwards, J. Stevens, C. Brem, A. Jameson, S. Burke-Spolaor, S. Johnston, N. D. R. Bhat, P. Chandra S. Kudale, and S. Bhandari. Identifying the source of perytons at the Parkes radio telescope. *Monthly Notices of the Royal Astronomical Society*, 451:3933–3940, 08 2015.

[89] E. Petroff and O. Yaron. Fast Radio Burst Catalogue on the TNS. *Transient Name Server AstroNote*, 160:1, August 2020.

[90] E. Platts, A. Weltman, A. Walters, S. P. Tendulkar, J. E. B. Gordin, and S. Kandhai. A living theory catalogue for fast radio bursts. *Physics Reports*, 821:1–27, 08 2019.

[91] Ziggy Pleunis, Deborah C. Good, Victoria M. Kaspi, Ryan Mckinven, Scott M. Ransom, Paul Scholz, Kevin Bandura, Mohit Bhardwaj, P. J. Boyle, Charanjot Brar, Tomas Cassanelli, Pragya Chawla, Fengqiu, Dong, Emmanuel Fonseca, B. M. Gaensler, Alexander Josephy, Jane F. Kaczmarek, Calvin Leung, Hsiu-Hsien Lin, Kiyoshi W. Masui, Juan Mena-Parra, Daniele Michilli, Cherry Ng, Chitrang Patel, Masoud Rafiei-Ravandi, Mubdi Rahman, Pranav Sanghavi, Kaitlyn Shin, Kendrick M. Smith, Ingrid H. Stairs, and Shriharsh P. Tendulkar. Fast Radio Burst Morphology in the First CHIME/FRB Catalog. *arXiv e-prints*, page arXiv:2106.04356, 06 2021.

[92] J. Xavier Prochaska, Jean-Pierre Macquart, Matthew McQuinn, Sunil Simha, Ryan M. Shannon, Cherie K. Day, Lachlan Marnoch, Stuart Ryder, Adam Deller,

Keith W. Bannister, Shivani Bhandari, Rongmon Bordoloi, John Bunton, Hyerin Cho, Chris Flynn, Elizabeth K. Mahony, Chris Phillips, Hao Qiu, and Nicolas Tejos. The low density and magnetization of a massive galaxy halo exposed by a fast radio burst. *Science*, 366(6462):231, 10 2019.

[93] J Xavier Prochaska and Yong Zheng. Probing Galactic haloes with fast radio bursts. *Monthly Notices of the Royal Astronomical Society*, 485(1):648–665, 3/2/2021 2019.

[94] Masoud Rafiei-Ravandi, Kendrick M. Smith, Dongzi Li, Kiyoshi W. Masui, Alexander Josephy, Matt Dobbs, Dustin Lang, Mohit Bhardwaj, Chitrang Patel, Kevin Bandura, Sabrina Berger, P. J. Boyle, Charanjot Brar, Tomas Cassanelli, Pragya Chawla, Fengqiu Adam Dong, Emmanuel Fonseca, B. M. Gaensler, Utkarsh Giri, Deborah C. Good, Mark Halpern, Jane Kaczmarek, Victoria M. Kaspi, Calvin Leung, Hsiu-Hsien Lin, Juan Mena-Parra, B. W. Meyers, D. Michilli, Moritz Münchmeyer, Cherry Ng, Emily Petroff, Ziggy Pleunis, Mubdi Rahman, Pranav Sanghavi, Paul Scholz, Kaitlyn Shin, Ingrid H. Stairs, Shriharsh P. Tendulkar, Keith Vanderlinde, and Andrew Zwaniga. CHIME/FRB Catalog 1 results: statistical cross-correlations with large-scale structure. *arXiv e-prints*, page arXiv:2106.04354, 06 2021.

[95] Masoud Rafiei-Ravandi, Kendrick M. Smith, and Kiyoshi W. Masui. Characterizing fast radio bursts through statistical cross-correlations. *Physical Review D*, 102(2):023528–, 07 2020.

[96] Mubdi Rahman, Brice Ménard, Ryan Scranton, Samuel J. Schmidt, and Christopher B. Morrison. Clustering-based Redshift Estimation: Comparison to Spectroscopic Redshifts. *Mon. Not. Roy. Astron. Soc.*, 447:3500, 2015.

[97] Anand Raichoor, Daniel J. Eisenstein, Tanveer Karim, Jeffrey A. Newman, John Moustakas, David D. Brooks, Kyle S. Dawson, Arjun Dey, Yutong Duan, Sarah Eftekharzadeh, Enrique Gaztañaga, Robert Kehoe, Martin Landriau, Dustin Lang, Jae H. Lee, Michael E. Levi, Aaron M. Meisner, Adam D. Myers, Nathalie Palanque-Delabrouille, Claire Poppett, Francisco Prada, Ashley J. Ross, David J. Schlegel, Michael Schubnell, Ryan Staten, Gregory Tarlé, Rita Tojeiro, Christophe Yèche, and Rongpu Zhou. Preliminary Target Selection for the DESI Emission Line Galaxy (ELG) Sample. *Research Notes of the American Astronomical Society*, 4:180, 10 2020.

[98] V. Ravi, M. Catha, L. D'Addario, S. G. Djorgovski, G. Hallinan, R. Hobbs, J. Kocz, S. R. Kulkarni, J. Shi, H. K. Vedantham, S. Weinreb, and D. P. Woody. A fast radio burst localized to a massive galaxy. *Nature*, 572:352–354, 08 2019.

[99] Vikram Ravi. Measuring the circum- and inter-galactic baryon contents with fast radio bursts. *Astrophys. J.*, 872:88, 2019.

[100] Vikram Ravi and Abraham Loeb. Explaining the Statistical Properties of Fast Radio Bursts with Suppressed Low-frequency Emission. *The Astrophysical Journal*, 874:72, 03 2019.

[101] Darren Reed, Richard Bower, Carlos Frenk, Adrian Jenkins, and Tom Theuns. The halo mass function from the dark ages through the present day. *Mon. Not. Roy. Astron. Soc.*, 374:2–15, 2007.

[102] Robert Reischke, Steffen Hagstotz, and Robert Lilow. Consistent equivalence principle tests with fast radio bursts. *arXiv e-prints*, page arXiv:2102.11554, 02 2021.

[103] Robert Reischke, Steffen Hagstotz, and Robert Lilow. Probing primordial non-Gaussianity with fast radio bursts. *Physical Review D*, 103:023517, 01 2021.

[104] A. Ridnaia, D. Svinkin, D. Frederiks, A. Bykov, S. Popov, R. Aptekar, S. Golenetskii, A. Lysenko, A. Tsvetkova, M. Ulanov, and T. L. Cline. A peculiar hard X-ray counterpart of a Galactic fast radio burst. *Nature Astronomy*, 5:372–377, 01 2021.

[105] Tridib Roy and R. T. Gangadhara. Radio Emission from Pulsars due to Relativistic Plasma. *The Astrophysical Journal*, 878(2):148, 2019.

[106] Omar Ruiz-Macias, Pauline Zarrouk, Shaun Cole, Peder Norberg, Carlton Baugh, David Brooks, Arjun Dey, Yutong Duan, Sarah Eftekharzadeh, Daniel J. Eisenstein, Jaime E. Forero-Romero, Enrique Gaztañaga, ChangHoon Hahn, Robert Kehoe, Martin Landriau, Dustin Lang, Michael E. Levi, John Lucey, Aaron M. Meisner, John Moustakas, Adam D. Myers, Nathalie Palanque-Delabrouille, Claire Poppett, Francisco Prada, Anand Raichoor, David J. Schlegel, Michael Schubnell, Gregory Tarlé, David H. Weinberg, M. J. Wilson, and Christophe Yèche. Preliminary Target Selection for the DESI Bright Galaxy Survey (BGS). *Research Notes of the American Astronomical Society*, 4:187, 10 2020.

[107] Emmanuel Schaan, Simone Ferraro, Stefania Amodeo, Nicholas Battaglia, Simone Aiola, Jason E. Austermann, James A. Beall, Rachel Bean, Daniel T. Becker, Richard J. Bond, Erminia Calabrese, Victoria Calafut, Steve K. Choi, Edward V. Denison, Mark J. Devlin, Shannon M. Duff, Adriaan J. Duivenvoorden, Jo Dunkley, Rolando Dünner, Patricio A. Gallardo, Yilun Guan, Dongwon Han, J. Colin Hill, Gene C. Hilton, Matt Hilton, Renée Hložek, Johannes Hubmayr, Kevin M.

Huffenberger, John P. Hughes, Brian J. Koopman, Amanda MacInnis, Jeff McMahon, Mathew S. Madhavacheril, Kavilan Moodley, Tony Mroczkowski, Sigurd Naess, Federico Nati, Laura B. Newburgh, Michael D. Niemack, Lyman A. Page, Bruce Partridge, Maria Salatino, Neelima Sehgal, Alessandro Schillaci, Cristóbal Sifón, Kendrick M. Smith, David N. Spergel, Suzanne Staggs, Emilie R. Storer, Hy Trac, Joel N. Ullom, Jeff Van Lanen, Leila R. Vale, Alexander van Engelen, Mariana Vargas Magaña, Eve M. Vavagiakis, Edward J. Wollack, and Zhilei Xu. Atacama Cosmology Telescope: Combined kinematic and thermal Sunyaev-Zel'dovich measurements from BOSS CMASS and LOWZ halos. *Physical Review D*, 103(6):063513–, 03 2021.

[108] F. W. Scholz and M. A. Stephens. K-Sample Anderson–Darling Tests. *Journal of the American Statistical Association*, 82(399):918–924, 09 1987.

[109] P. Scholz et al. The repeating Fast Radio Burst FRB 121102: Multi-wavelength observations and additional bursts. *Astrophys. J.*, 833(2):177, 2016.

[110] R. M. Shannon, J.-P. Macquart, K. W. Bannister, R. D. Ekers, C. W. James, S. Osłowski, H. Qiu, M. Sammons, A. W. Hotan, M. A. Voronkov, R. J. Beresford, M. Brothers, A. J. Brown, J. D. Bunton, A. P. Chippendale, C. Haskins, M. Leach, M. Marquarding, D. McConnell, M. A. Pilawa, E. M. Sadler, E. R. Troup, J. Tuthill, M. T. Whiting, J. R. Allison, C. S. Anderson, M. E. Bell, J. D. Collier, G. Gürkan, G. Heald, and C. J. Riseley. The dispersion-brightness relation for fast radio bursts from a wide-field survey. *Nature*, 562:386–390, October 2018.

[111] Erin S. Sheldon, Carlos Cunha, Rachel Mandelbaum, J. Brinkmann, and Benjamin A. Weaver. Photometric Redshift Probability Distributions for Galaxies in the SDSS DR8. *Astrophys. J. Suppl.*, 201:32, 2012.

[112] Ravi K. Sheth and Giuseppe Tormen. An Excursion set model of hierarchical clustering : Ellipsoidal collapse and the moving barrier. *Mon. Not. Roy. Astron. Soc.*, 329:61, 2002.

[113] Masato Shirasaki, Kazumi Kashiyama, and Naoki Yoshida. Large-scale clustering as a probe of the origin and the host environment of fast radio bursts. *Physical Review D*, 95:083012, 04 2017.

[114] Sunil Simha, Joseph N. Burchett, J. Xavier Prochaska, Jay S. Chittidi, Oskar Elek, Nicolas Tejos, Regina Jorgenson, Keith W. Bannister, Shivani Bhandari, Cherie K. Day, Adam T. Deller, Angus G. Forbes, Jean-Pierre Macquart, Stuart D. Ryder,

and Ryan M. Shannon. Disentangling the Cosmic Web toward FRB 190608. *The Astrophysical Journal*, 901:134, 10 2020.

[115] Sukhdeep Singh, Rachel Mandelbaum, and Surhud More. Intrinsic alignments of SDSS-III BOSS LOWZ sample galaxies. *Mon. Not. Roy. Astron. Soc.*, 450(2):2195–2216, 2015.

[116] Kendrick M. Smith, Mathew S. Madhavacheril, Moritz Münchmeyer, Simone Ferraro, Utkarsh Giri, and Matthew C. Johnson. KSZ tomography and the bispectrum. 2018.

[117] M. Sokolowski et al. No low-frequency emission from extremely bright Fast Radio Bursts. *Astrophys. J.*, 867(1):L12, 2018.

[118] Apostolos Spanakis-Misirlis. Frbstats: A web-based platform for visualization of fast radio burst properties. *Astrophysics Source Code Library*, page ascl:2106.028, 06 2021.

[119] L. G. Spitler, J. M. Cordes, J. W. T. Hessels, D. R. Lorimer, M. A. McLaughlin, S. Chatterjee, F. Crawford, J. S. Deneva, V. M. Kaspi, R. S. Wharton, B. Allen, S. Bogdanov, A. Brazier, F. Camilo, P. C. C. Freire, F. A. Jenet, C. Karako-Argaman, B. Knispel, P. Lazarus, K. J. Lee, J. van Leeuwen, R. Lynch, S. M. Ransom, P. Scholz, X. Siemens, I. H. Stairs, K. Stovall, J. K. Swiggum, A. Venkataraman, W. W. Zhu, C. Aulbert, and H. Fehrmann. FAST RADIO BURST DISCOVERED IN THE ARECIBO PULSAR ALFA SURVEY. *The Astrophysical Journal*, 790(2):101, jul 2014.

[120] L. G. Spitler, P. Scholz, J. W. T. Hessels, S. Bogdanov, A. Brazier, F. Camilo, S. Chatterjee, J. M. Cordes, F. Crawford, J. Deneva, R. D. Ferdman, P. C. C. Freire, V. M. Kaspi, P. Lazarus, R. Lynch, E. C. Madsen, M. A. McLaughlin, C. Patel, S. M. Ransom, A. Seymour, I. H. Stairs, B. W. Stappers, J. van Leeuwen, and W. W. Zhu. A repeating fast radio burst. *Nature*, 531:202–205, 03 2016.

[121] Ryuichi Takahashi, Kunihito Ioka, Asuka Mori, and Koki Funahashi. Statistical modelling of the cosmological dispersion measure. *Monthly Notices of the Royal Astronomical Society*, 502:2615–2629, 04 2021.

[122] S. P. Tendulkar, C. G. Bassa, J. M. Cordes, G. C. Bower, C. J. Law, S. Chatterjee, E. A. K. Adams, S. Bogdanov, S. Burke-Spolaor, B. J. Butler, P. Demorest, J. W. T. Hessels, V. M. Kaspi, T. J. W. Lazio, N. Maddox, B. Marcote, M. A. McLaughlin, Z. Paragi, S. M. Ransom, P. Scholz, A. Seymour, L. G. Spitler, H. J. van Langevelde,

and R. S. Wharton. The Host Galaxy and Redshift of the Repeating Fast Radio Burst FRB 121102. *The Astrophysical Journal*, 834(2):L7, 2017.

[123] Keith Vanderlinde, Adrian Liu, Bryan Gaensler, Dick Bond, Gary Hinshaw, Cherry Ng, Cynthia Chiang, Ingrid Stairs, Jo-Anne Brown, Jonathan Sievers, Juan Mena, Kendrick Smith, Kevin Bandura, Kiyoshi Masui, Kristine Spekkens, Leo Belostotski, Matt Dobbs, Neil Turok, Patrick Boyle, Michael Rupen, Tom Landecker, Ue-Li Pen, and Victoria Kaspi. The canadian hydrogen observatory and radio-transient detector (chord). *Canadian Long Range Plan for Astronomy and Astrophysics White Papers*, 2020:28, 10 2019.

[124] A. Vikhlinin, A. Kravtsov, W. Forman, C. Jones, M. Markevitch, S. S. Murray, and L. Van Speybroeck. Chandra Sample of Nearby Relaxed Galaxy Clusters: Mass, Gas Fraction, and Mass-Temperature Relation. *The Astrophysical Journal*, 640(2):691–709, Apr 2006.

[125] Risa H. Wechsler and Jeremy L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. *Annual Review of Astronomy and Astrophysics*, 56(1):435–487, Sep 2018.

[126] J. M. Yao, R. N. Manchester, and N. Wang. A New Electron-density Model for Estimation of Pulsar and FRB Distances. *The Astrophysical Journal*, 835:29, 01 2017.

[127] Rachel C. Zhang, Bing Zhang, Ye Li, and Duncan R. Lorimer. On the energy and redshift distributions of fast radio bursts. *Monthly Notices of the Royal Astronomical Society*, 501:157–167, 01 2021.

[128] Z. J. Zhang, K. Yan, C. M. Li, G. Q. Zhang, and F. Y. Wang. Intergalactic Medium Dispersion Measures of Fast Radio Bursts Estimated from IllustrisTNG Simulation and Their Cosmological Applications. *The Astrophysical Journal*, 906:49, 01 2021.

[129] Rongpu Zhou, Jeffrey A. Newman, Kyle S. Dawson, Daniel J. Eisenstein, David D. Brooks, Arjun Dey, Biprateep Dey, Yutong Duan, Sarah Eftekharzadeh, Enrique Gaztañaga, Robert Kehoe, Martin Landriau, Michael E. Levi, Timothy C. Licquia, Aaron M. Meisner, John Moustakas, Adam D. Myers, Nathalie Palanque-Delabrouille, Claire Poppett, Francisco Prada, Anand Raichoor, David J. Schlegel, Michael Schubnell, Ryan Staten, Gregory Tarlé, and Christophe Yèche. Preliminary Target Selection for the DESI Luminous Red Galaxy (LRG) Sample. *Research Notes of the American Astronomical Society*, 4:181, 10 2020.

[130] Rongpu Zhou, Jeffrey A Newman, Yao-Yuan Mao, Aaron Meisner, John Moustakas, Adam D Myers, Abhishek Prakash, Andrew R Zentner, David Brooks, Yutong Duan, Martin Landriau, Michael E Levi, Francisco Prada, and Gregory Tarle. The Clustering of DESI-like Luminous Red Galaxies Using Photometric Redshifts. *Monthly Notices of the Royal Astronomical Society*, 12/9/2020 2020.

APPENDICES

Appendix A

Halo model

In this appendix, we describe the model for spatial clustering of FRBs and galaxies used throughout §2. We use a halo model approach: first we specify the clustering of dark matter halos, then specify how halos are populated by FRBs and galaxies.

A.1 Dark matter halos

We define $\sigma(R, z)$ to be the RMS amplitude of the *linear* density field at redshift z, smoothed with a tophat filter of comoving radius R:

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

where $W(x)$ is the Fourier transform of a unit-radius tophat:

$$W(x) = 3\,\frac{\sin(x) - x\cos(x)}{x^3} \tag{A.2}$$

and $P_{\rm lin}(k, z)$ is the matter power spectrum in linear perturbation theory, which we compute numerically with CAMB [61]. Throughout, we adopt a flat ΛCDM cosmology with $h = 0.67$, $\Omega_m = 0.315$, $\Omega_b = 0.048$, $A_s = 2 \times 10^{-9}$, $n_s = 0.965$, $\sum_\nu m_\nu = 0.06$ eV, and $T_{\rm CMB} = 2.726$ K.

If M is a halo mass, we define

$$R_M = \left(\frac{3M}{4\pi\rho_m} \right)^{1/3} \tag{A.3}$$

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

Let $n_h(M,z)$ be the halo mass function, i.e. the number density of halos per comoving volume per unit halo mass. We use the Sheth-Tormen mass function [112, 101], given by:

$$\begin{aligned} n_h(M) &= \frac{\rho_{m,0}}{M}\frac{d\log\sigma^{-1}}{dM}f(\sigma) \\ f(\sigma) &= A\frac{\delta_c}{\sigma}\sqrt{\frac{2a}{\pi}}\left(1+\left(\frac{\sigma^2}{a\delta_c^2}\right)^p\right)\exp\left(-\frac{a\delta_c^2}{2\sigma^2}\right) \end{aligned} \tag{A.4}$$

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

$$a = 0.707 \qquad \delta_c = 1.686 \qquad p = 0.3 \tag{A.5}$$

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

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

$$b_h(M) = 1 + \frac{1}{\delta_c}\frac{d\log f}{d\log\sigma} \tag{A.6}$$

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

We assume that halos have NFW (Navarro-Frenk-White) density profiles [82]. Recall that the NFW profile $\rho(r)$ has two parameters: the virial radius $r_{\rm vir}$ where the profile is truncated, and the scale radius r_s which appears in the functional form of the profile. Sometimes, we reparameterize by replacing one of these parameters by the concentration $c = r_{\rm vir}/r_s$. The NFW profile $u(r)$ and its Fourier transform $\tilde{u}(k)$ are given by:

$$u(r) = \frac{A}{(r/r_s)(1+r/r_s)^2} \qquad (r \le r_{\rm vir}) \tag{A.7}$$

$$\begin{aligned} \tilde{u}(k) = 4\pi A r_s^3\Bigg(&-\frac{\sin(\kappa c)}{\kappa(1+c)} \\ &+(\cos\kappa)\Big[\mathrm{Ci}(\kappa(1+c)) - \mathrm{Ci}(\kappa)\Big] \\ &+(\sin\kappa)\Big[\mathrm{Si}(\kappa(1+c)) - \mathrm{Si}(\kappa)\Big]\Bigg) \end{aligned} \tag{A.8}$$

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

$$\begin{aligned} \mathrm{Si}(x) &= \int_0^x dt\, \frac{\sin t}{t} && \text{(A.9)} \\ \mathrm{Ci}(x) &= -\int_x^\infty dt\, \frac{\cos t}{t} \\ &= \gamma + \log(x) + \int_0^x dt\, \frac{\cos t - 1}{t} && \text{(A.10)} \end{aligned}$$

and $\gamma = 0.577216\ldots$ is Euler's constant. We choose the normalizing constant A in Eqs. (A.7), (A.8) to be:

$$A = \frac{1}{4\pi r_s^3}\left(\log(1+c) - \frac{c}{1+c}\right)^{-1} \qquad \text{(A.11)}$$

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

To use the NFW profile, we need expressions for the virial radius $r_{\rm vir}(M,z)$ and halo concentration $c(M,z)$, as functions of halo mass and redshift. For the concentration, we use the fitting function from [38]:

$$\begin{aligned} \log_{10} c(M,z) &= \alpha(z) + \beta(z)\log_{10}\left(\frac{M}{10^{12}\,h^{-1}M_\odot}\right) \\ \alpha(z) &= 0.537 + 0.488\exp\left(-0.718 z^{1.08}\right) \\ \beta(z) &= -0.097 + 0.024 z \end{aligned} \qquad \text{(A.12)}$$

For the virial radius, we reparameterize by defining a virial *density*:

$$\rho_{\rm vir} = \frac{3M(1+z)^3}{4\pi r_{\rm vir}^3} \qquad \text{(A.13)}$$

then use the fitting function for $\rho_{\rm vir}$ from [40]:

$$\begin{aligned} \rho_{\rm vir}(z) &= 178\,\Omega_m(z)^{0.45}\rho_{\rm crit}(z) \\ &= 178\,\Omega_m(z)^{0.45}\left(\frac{3}{8\pi G}H(z)^2\right) \end{aligned} \qquad \text{(A.14)}$$

The factor $(1+z)^3$ in Eq. (A.13) arises because $\rho_{\rm vir}$ is a physical density, whereas $r_{\rm vir}$ is a comoving distance.

A.2 Galaxies

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

$$\bar{N}_g(M,z) = \begin{cases} (M/M_g(z)) & \text{if } M \geq M_g(z) \\ 0 & \text{if } M < M_g \end{cases} \tag{A.15}$$

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

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

$$\frac{dn_g^{2d}}{dz} = \Omega \frac{\chi(z)^2}{H(z)} \int_{M_g(z)}^{\infty} dM\, n_h(M) \frac{M}{M_g(z)} \tag{A.16}$$

for $M_g(z)$. (This procedure for reverse-engineering a threshold halo mass $M_g(z)$ from an observed redshift distribution is sometimes called "abundance matching".) The redshift distribution dn_g^{2d}/dz is taken from [111, 17, 2] for SDSS-DR8, 2MPZ, and DESI-ELG respectively. For each survey, the redshift distribution dn_g^{2d}/dz and threshold halo mass $M_g(z)$ are shown in Figures 2.1, 2.2.

A.3 FRBs

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

$$\frac{dn_f^{2d}}{dz} \propto z^2 e^{-\alpha z} \tag{A.17}$$

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

$$\alpha = \begin{cases} 3.5 & \text{(high-}z\text{ FRB model)} \\ 120 & \text{(low-}z\text{ FRB model)} \end{cases} \tag{A.18}$$

$$z_{\rm max} = \begin{cases} 5 & \text{(high-}z\text{ FRB model)} \\ 0.12 & \text{(low-}z\text{ FRB model)} \end{cases} \tag{A.19}$$

for our fiducial high-z and low-z FRB models respectively. The FRB redshift distribution in both models is shown in Figure 2.1.

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

$$\bar{N}_f(M,z) = \begin{cases} \xi(z)\,(M/M_f) & \text{if } M \geq M_f \\ 0 & \text{if } M < M_f \end{cases} \tag{A.20}$$

where M_f is the threshold halo mass for hosting an FRB, and $\xi(z)$ is an FRB event rate per threshold halo mass. In the FRB case, we take M_f to be a free parameter, and determine $\xi(z)$ by abundance-matching to the FRB redshift distribution in Eq. (A.17). In detail, we take:

$$M_f = 10^9\ h^{-1}\ M_\odot \tag{A.21}$$

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

$$\xi(z) = \frac{dn_f^{2d}}{dz}\left(\Omega\frac{\chi(z)^2}{H(z)}\int_{M_f}^{\infty} dM\, n_h(M)\frac{M}{M_f(z)}\right)^{-1} \tag{A.22}$$

Thus, our FRB redshift distribution and HOD are parameterized by $(\alpha, z_{\rm max}, M_f)$, and the total number of observed FRBs N_f which determines the proportionality constant in Eq. (A.17).

We model dispersion measures by assuming that the host DM is a lognormal random variable. That is, the probability distribution is:

$$p(\mathrm{DM_{host}}) = \frac{1}{\mathrm{DM_{host}}\sqrt{2\pi\sigma_{\log}^2}}\exp\left(-\frac{(\log \mathrm{DM_{host}} - \mu_{\log})^2}{2\sigma_{\log}^2}\right) \tag{A.23}$$

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

$$\mu_{\log} = \begin{cases} 4 & \text{(high-}z\text{ FRB model)} \\ 6.78 & \text{(low-}z\text{ FRB model)} \end{cases} \tag{A.24}$$

$$\sigma_{\log} = \begin{cases} 1 & \text{(high-}z\text{ FRB model)} \\ 0.63 & \text{(low-}z\text{ FRB model)} \end{cases} \tag{A.25}$$

The FRB DM distribution in both models is shown in Figure 2.1.

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

$$p(\theta_x, \theta_y) = \frac{4\log 2}{\pi\theta_f^2}\exp\left(-4\log 2\frac{\theta_x^2+\theta_y^2}{\theta_f^2}\right) \tag{A.26}$$

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

A.4 Power spectra

Given the model for halos, FRBs and galaxies from the previous sections, we are interested in angular power spectra of the form C_ℓ^{XY}, where each 2D field X, Y could be either a galaxy field (denoted g) or an FRB field (denoted f). We are primarily interested in cross power spectra C_ℓ^{fg}, but auto spectra (C_ℓ^{ff}, C_ℓ^{gg}) also arise when forecasting signal-to-noise (e.g. Eq. 2.46).

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

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

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

$$\bar{N}_g(M,z) = \begin{cases} \frac{M}{M_g(z)} & \text{if } M \geq M_g(z) \text{ and } z \in [z_{\rm min}, z_{\rm max}] \\ 0 & \text{otherwise.} \end{cases} \tag{A.27}$$

generalizing Eq. (A.15) for an unbinned galaxy field. If X is a binned FRB field, in DM bin $({\rm DM}_{\rm min}, {\rm DM}_{\rm max})$, then:

$$\bar{N}_f(M,z) = \begin{cases} \xi(z) \frac{M}{M_f} \int_{{\rm DM}_{\rm min} - {\rm DM}_{\rm IGM}(z)}^{{\rm DM}_{\rm max} - {\rm DM}_{\rm IGM}(z)} d({\rm DM}_{\rm host})\, p({\rm DM}_{\rm host}) \\ \qquad\qquad \text{if } M \geq M_f \\ 0 \qquad \text{if } M < M_f \end{cases} \tag{A.28}$$

generalizing Eq. (A.20) for an unbinned FRB field. Here, $p({\rm DM}_{\rm host})$ is the host DM probability distribution in Eq. (A.23), and ${\rm DM}_{\rm IGM}(z)$ is the IGM contribution to the DM at redshift z (Eq. 3.3).

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

$$n_X^{3d}(z) = \int dM\, n_h(M) \bar{N}_X(M,z) \tag{A.29}$$

$$n_X^{2d} = \int dz\, \frac{\chi(z)^2}{H(z)} n_X^{3d}(z) \tag{A.30}$$

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

$$n^{2d}_{f_i f_j} = n^{2d}_{f_i}\delta_{ij} \quad n^{2d}_{g_i g_j} = n^{2d}_{g_i}\delta_{ij} \quad n^{2d}_{f_i g_j} = 0 \tag{A.31}$$

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

$$u^g_\ell(M, z) = \tilde{u}(M, k, z)_{k=\ell/\chi(z)} \tag{A.32}$$

$$u^f_\ell(M, z) = b_\ell \tilde{u}(M, k, z)_{k=\ell/\chi(z)} \tag{A.33}$$

in the galaxy and FRB cases respectively. Here, $\tilde{u}$ is the 3D NFW profile in Eq. (A.8), and

$$b_\ell \equiv \exp\left(-\frac{\theta_f^2 \ell^2}{16 \log 2}\right) \tag{A.34}$$

is the Fourier-transformed FRB error distribution from Eq. (A.26).

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

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

$$X(\boldsymbol{\theta}) = \frac{1}{n^{2d}_X} \sum_j \delta^2(\boldsymbol{\theta} - \boldsymbol{\theta}^X_j) \tag{A.35}$$

$$X(\boldsymbol{\ell}) = \frac{1}{n^{2d}_X} \sum_j e^{-i\boldsymbol{\ell}\cdot\boldsymbol{\theta}^X_j} \tag{A.36}$$

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

$$\begin{aligned} \langle X(\boldsymbol{\ell})^* Y(\boldsymbol{\ell}')\rangle &= \frac{1}{n^{2d}_X n^{2d}_Y} \left\langle \sum_{jk} e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}^X_j - i\boldsymbol{\ell}'\cdot\boldsymbol{\theta}^Y_k} \right\rangle \\ &= C^{XY}_\ell (2\pi)^2 \delta^2(\boldsymbol{\ell} - \boldsymbol{\ell}') \end{aligned} \tag{A.37}$$

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

$$C_\ell^{XY} = C_\ell^{XY(2h)} + C_\ell^{XY(1h)} + C_\ell^{XY(p)} \tag{A.38}$$

which are given explicitly as follows:

$$\begin{aligned} C_\ell^{XY(2h)} &= \frac{1}{n_X^{2d} n_Y^{2d}} \int dz \frac{\chi(z)^2}{H(z)} n_X^{3d}(z) n_Y^{3d}(z) \\ &\quad \times b_X(z,\ell) b_Y(z,\ell) P_{\rm lin}(k,z) \\ C_\ell^{XY(1h)} &= \frac{1}{n_X^{2d} n_Y^{2d}} \int dz\, dM\, \frac{\chi(z)^2}{H(z)} n_h(M,z) \\ &\quad \times \bar{N}_X(M,z) \bar{N}_Y(M,z) u_\ell^X(M,z) u_\ell^Y(M,z) \\ C_\ell^{XY(p)} &= \frac{n_{XY}^{2d}}{n_X^{2d}\, n_Y^{2d}} \end{aligned} \tag{A.39}$$

where in the first line we have defined:

$$b_X(z,\ell) \equiv \frac{1}{n_X^{3d}(z)} \int dM b_h(M,z) n_h(M,z) \bar{N}_X(M,z) u_\ell^X(M,z) \tag{A.40}$$

On large scales (where $u_\ell = 1$), the quantity $b_X(z,\ell)$ reduces to the bias parameter $b_X(z)$ defined in §2.3.

Throughout this thesis, we have generally neglected the Poisson term in C_ℓ^{fg}, which arises if FRBs are actually located in survey galaxies (in contrast to the 1-halo term, which arises if FRBs are in the same halos as the survey galaxies). This is equivalent to our assumption in Eq. (A.31) that $n_{fg}^{2d} = 0$. If this assumption is relaxed, then $C_\ell^{fg(p)}$ will be given by:

$$C_\ell^{fg(p)} = b_\ell \frac{n_{fg}^{2d}}{n_f^{2d}\, n_g^{2d}} \tag{A.41}$$

where the FRB beam convolution b_ℓ has been inserted by hand into the general expression in Eq. (A.39), since the FRB beam displaces FRBs relative to their host galaxies.

A.5 Free electrons

When modeling propagation effects (§2.5), the 3D galaxy-electron power spectrum $P_{ge}(k,z)$ appears. This can also be computed in the halo model.

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

We will also make the approximation that electrons have the same halo profiles as dark matter. This is a good approximation on large scales, but may overpredict P_{ge} on small scales by a factor of a few. This happens because dark matter is pressureless, whereas electrons have associated gas pressure, which "puffs out" the profile. In §2 our goal is modeling propagation effects at the order-of-magnitude level, and it suffices to approximate electron profiles by dark matter profiles. For a more precise treatment, fitting functions for electron profiles could be used [12].

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

$$\begin{aligned} P_{ge}^{1h}(k,z) &= \frac{1}{\rho_{m,0} n_g^{3d}(z)} \int dM\, M n_h(M,z) \bar{N}_g(M,z) \tilde{u}(M,k,z)^2 \\ P_{ge}^{2h}(k,z) &= b_g(k,z) b_e(k,z) P_{\rm lin}(k,z) \end{aligned} \tag{A.42}$$

where:

$$\begin{aligned} b_e(k,z) &= \frac{1}{\rho_{m,0}} \int dM\, M b_h(M,z) n_h(M,z) \tilde{u}(M,k,z) \\ b_g(k,z) &= \frac{1}{n_g^{3d}(z)} \int dM\, b_h(M,z) n_h(M,z) \bar{N}_g(M,z) \tilde{u}(M,k,z) \end{aligned} \tag{A.43}$$

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

Appendix B

Accuracy of the Limber approximation

Throughout the thesis, angular power spectra have been calculated using the Limber approximation [64, 52, 53]. Let X, Y be 2D fields which are obtained from the 3D density field $\delta_{\rm lin}(\mathbf{x})$ by line-of-sight integration:

$$\begin{aligned} X(\boldsymbol{\theta}) &= \int d\chi\, W_X(\chi)\delta_{\rm lin}(\chi\boldsymbol{\theta}) \\ Y(\boldsymbol{\theta}) &= \int d\chi\, W_Y(\chi)\delta_{\rm lin}(\chi\boldsymbol{\theta}) \end{aligned} \tag{B.1}$$

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

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

For example, the 2-halo power spectrum $C_\ell^{fg(2h)}$ in Eq. (2.13) was calculated by applying the Limber approximation with weight functions:

$$W_X(z) = b_X(z)\frac{dn_X^{2d}}{dz} \qquad (X \in \{f, g\}) \tag{B.3}$$

How accurate is the Limber approximation? For a detailed analysis, including explicit calculation of subleading terms, see [67]. At back-of-the-envelope level, the Limber approximation is accurate if

$$\ell \gg \left(\frac{d \log W_X}{d \log \chi} \frac{d \log W_Y}{d \log \chi} \right)^{1/2} \tag{B.4}$$

In most of this thesis, the factors $(d \log W / d \log \chi)$ are of order 1, and therefore the Limber approximation is accurate for $\ell \gg 1$. There is one exception: when we calculate C_ℓ^{fg} for a narrow FRB DM slice in our fiducial high-z FRB model (§2.2), we have $(d \log W / d \log \chi) \approx z_f / (\Delta z_f)$, where the mean FRB redshift z_f can be as large as 3, and the width (Δz_f) of the FRB redshift distribution can be as small as 0.1. In this case, the Limber approximation will still be accurate for $\ell \gg (3/0.1)^{1/2} \approx 6$, which is sufficient for purposes of this thesis.

One more subtle point. We sometimes consider the limit of narrow redshift bins, for example when computing Fisher matrix forecasts in §2.6. Generally speaking, the Limber approximation for C_ℓ^{fg} breaks down when the redshift bin width (Δz) is taken to zero. However, the Fisher forecast converges as $(\Delta z) \to 0$: the Fisher matrix with $(\Delta z) = 0$ is nearly equal to the Fisher matrix with $(\Delta z) = 0.1$ (or smaller), and the Limber approximation is still accurate at $(\Delta z) = 0.1$. This is partly because the low-ℓ end of harmonic space contains a small area (see e.g. Figure 2.8). Therefore, the Limber-approximated narrow-bin Fisher matrix is a good approximation to the exact Fisher matrix.

Appendix C

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 $C_\ell^{fg} \to b_\ell C_\ell^{fg}$, where b_ℓ is the "beam" transfer function. Throughout §3, 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 §3, where a Gaussian transfer function $b_\ell = e^{-\ell^2/L^2}$ is used, but L is a free parameter which we fit to the data, varying L over the range $315 \leq L \leq 1396$.

C.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 non-detection) in each beam, using the localization pipeline described by [25, 29]. For an FRB which is detected in a single beam, the localization pipeline assigns sky location equal to the beam center. For a multi-beam 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 section.) 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:

$$S = \left[-\frac{\theta_0}{2}, \frac{\theta_0}{2}\right] \times \left[-\frac{\theta_0}{2\sin\Theta_e}, \frac{\theta_0}{2\sin\Theta_e}\right] , \tag{C.1}$$

where $\theta_0 = 23'.4$ in CHIME. If the detected FRBs are uniformly distributed on the sky, then the effective beam is:

$$b_\ell = \frac{\int_S d^2\theta\, J_0(\ell\theta)}{\int_S d^2\theta\, 1} , \tag{C.2}$$

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 C.1, and agrees well with a Gaussian transfer function $b_\ell = e^{-\ell^2/L^2}$ with $L = 670$.

C.2 Toy beam model 2: including selection bias

In the previous section, 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

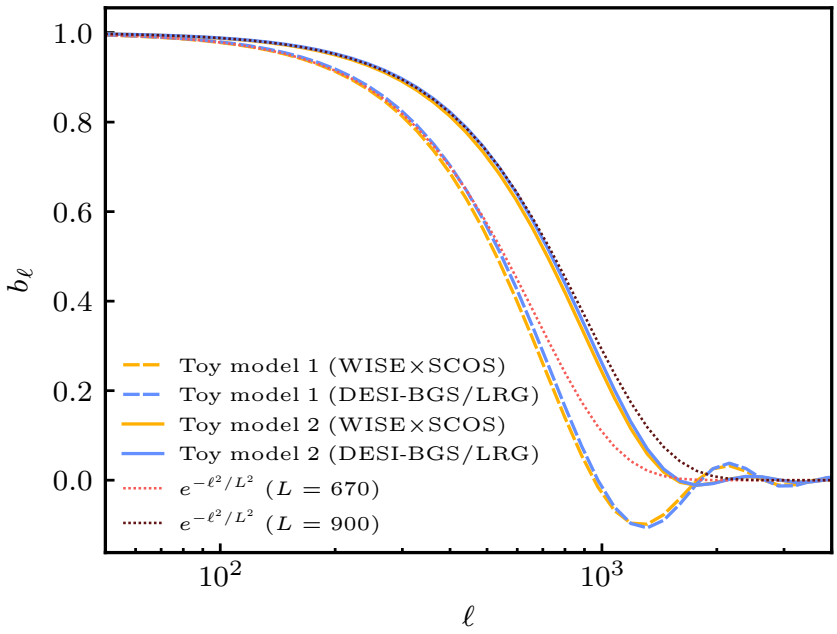

Figure C.1: CHIME/FRB beam transfer function b_ℓ in a toy beam model, without (model 1, §C.1) and with (model 2, §C.2) selection bias included. Since b_ℓ is elevation-dependent, the result is slightly different after averaging over FRBs in the WISE×SCOS (orange) and DESI-BGS/LRG (blue) sky regions. For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

account for this selection bias, we define the unnormalized intensity beam:

$$B(\theta_x, \theta_y) = \frac{\mathrm{sinc}^2(\theta_x D_x/\lambda)\,\mathrm{sinc}^2((\theta_y D_y \sin\Theta_e)/\lambda)}{\mathrm{sinc}^2(\theta_x D_x/4\lambda)}\,, \tag{C.3}$$

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

Assuming a Euclidean FRB fluence distribution $N(\geq F) \propto F^{-3/2}$ (consistent with statistical analysis of the CHIME/FRB Catalog 1 [29]), 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:

$$b_\ell = \frac{\int_S d^2\theta\, B(\theta)^{3/2} J_0(\ell\theta)}{\int_S d^2\theta\, B(\theta)^{3/2}} \tag{C.4}$$

averaged over catalog elevations Θ_e as in the previous section. The resulting transfer function b_ℓ is shown in Figure C.1 and agrees well with a Gaussian transfer function $b_\ell = e^{-\ell^2/L^2}$ with $L = 900$.

C.3 Plausible range of L-values

Comparing the last two sections, we see that the selection bias considered in §C.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 thesis. 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 multi-beam 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 which lead to the largest plausible localization errors. We start with the toy beam model b_ℓ from §C.2, with $\lambda = 0.75\,\mathrm{m}$ (the longest wavelength in CHIME). We then convolve with a halo profile ($b_\ell \to b_\ell u_\ell(M, z)^2$, where $u_\ell(M, z)$ is a Navarro-Frenk-White (NFW) density profile [82]), 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 $L_{\mathrm{min}} = 315$.

Similarly, to assign a largest plausible L-value, we make assumptions which lead to the smallest plausible localization errors. We use the smallest toy model from §C.2 with

$\lambda = 0.375\,\mathrm{m}$ (the shortest wavelength in CHIME). We assume that 40% of the events are multi-beam detections, and that multi-beam detections have localization errors which 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 $L_{\max} = 1396$.

Appendix D

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:

$$d\hat{C}_\ell^{fg} \equiv \hat{C}_\ell^{fg\,(\mathrm{low})} - \hat{C}_\ell^{fg\,(\mathrm{high})} \tag{D.1}$$

Recall that for a non-null power spectrum $\hat{C}_\ell$, we compressed the ℓ-dependence into a scalar summary statistic $\hat{\alpha}_L$ by taking a weighted ℓ-average (Eq. 3.11). Analogously, we compress the difference spectrum $d\hat{C}_\ell^{fg}$ into a summary statistic $\hat{\beta}_L$, defined by:

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

where L is an angular scale parameter. Next, by analogy with SNR_L (defined previously in Eq. 3.14), we define:

$$\Delta_L = \frac{\hat{\beta}_L}{\mathrm{Var}(\hat{\beta}_L)^{1/2}} \tag{D.3}$$

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×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 Figures 3.7–3.9), 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×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 D.1, we identify 12 such variables, and denote the corresponding Δ_L values (with $L = 1000$) by Δ_i, where $i \in \{1, 2, 3, \ldots, 12\}$. We note that these 12 tests are non-independent, 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 D.1, but a few outliers are unsurprising, so it is not immediately clear whether the Δ_i values in Table D.1 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 D.1 is statistically significant. We define:

$$\Delta_{\max} = \max_i |\Delta_i| \tag{D.4}$$

We then compare these values of $\Delta_{\max}$ 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.

Parameter	Median (WISE×SCOS)	Δ_i WISE×SCOS	Median (DESI)	Δ_i DESI-BGS	Δ_i 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 $\lvert b\rvert$	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 D.1: Null tests in Appendix D. 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. D.3), which measures consistency of the FRB-galaxy correlation in "sigmas".

In Table D.2, we report the p-value for each $\Delta_{\rm max}$, i.e. the fraction of mocks whose $\Delta_{\rm max}$ exceeds the "data" value. No statistically significant deviation from $\Delta_{\rm max} = 0$ 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:

$$\chi^2 = \sum_{i,i'} \Delta_i \, {\rm Cov}\left(\Delta_i, \Delta_{i'}\right)^{-1} \Delta_{i'} \, , \tag{D.5}$$

where the covariance ${\rm Cov}(\Delta_i, \Delta_{i'})$ 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 D.2. We find borderline evidence for $\chi^2 \neq 0$ for DESI-BGS ($p = 0.030$), but interpret this as inconclusive, since Table D.2 contains six p-values, so one p-value as small as 0.03 is unsurprising (this happens with probability $\approx$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, [44]) and Anderson-Darling (AD, [108]) tests. Figure D.1 compares the two distributions for the three galaxy samples, and the last two columns of Table D.2 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 §3 depends on any of the parameters in Table D.1.

Galaxy sample	$\Delta_{\max}$	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 D.2: Summary statistics for the twelve null tests in Table D.1. As described in Appendix D, we reduce the twelve-component vector Δ_i into two scalar summary statistics $\Delta_{\max}, \chi^2$, 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.

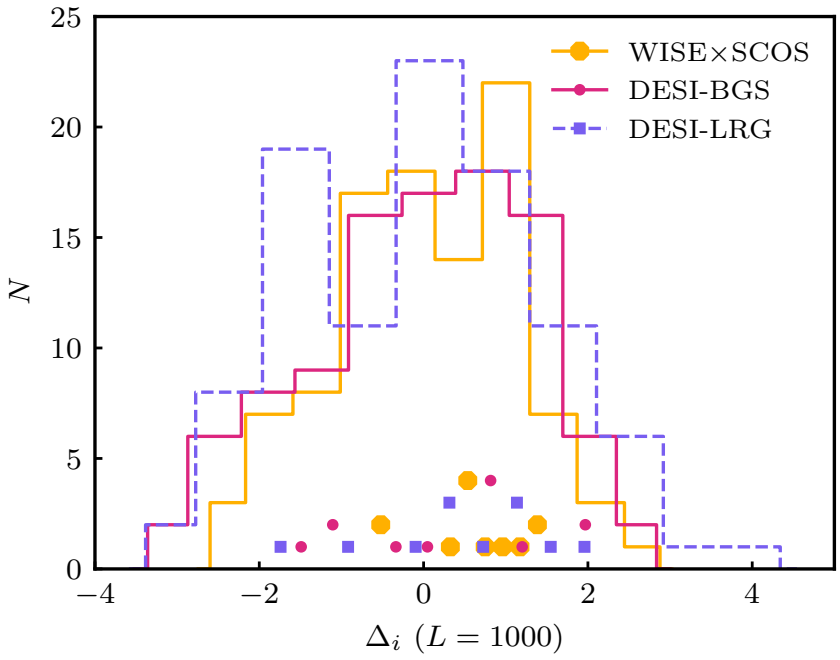

Figure D.1: Histograms of the statistic Δ_i for the 12 null tests (filled markers) and 100 jackknives (lines). Using the KS and AD tests, the distributions are found to be consistent (Appendix D).

Appendix E

Tail-fitting procedure

In §3.3.5, we assign statistical significance of the FRB-galaxy detection, by defining a frequentist statistic $\mathrm{SNR}_{\mathrm{max}}$, and ranking the "data" value $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{data})}$ within a histogram of simulated values $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{mock})}$. This procedure is conceptually straightforward, but there is a technical challenge: because $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{data})}$ turns out to be an extreme outlier, a brute-force approach requires an impractical number of simulations. Therefore, we fit the tail of the $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{mock})}$ distribution to an analytic distribution, and assign statistical significance (or p-value) analytically.

Empirically, we find that the top 10% of the $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{mock})}$ distribution agrees well with the top 10% of a Gaussian distribution, as shown in Figure E.1. The parameters of the Gaussian distribution were determined as follows. Let $p(x|\mu,\sigma)$ denote a Gaussian distribution with mean μ and variance σ^2:

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

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

$$\begin{aligned}\log \mathcal{L}(x|\mu,\sigma) = &\left(\sum_{x\in\Sigma_+} \log p(x|\mu,\sigma) \right) \\ &+ \left|\Sigma_-\right| \log \int_{-\infty}^{\Sigma_0} p(x|\mu,\sigma)\,,\end{aligned} \tag{E.2}$$

where x denotes mock realizations. This likelihood function has been constructed to fit

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

Table E.1: "Brute-force" and analytic p-values, computed as described in Appendix E.

parameters to the details of the Σ_+ values, while putting all Σ_- values into a single coarse bin.

Figure E.1 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 E.1, 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 $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{mock})}$ values (out of 10^4 total simulations) that exceed $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{data})}$. The "analytic" p-value is obtained by fitting the top 10% of the simulated $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{mock})}$ values to a Gaussian distribution, as described above, and evaluating the CDF of the distribution at $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{data})}$. 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 §3.

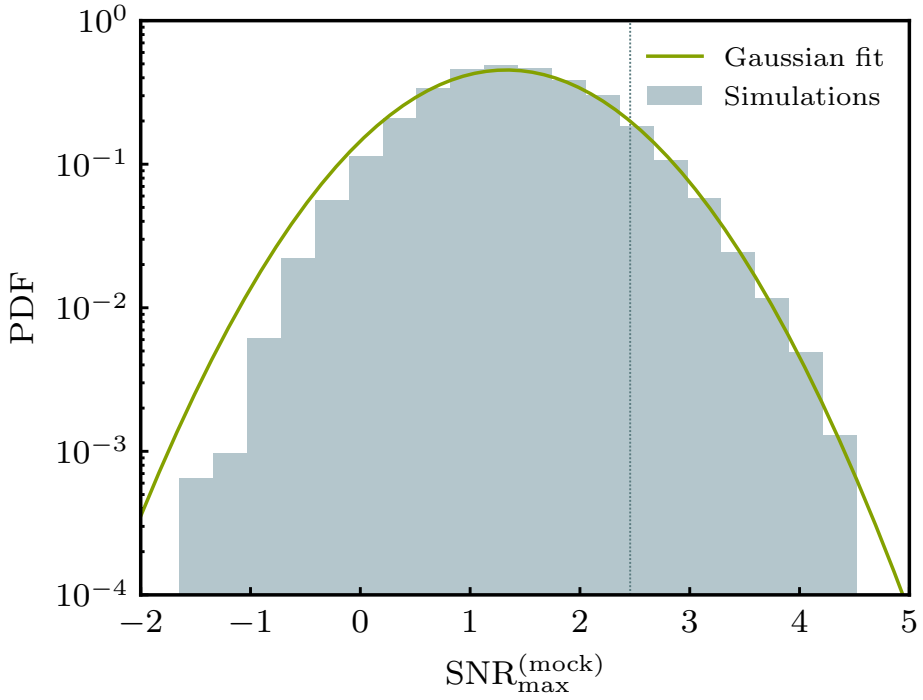

Figure E.1: Gaussian fit to the tail of the $\mathrm{SNR}_{\mathrm{max}}^{(\mathrm{mock})}$ distribution from Appendix E. For the top ∼10% of the samples (i.e. to the right of the dotted line) the agreement between the fit and the simulations is excellent. This plot is for WISE×SCOS; the other two cases (DESI-BGS, DESI-LRG) are similar.