Constraining the Dispersion Measure of the Galactic Halo using FRBs and Density Estimation

E. Platts[1⋆], J. Xavier Prochaska[2,3] and Casey J. Law[4]

[1] *High Energy Physics, Cosmology & Astrophysics Theory (HEPCAT) group, Department of Mathematics and Applied Mathematics, University of*
[2] *Department of Astronomy & Astrophysics, UC Santa Cruz, USA*
[3] *Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU; WPI), The University of Tokyo, Japan*
[4] *Department of Astronomy and Radio Astronomy Lab, University of California, Berkeley, CA 94720, USA*

30 September 2019

ABSTRACT

It has recently been shown that Fast Radio Bursts (FRBs) present a new means to constrain the distribution of baryons in the Universe (Chatterjee et al. 2017; Tendulkar et al. 2017). As these short, coherent bursts travel through space, the refractive index of intervening plasma causes a delay in arrival time between photons of different energies. This characteristic frequency sweep, described by a dispersion measure (DM), provides a constraint on the electron distribution along the line of sight between the source and the observer. On this basis, a signature of the intervening matter will be imprinted on the DM, presenting an opportunity to probe the illusive circumgalactic medium (CGM). Using Kernel Density Estimation (KDE) and Density Estimation using Field Theory (DEFT), we place constraints on the contribution of the Galactic halo to the DM of FRBs.

1 INTRODUCTION

There are two primary challenges that this paper explores: how one might use Fast Radio Bursts (FRBs) to probe the dispersion measure (DM)—and hence the the baryon mass density—of the Galactic halo; and how to do so with a limited dataset. The first problem is addressed by placing a constraint on the DM value where the Galactic and extragalactic contributions are separated. This value is referred to as the 'gap'. To date, however, there are only ~ 100 observed FRBs. This necessitates techniques that are well suited to dealing with small datasets. Further, given the complexity of modelling FRB dynamics, one must invoke a non-parametric technique. We propose the use of probability density estimations—in particular, Kernel Density Estimation (KDE) and Density Estimation using Field Theory (DEFT)—to find an ensemble of DM distributions that describe current data. Intrinsically, however, probability density functions (PDFs) have tails that exponentially decrease to zero, and therefore do not necessarily predict the true minimum/maximum of the model. As such, a metric must be established to estimate a feasible cut-off value. To do this, the DM distribution of FRBs is simulated with a known boundary between the extragalactic and galactic components of the DM. This metric is then applied to the PDF that describes the observed data.

1.1 The Galactic Halo

Rapid advancements in technology and a multitude of multi-wavelength surveys have provided potential to map the baryonic mass density of the universe (Fukugita et al. 1998; Prochaska & Tumlinson 2009; Prochaska & Zheng 2019a). Measurements of light element ratios (Burles & Tytler 1996; O'Meara et al. 2001; Cooke et al. 2018) and observations of the cosmic microwave background (CMB) (Spergel et al. 2007; Planck Collaboration et al. 2016) provide a snapshot of the universe at very early times. These observations provide tight constraints on the baryon mass fraction of the universe, $f_b = \Omega_b/\Omega_m = 0.175 \pm 0.012$. Since this ratio is independent of the Hubble constant, it aids in probing the baryonic distribution of the universe through time and space.

In the early universe ($z \sim 3$), the majority of baryons resided in a cool ($T \sim 10^4$ K), diffuse plasma. These baryons are predicted to have collapsed into sheetlike and filamentary structures that make up the intergalactic medium (IGM). These give rise to the HI Lyman-α forest—a collection of absorption lines that are observed in the spectra of quasistellar objects (QSOs) (Miralda-Escude et al. 1996; Rauch 1998). Around the time of structure formation, baryons are pulled by gravitationally-dominant dark matter as the dark matter collapses into halos. As the gas falls inwards, it is shock-heated to form a hot, diffuse gas, known as circumgalactic medium (CGM) (White & Rees 1978). $\sim 10\%$ of the gas cools and falls into the center of the halo to form stars and the interstellar medium (ISM). Comparing the baryonic mass fraction of galactic halos ($M_b/M_{\rm halo}$) to the cosmic mean (Ω_b/Ω_m), however, reveals a baryonic deficit (Dai et al. 2010). The missing baryons may have been ejected back into the IGM before forming stars (Prochaska et al. 2011; Booth et al. 2012), or perhaps they simply have yet to be detected. In the latter case, the CGM presents itself as a possible refuge for illusive baryons, and is the focus of this paper.

The CGM is a massive, extended reservoir (Chen et al. 2001; Werk et al. 2014; Lehner et al. 2015) of metal-enriched and multiphase gas (Werk et al. 2013; Lehner et al. 2014; Liang & Chen 2014; Prochaska et al. 2017) that pervades the dark matter halo. It comprises cool ($T \sim 10^4$ K) and dense gas clumps embedded in a hot ($T \sim 10^6$ K), diffuse plasma (Heitsch & Putman 2009; Stocke et al. 2013; Prochaska et al. 2017; Armillotta et al. 2017; Hani et al. 2019). It plays a key role in galaxy evolution: it provides a source of star-forming fuel, facilitates galactic feedback and recycling, and is the fundamental liaison between galactic baryons and the IGM (Putman et al. 2012; Tumlinson et al. 2017). As with the IGM, CGM can be observed in QSO spectra: when a galaxy falls along the line of sight of a QSO, it creates a set of characteristic absorption lines (Bergeron 1986; Bergeron & Boissé 1991; Lanzetta et al. 1995; Tripp et al. 2000; Chen et al. 2001; Prochaska et al. 2011; Tumlinson et al. 2013). The cool CGM ($< 10^5$ K) can be measured with ultraviolet (UV) absorption lines (Savage et al. 2011; Burchett et al. 2019), and the hot CGM ($> 10^6$ K) with X-ray emission (Fang et al. 2015; Nicastro et al. 2018) or Sunyaev-Zeldovich (SZ) signals (Lim et al. 2018; Hill et al. 2018). Different observations and analyses, however, produce significantly different results (Anderson et al. 2013; Planck Collaboration et al. 2013; Werk et al. 2014; Keeney et al. 2017; Lim et al. 2018), and current telescope sensitivities are insufficient to probe lower mass galaxies. This has encouraged observers to seek other avenues. Recently, Chatterjee et al. (2017) and Tendulkar et al. (2017) proposed that the CGM may be evidenced by the DM of FRBs. We endeavour to explore this possibility.

1.2 Fast Radio Bursts

Fast Radio Bursts (FRBs) are very bright (∼Jy), brief (∼ms) extragalactic radio transients that have sparked widespread excitement and intrigue in astrophysics and cosmology communities. The first FRB was discovered a decade ago in archival Parkes telescope data (Lorimer et al. 2007), and only ∼ 100 events have been observed since (see the online FRB catalogue for up to date and open source data (Petroff et al. 2016)). Such elusiveness has made it difficult to ascertain FRB origins or mechanisms, and as the few observations have come in, the plot seems to have thickened. FRBs exhibit a characteristic dispersion in their arrival time—going as $\Delta t \sim \Delta\nu^{-2}$—however there is huge variation in other observed properties. Some bursts have circular (Ravi et al. 2015; Petroff et al. 2015; Caleb et al. 2018) and/or linear (Masui et al. 2015; Ravi et al. 2016; Michilli et al. 2018; Gajjar et al. 2018) polarizations; rotation measures (RMs) have been observed between ~ 10 rad m^{-2} (Masui et al. 2015) and $\sim 10^5$ rad m^{-2} (Michilli et al. 2018); DMs have been observed between 103.5 cm^{-3}pc (Andersen et al. 2019) and 2596.1 cm^{-3}pc (Caleb et al. 2018); some bursts different pulse profiles (Champion et al. 2016; Farah et al. 2018); and 11 only have been observed to repeat (Spitler et al. 2016, 2018; Andersen et al. 2019; Kumar et al. 2019). Further, only 3 FRBs have been localised to host galaxies, and all are different: FRB 121102 resides in a low-metalicity, star-forming dwarf galaxy at redshift $z = 0.19$ (Chatterjee et al. 2017; Tendulkar et al. 2017; Bassa et al. 2017); FRB 190523 in a massive galaxy with a low star-formation rate at redshift $z = 0.66$ (Ravi et al. 2019); and FRB 180924 at the center of a medium-sized luminous galaxy at redshift $z = 0.3214$. As such, nearly 50 theories have been proposed purporting to explain the phenomena. For a detailed description of these theories, see Platts et al. (2018) and the associated online repository for updates. For a comprehensive review of FRBs and our current understanding, see Petroff et al. (2019). Further reviews are given by Katz (2016); Petroff (2017); Katz (2018); Popov et al. (2018); Cordes & Chatterjee (2019).

While FRBs are compelling in their own right, they also offer huge potential as probes of the Universe. Proposals include (but aren't limited to) studying the intergalactic medium (IGM) (Deng & Zhang 2014; Zheng et al. 2014; Macquart et al. 2015; Akahori et al. 2016; Shull & Danforth 2018; Ravi 2019), the baryonic and dark matter distribution (Gao et al. 2014; Muñoz et al. 2016; Wang & Wang 2018), and the cosmic web (Ravi et al. 2016), and tightening constraints on cosmological parameters (Zhou et al. 2014; Yang & Zhang 2016; Walters et al. 2018; Yu & Wang 2017; Li et al. 2018; Zitrin & Eichler 2018; Wei et al. 2018; Walters et al. 2019). In this paper, we investigate how one can use FRBs to probe the Galactic halo of the Milky Way (Prochaska & Zheng 2019b).

2 OVERVIEW

A brief pulse of radiation will be scattered by free electrons in accordance with the photon energy and the refractive index of intervening plasma. This results in a delay in photon arrival time, observed as a frequency sweep and described by a dispersion measure (DM). The DM is defined as the integrated column density of free electrons (n_e) between an observer and a source:

$$DM = \int \frac{n_e ds}{1+z}. \tag{2.1}$$

The DM of each FRB therefore harbours information about the baryonic density between the FRB source and Earth. Further, FRBs appear to be isotropic across the sky (Champion et al. 2016), making them ideal to probe the Galaxy from all directions. The observed DMs of FRBs have contributions from the ISM ($DM_{\rm ISM}$), the Galactic halo ($DM_{\rm halo}$), the IGM and intervening galaxies ($DM_{\rm cosmic}$), and the host galaxy ($DM_{\rm host}$). $DM_{\rm ISM}$ can be obtained with observations of pulsars in and around the Galaxy using the publicly available Cordes-Lazio NE2001 Galactic Free Electron Density Model (Cordes & Lazio 2002, 2003). For now this contribution is absorbed into the Galactic halo DM. One therefore has:

$$DM_{\rm FRB} = DM_{\rm MW_{halo}} + DM_{\rm cosmic} + DM_{\rm host}. \tag{2.2}$$

We next define the gap: the maximum value for $DM_{\rm halo}$ and the minimum value for $DM_{\rm FRB}$, ie. a constraint on the region dividing Galactic and extragalactic components of the DM:

$$DM_{\rm gap} = DM_{\rm MW_{halo}} + \min(DM_{\rm cosmic}) + \min(DM_{\rm host}). \tag{2.3}$$

The complex and largely unknown dynamics of FRBs suggest a non-parametric approach, and we thus consider PDFs of the $DM_{\rm FRB}$ distribution to estimate $DM_{\rm gap}$. PDFs, however, have exponentially decreasing tails that do not necessarily describe the true minimum of the data. One must thus enforce a cut-off value. Establishing this point, however, is tricky: there are no statistically rigorous ways in which it can be done. One might choose a percentile at which to trim the tail, but this will give a very loose estimate. By simulating the PDF of a mock universe, one has a known $DM_{\rm gap}$ value. One can then takes random draws from the PDF of the mock universe. Those draws that fall below $DM_{\rm gap}$ are incorrect predictions of the model, and can thus be used to establish a metric to determine a cut-off value. The region that lies before $DM_{\rm gap,sim}$ can therefore be used as a proxy to determine $DM_{\rm gap,obs}$.

3 DENSITY ESTIMATION TECHNIQUES

With limited data, it is difficult to establish a representative probability density function (PDF), and so a number of tools have been developed to deal with the challenge. This paper considers two methods: Kernel Density Estimation (KDE) (Silverman 1986) and Density Estimation using Field Theory (DEFT) (Kinney 2014, 2015; Chen et al. 2018).

3.1 Kernel Density Estimation

This non-parametric technique estimates an unknown density by constructing a kernel at each data point and summing their contributions. Consider an independent and identically distributed sample $(x_1, x_2, ..., x_n)$ drawn from some unknown distribution $Q_{\rm true}(x)$. We wish to obtain an estimate $\hat{Q}(x)$ of this distribution using KDE:

$$\hat{Q}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \quad (3.1)$$

where K is the kernel and $h > 0$ is the bandwidth. The kernel is the underlying distribution function (most commonly chosen to be uniform, triangular, biweight, triweight, Epanechnikov or Gaussian), and the bandwidth is a smoothing parameter. Bandwidth selection is thus a trade-off between the bias of the KDE and its variance. Most commonly, the bandwidth is chosen to minimize the mean integrated squared error (MISE) over the entire dataset. In the small data regime, however, this is often inappropriate. Instead, we invoke a k-nearest neighbours (kNN) approach with cross-validation. Here a data point and a number of its neighbours (k) are assigned a bandwidth based on the MISE. This is done for each data point, and a voting system is used to select the best bandwidth. Cross-validation is performed to find the optimal k. For kernel selection though, there are very few statistical techniques available, and it is primarily done by inspection. The most commonly invoked kernel is Gaussian, and is used in this work. One must also decide how to treat points near the boundary Jones (1993), for which there is currently no consensus. Due to time restrictions, we do not investigate this aspect here, but leave to future work.

An ensemble of plausible $\hat{Q} = DM_{\rm FRB}$ distributions are found by resampling with replacement, or bootstrapping. Each resampled dataset has its bandwidth independently optimized. The best estimate $Q^* = DM^*_{\rm FRB}$ is given by the mean of the KDE ensemble. $DM_{\rm gap}$ is calculated with $DM^*_{\rm FRB}$ and the uncertainty of $DM_{\rm gap}$ with the ensemble of $DM_{\rm FRB}$ distributions.

3.2 Density Estimation Using Field Theory

Density Estimation using Field Theory (DEFT) takes a Bayesian field theory approach to density estimation in small datasets (Kinney 2014, 2015; Chen et al. 2018) using a Laplace approximation of the Bayesian posterior (also see Riihimaki & Vehtari (2014)). An advantage of DEFT over commonly invoked density estimation methods—including KDE—is that the method does not require the manual identification of critical parameters nor requires the specification of boundary conditions. The DEFT simulations in this report use the Python package SUFTware (Statistics Using Field Theory) Chen et al. (2018).

Consider n data points $(x_1, x_2, ..., x_n)$ drawn from known probability distribution $Q_{\rm true}(x)$ with x intervals of length L. We wish to find the best estimate $Q^*(x)$ of this distribution and the accompanying ensemble of other plausible estimates. Each distribution $Q(x)$ is parameterized by a real field $\phi(x)$, ensuring that $Q(x)$ is positive and normalized:

$$Q(x) = \frac{e^{-\phi(x)}}{\int dx' e^{-\phi(x')}}. \quad (3.2)$$

Using scalar field theory, a prior $p(\phi|\ell)$ is formulated that favours smooth probability densities. Specifically, Kinney (2015) consider priors of the form

$$p(\phi|\ell) = \frac{e^{-S^0_\ell[\phi]}}{Z^0_\ell} \quad (3.3)$$

with action

$$S^0_\ell[\phi] = \int \frac{dx}{L}\frac{\ell^{2\alpha}}{2}\left(\partial^\alpha\phi\right)^2, \quad (3.4)$$

and partition function

$$Z^0_\ell = \int \mathcal{D}\phi e^{-S^0_\ell[\phi]}. \quad (3.5)$$

ℓ gives the length scale below which ϕ fluctuations are strongly damped, and $\alpha > 0$ (chosen by hand) is an integer that determines the smoothness. The resultant posterior is given by

$$p(\phi|\text{data}, \ell) = \frac{e^{-S_\ell[\phi]}}{Z_\ell}, \quad (3.6)$$

with nonlinear action

$$S_\ell[\phi] = \int \frac{dx}{L}\left\{\frac{\ell^2\alpha}{2}\left(\partial^\alpha\phi\right)^2 + nLR\phi + ne^{-\phi}\right\}, \quad (3.7)$$

and partition function

$$Z_\ell = \int \mathcal{D}\phi e^{-S_\ell[\phi]}, \quad (3.8)$$

where $R(x) = \frac{1}{n}\sum_{i=1}^{n}\partial(x - x_i)$ is a histogram that summarizes the data.

Maximum *a posteriori* (MAP) density estimation approximates the posterior $p(\phi|\text{data}, \ell)$ as a δ function given by the mode of the posterior, at which the action $S_\ell[\phi]$ is then minimized. It has been shown that even without imposing boundary conditions on ϕ, $S_\ell[\phi]$ has a unique minimum (Kinney 2015). The optimal length scale ℓ^* is identified by maximizing the Bayesian evidence $p(\text{data}|\ell)$.

The uncertainty in the DEFT estimate Q^* is determined by sampling the Bayesian posterior,

$$p(Q|\text{data}) = \int dl p(\ell|\text{data}) p(Q|\text{data}, \ell), \tag{3.9}$$

by first drawing ℓ from $p(\ell|\text{data})$ and then drawing Q from $p(Q|\text{data}, \ell)$. Laplace approximation is used to estimate $p(Q|\text{data}, \ell)$ by constructing a Gaussian centered at its MAP value. This gives the Laplace posterior,

$$p_{\text{Lap}}(Q|\text{data}) = \int dl p(\ell|\text{data}) p_{\text{Lap}}(Q|\text{data}, \ell), \tag{3.10}$$

from which an ensemble of distributions Q can be sampled. Some of the Qs generated, however, have wisps that clearly do not represent the data. Importance resampling is used to resolve this, where each ϕ is given a weight

$$w_\ell[\phi] = \exp\left(S_\ell^{\text{Lap}}[\phi] - S_\ell[\phi]\right) \tag{3.11}$$

proportional to its probability of being drawn (Chen et al. 2018).

DM_{gap} is calculated from the best estimate $Q^* = DM^*_{\text{FRB}}$, and the uncertainty of DM_{FRB} from the $Q(x) = DM_{\text{FRB}}$ ensemble.

4 METHODOLOGY AND RESULTS

The core idea behind this work is that if the PDF of $DM_{\text{FRB,sim}}$ can be reasonably approximated by a small subset of data, one can use this as a proxy to imply a value for $DM_{\text{gap,obs}}$. A density function for DM_{FRB} is simulated, from which a small number of samples are drawn. These samples are used to estimate $DM_{\text{FRB,sim}}$ using KDE and DEFT. Since $DM_{\text{FRB,sim}}$ is a known PDF, we have a known value for $DM_{\text{gap,sim}}$. Samples randomly drawn from the distribution that fall below $DM_{\text{gap,sim}}$ are incorrect predictions for the gap. One can therefore use the area of the $DM_{\text{FRB,sim}}$ PDF below $DM_{\text{gap,sim}}$ to determine a cut-off value for the PDF. If $DM_{\text{FRB,obs}}$ is reasonably similar to $DM_{\text{FRB,sim}}$, one can apply metric to find the cut-off value of $DM_{\text{FRB,obs}}$. For this to be applicable to $DM_{\text{FRB,obs}}$, the number of DM samples drawn from the $DM_{\text{FRB,sim}}$ PDF should equal the number of observed FRB DMs.

4.1 The Simulated Universe

A PDF is built for each of the three DM contributions—$DM_{\text{MW}_{\text{halo}}}$, DM_{cosmic} and DM_{host}—which are then summed to get the total DM_{FRB} density. DM_{host} is chosen to be some lognormal distribution with a minimum value of 40pc cm^{-3} and $DM_{\text{MW}_{\text{halo}}}$ is chosen to be a delta function at 60pc cm^{-3}. These PDFs are rough estimates, but will suffice: for now we assume DM_{FRB} is a function of z and thus $DM_{\text{MW}_{\text{halo}}}$ and $DM_{\text{MW}_{\text{host}}}$ are unlikely to affect the overall shape of the DM_{FRB} PDF. This is because the redshift distance of the Milky Way and of the host galaxy are both small

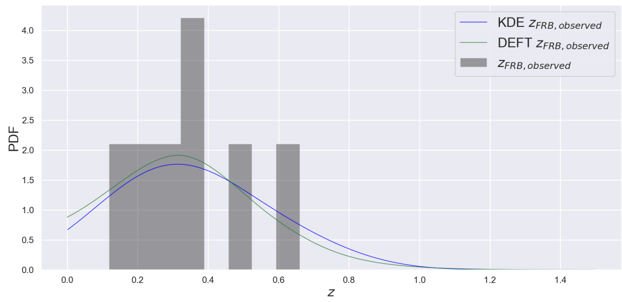


Figure 1. Simulated and observed PDFs of FRB redshift distributions using z_{FRB} obtained with KDE and DEFT

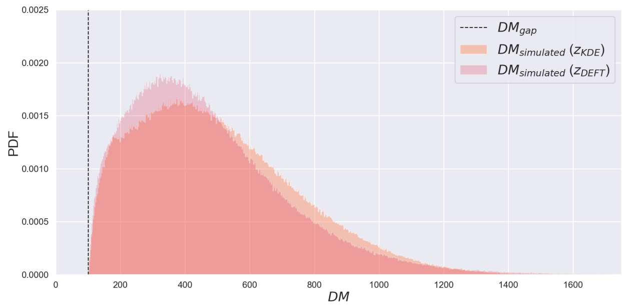


Figure 2. PDF of $DM_{FRB,simulated}$ and DM_{gap}.

compared to that of the IGM. Together they set the simulated gap value (Equation 2.3) to $DM_{\text{gap}} = 100$pc cm^{-3}.

To find $DM_{\text{cosmic}}(z)$ one must first obtain the z density for FRBs, ie. the likelihood of an FRB event taking place within a z bin. Both density estimation techniques are used to obtain independent DM_{FRB} densities for each analysis[1].

Random draws from the z density function are used to find the average cosmic contribution to the DM, $\langle DM_{\text{cosmic}}(z)\rangle$. $\langle DM_{\text{cosmic}}(z)\rangle$ is based on current empirical knowledge of baryon distributions and ionization states, including the IGM and galactic halos (Prochaska & Zheng 2019b).

$$\langle DM_{\text{cosmic}}\rangle = \int \frac{\bar{n}_e ds}{1+z}, \tag{4.1}$$

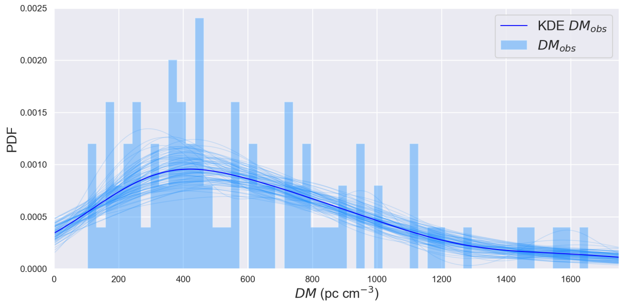
where $\bar{n}_e = f_d(z)\rho_b(z)\mu_e/\mu_m m_p$ is the average electron density, f_d is the fraction of cosmic baryons in diffuse ionised gas, $\rho_b \equiv \Omega_b \rho_c$ is the cosmic baryonic mass density, and μ_m and μ_e describe properties of Helium[2].

To obtain DM_{cosmic}, the variance $\sigma = f_d(z)z^{-1/2}$ and random Gaussian noise z_{noise} is added:

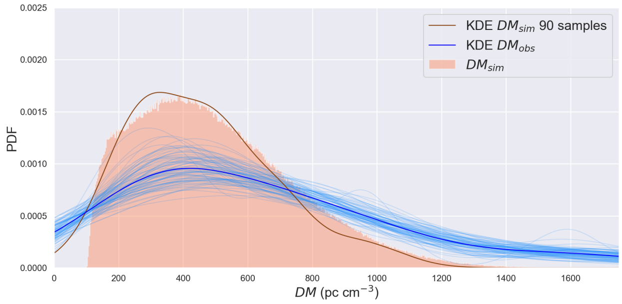
$$DM_{\text{cosmic}} = \langle DM_{\text{cosmic}}\rangle + \sigma z_{\text{noise}}\langle DM_{\text{cosmic}}\rangle. \tag{4.2}$$

The resultant PDF is added to those of $DM_{\text{MW}_{\text{halo}}}$ and DM_{host} to give the simulated density function of DM_{FRB} (Fig. 2).

[1] See the Appendix for an alternative, simulation-based approach that requires further investigation.
[2] The code and details on obtaining these parameters is available at: https://github.com/FRBs by J. Xavier Prochaska, Sunil Simha, Nicholas Tejos, Casey J. Law and others.

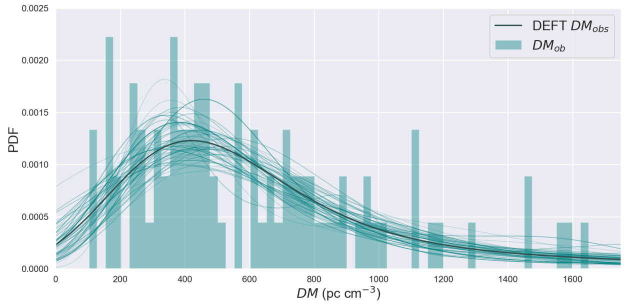


(a) Histogram and PDF ensemble for $DM_{\mathrm{FRB,obs}}$ using KDE.

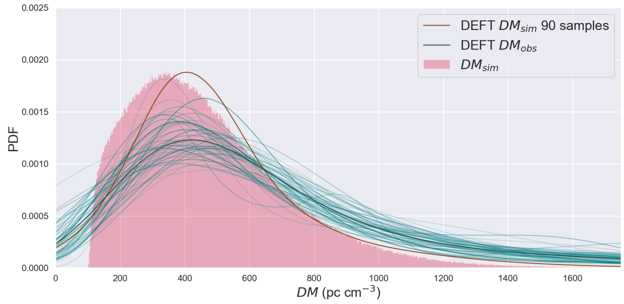


(b) Ensemble of $DM_{\mathrm{FRB,obs}}$ PDFs using KDE shown against the histogram of $DM_{\mathrm{FRB,sim}}$ and its KDE with sample size $n = n_{\mathrm{FRB}} = 90$.

Figure 3. KDE Ensembles.

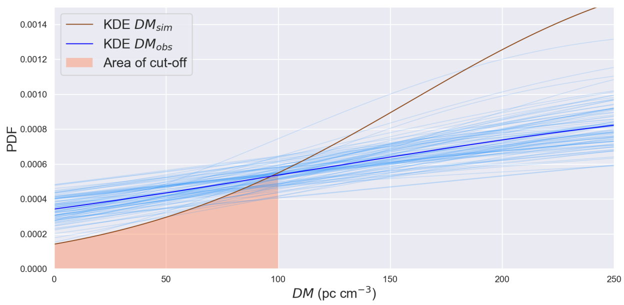


(a) Histogram and PDF ensemble for $DM_{\mathrm{FRB,obs}}$ using DEFT.

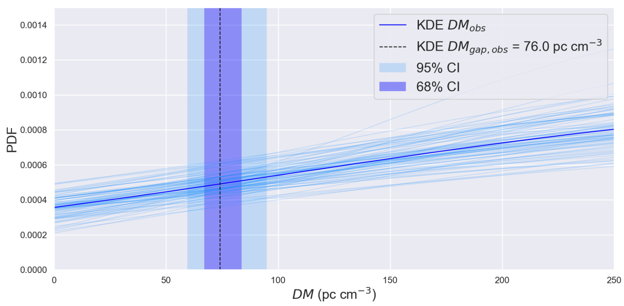


(b) Ensemble of $DM_{\mathrm{FRB,obs}}$ PDFs using DEFT shown against the histogram of $DM_{\mathrm{FRB,sim}}$ and its DEFT with sample size $n = n_{\mathrm{FRB}} = 90$.

Figure 5. DEFT Ensembles.

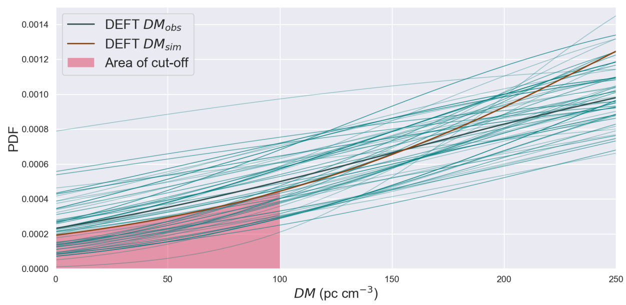


(a) Area below $DM_{\mathrm{gap,sim}} = 100$ cm^{-3}pc and ensemble of KDEs for $DM_{\mathrm{FRB,obs}}$.

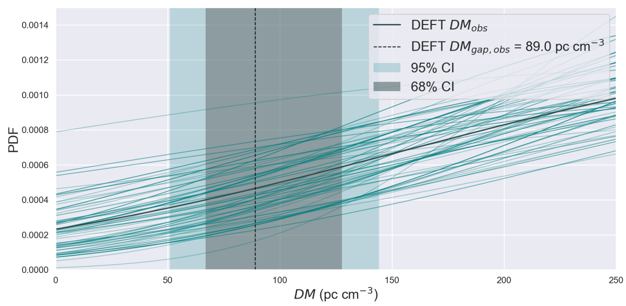


(b) $DM_{gap} = 76.0$ cm^{-3}pc with 68% and 95% CIs.

Figure 4. KDE Approximation of DM_{gap}.



(a) Area below simulated gap $DM_{gap} = 100$ and ensemble of PDFs for $DM_{\mathrm{FRB,obs}}$.



(b) $DM_{\mathrm{gap}} = 89.0$ cm^{-3}pc with 68% and 95% CIs.

Figure 6. DEFT Approximation of DM_{gap}.

4.2 The Observed Universe: KDE

A histogram is constructed from the observed data, from which a KDE of $DM_{\mathrm{FRB,obs}}$ is made. An ensemble of other plausible $DM_{\mathrm{FRB,obs}}$ PDFs is made via bootstrapping (Fig. 3a). A Gaussian kernel is used, and the bandwidth for each dataset is selected via cross-validation with 30 folds, giving bandwidths lying in the range $h = [150, 330]$. A KDE for the $DM_{\mathrm{FRB,obs}}$ density is also made, with a sample size $n = n_{\mathrm{obs}} = 90$ (Fig. 3b). A Gaussian kernel is used and the bandwidth is found to be $h = 177$.

The area below $DM_{\mathrm{gap,sim}} = 100\mathrm{pc\ cm}^{-3}$ is then calculated. $DM_{\mathrm{gap,obs}}$ is given by the point below which the $DM_{\mathrm{FRB,obs}}$ PDF area equals the area below $DM_{\mathrm{gap,sim}}$ (Fig. 4a). This is done for all realizations of the $DM_{\mathrm{FRB,obs}}$ PDFs, from which confidence intervals can be derived (Fig. 4b). At 68% confidence $DM_{\mathrm{FRB,obs}} = 76.0^{+11.0}_{-8.0}\mathrm{pc\ cm}^{-3}$, and at 95% $DM_{\mathrm{FRB,obs}} = 76.0^{+22.5}_{-15.3}\mathrm{pc\ cm}^{-3}$.

4.3 The Observed Universe: DEFT

A DEFT approximation is used to construct the best PDF estimate and an ensemble of plausible PDFs for $DM_{\mathrm{FRB,obs}}$. Recall from Section 3.2 that there is no optimization routine to select smoothness parameter α, and thus it must be manually chosen within the range $\alpha = [1, ..., 4]$ Chen et al. (2018). Depending on this choice, results are significantly different, but because there are only four possible values, it is possible to select the most appropriate value by eye. $\alpha = 2$ is chosen for $DM_{\mathrm{FRB,sim}}$, and $\alpha = 4$ for $DM_{\mathrm{FRB,obs}}$.

As before, the area of the $DM_{\mathrm{FRB,sim}}$ PDF below $DM_{\mathrm{gap,sim}}$ is used as a proxy to obtain $DM_{\mathrm{gap,obs}}$ and the confidence intervals. At 68% confidence $DM_{\mathrm{FRB,obs}} = 89.0^{+38.6}_{-22.2}\mathrm{pc\ cm}^{-3}$, and at 95% $DM_{\mathrm{FRB,obs}} = 89.0^{+55.2}_{-38.0}\mathrm{pc\ cm}^{-3}$.

4.4 The Impact of Sample Size

It is important to see how well the density estimations perform on small datasets, and to determine at what number of samples one can expect the density estimation to be a reasonable representation of the true PDF. As such, the simulated DM_{FRB} PDF is approximated by both methods using draws of $n = 100$, $n = 1000$ and $n = 2000$ (Fig. 7). While DEFT has a larger uncertainty region than KDE for $n = 100$, it tightens dramatically for $n = 1000$ and $n = 2000$. The DEFT approximation quickly converges towards DM_{gap} as n increases, and the front of the simulated density is reasonably well approximated. The KDE approximation, however, shows no improvement in this regard. The KDE becomes more smooth as n increases, likely over-smoothing and missing small structures in the data. The shape of the DEFT approximation appears to become more nuanced as n increases, however this could be a result of over-fitting.

5 DISCUSSION

One might note that the assumption that $DM_{\mathrm{FRB}}(z) \propto z_{FRB}$ is not a true reflection of the true DM_{FRB}—host galaxies and local environments are expected to play significant roles in the DM, and already FRBs have been observed in different galaxy types with different offsets from galaxy centres. As an example, FRB 121102 and FRB 180924 have different hosts (see Section 1.2), and although FRB 121102 is at a lower redshift ($z \approx 0.19$ vs $z \approx 0.32$), it has a higher DM ($DM \approx 557$ pc cm^{-3} vs $DM \approx 361$ pc cm^{-3}). The assumption, however, is that for low DM_{FRB} values, the shape of the density function is not too strongly influenced by the overall shape, and reasonable constraints can be placed on DM_{gap}.

In terms of manual intervention: for KDE, the optimization routine for the bandwidth means that only the kernel must be manually selected; for DEFT, one must choose the smoothness parameter. There are only a small number of options for each of these parameters, making the task doable.

When comparing the performance of the estimation methods as the sample size increases, DEFT stands out. As is stands, with $n_{\mathrm{obs}} = 90$, the confidence interval given by DEFT is far wider than that given by KDE. However it is shown that as the sample size increases, DEFT quickly outperforms KDE. The DEFT density function rapidly converges to a good approximation of the simulated density, while preserving small structures in the data. KDE on the other hand becomes increasingly smooth, suggesting that small structures in the data are lost. Especially notable is that the tail of the DEFT density function converges quickly towards the DM_{gap}, whereas the KDE tail remains roughly unchanged. As more FRB data comes in, DEFT is likely a better approach to take.

The constraints on DM_{gap} from KDE and DEFT are in agreement. Because the uncertainties are so large it is hard to draw any definite conclusions from this, however it is a promising start. As a further exercise, a subset of the observed data ($n = 60$) is analysed by both methods. If the metric used to determine the cut off value is valid, one expects the estimated DM_{gap} for the smaller dataset should be consistent with $DM_{\mathrm{gap,obs}}$. For KDE: at 68% confidence $DM_{\mathrm{FRB,obs}} = 103.0^{+34.4}_{-25.1}\mathrm{pc\ cm}^{-3}$, and at 95% $DM_{\mathrm{FRB,obs}} = 103.0^{+72.9}_{-38.6}\mathrm{pc\ cm}^{-3}$, which is in agreement with DM_{gap} given by the full FRB dataset. For DEFT: at 68% confidence $DM_{\mathrm{FRB,obs}} = 93.0^{+27.0}_{-31.2}\mathrm{pc\ cm}^{-3}$, and at 95% $DM_{\mathrm{FRB,obs}} = 93.0^{+50.0}_{-44.0}\mathrm{pc\ cm}^{-3}$. These results are in good agreement with the DM_{gap} values given by KDE and DEFT using the full FRB dataset. This is promising, however the uncertainties are large, and a more conclusive statement can only be made once more data becomes available. Figures can be found in the Appendix.

6 CONCLUSION AND OUTLOOK

Thousands of FRBs are expected to be observed in the near future, offering a great opportunity study the Galactic halo. In this work, we motivate density estimates as a tool to successfully constrain the region that divides galactic and extragalactic components of DM_{FRB}. With only a small increase in the number of samples, we find that DEFT in particular is capable of approximating a PDF reasonably well, especially near the DM_{gap}. As well as increased data, the simulated DM_{FRB} will become better informed over time, making the metric used in determining the cut-off value more reliable. Another important aspect that hasn't been considered yet

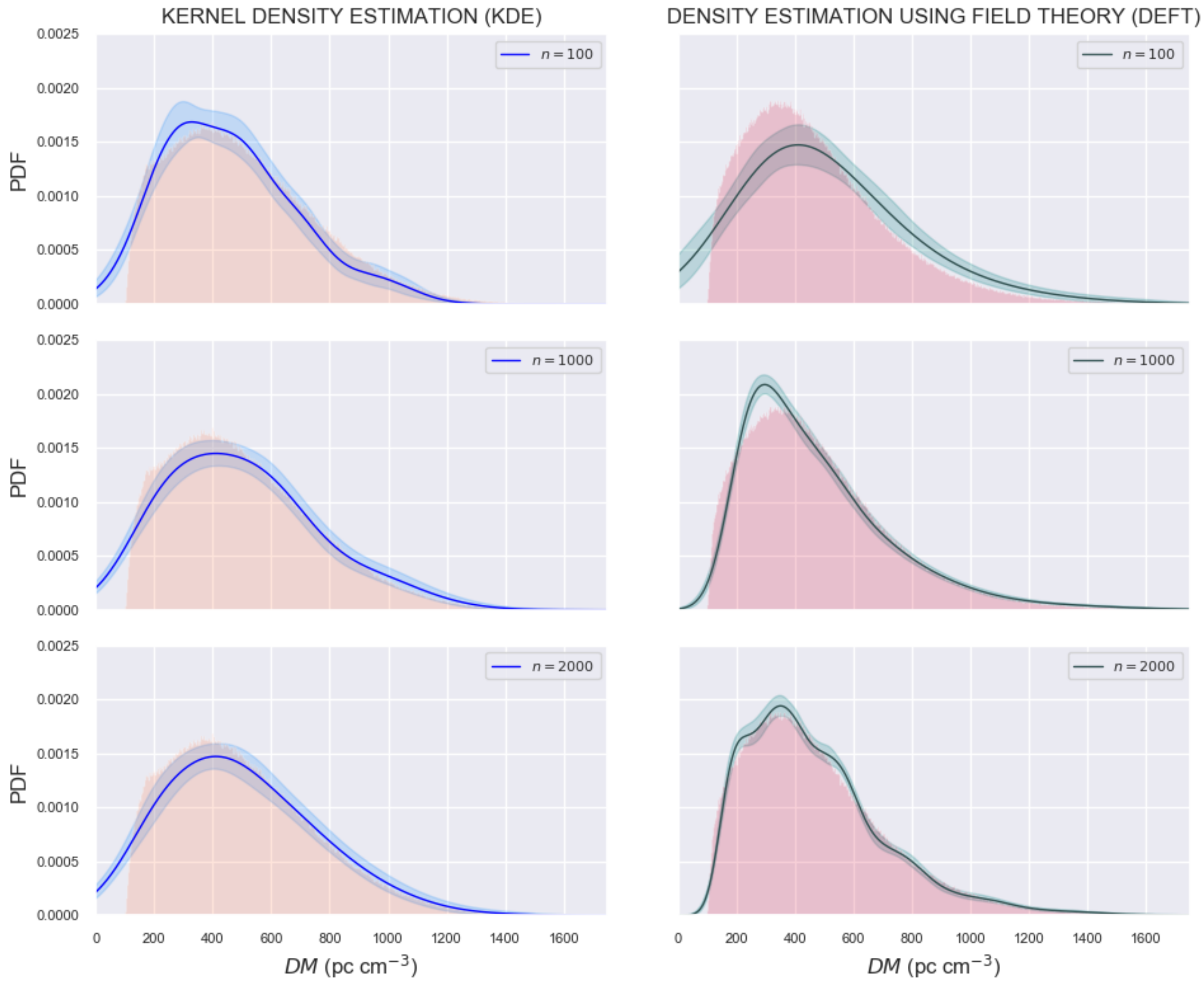


Figure 7. KDE and DEFT approximations of $DM_{\rm FRB,sim}$ with sample sizes $n = 100$, $n = 1000$ and $n = 2000$.

is the effect of boundary conditions on KDE. This requires follow-up: it could be that we have unwittingly placed restrictions that result in overly optimistic bounds for DM_{gap}.

7 ACKNOWLEDGEMENTS

This work was initiated as a project for the Kavli Summer Program in Astrophysics held at the University of California, Santa Cruz in 2019. The program was funded by the Kavli Foundation, The National Science Foundation, UC Santa Cruz, and the Simons Foundation. We thank them for their generous support. E. Platts would like to thank J. Xavier Prochaska for the invaluable training and support. E. Platts is supported by a PhD fellowship from the South African National Institute for Theoretical Physics (NITheP), and a grantholder bursary from the South African Research Chairs Initiative of the Department of Science and Technology (SARChI) and the National Research Foundation (NRF) of South Africa. Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

8 ACRONYMS

CGM circumgalactic medium. 1
CMB cosmic microwave background. 1

DEFT Density Estimation using Field Theory. 1, 3, 4, 6, 9
DM dispersion measure. 1, 2, 4, 6

FRB Fast Radio Burst. 1, 2, 4, 6, 9

IGM intergalactic medium. 1, 2, 4
ISM interstellar medium. 1, 2

KDE Kernel Density Estimation. 1, 3, 4, 6, 9
kNN k-nearest neighbours. 3

MAP Maximum *a posteriori*. 3, 4
MISE mean integrated squared error. 3

PDF probability density function. 1, 2, 3, 4, 6, 9

QSO quasistellar object. 1

RM rotation measure. 2

SZ Sunyaev-Zeldovich. 1

UV ultraviolet. 1

REFERENCES

Akahori T., Ryu D., Gaensler B. M., 2016, Astrophys. J., 824, 105
Andersen B. C., et al., 2019
Anderson M. E., Bregman J. N., Dai X., 2013, ApJ, 762, 106
Armillotta L., Fraternali F., Werk J. K., Prochaska J. X., Marinacci F., 2017, Monthly Notices of the Royal Astronomical Society, 470, 114
Bassa C. G., et al., 2017, Astrophys. J., 843, L8
Bergeron J., 1986, A&A, 155, L8
Bergeron J., Boissé P., 1991, A&A, 243, 344
Booth C. M., Schaye J., Delgado J. D., Dalla Vecchia C., 2012, Monthly Notices of the Royal Astronomical Society, 420, 1053
Burchett J. N., et al., 2019, Astrophys. J., 877, L20
Burles S., Tytler D., 1996, ApJ, 460, 584
Caleb M., et al., 2018, Mon. Not. Roy. Astron. Soc., 478, 2046
Champion D. J., et al., 2016, Mon. Not. Roy. Astron. Soc., 460, L30
Chatterjee S., et al., 2017, Nature, 541, 58
Chen H.-W., Lanzetta K. M., Webb J. K., 2001, ApJ, 556, 158
Chen W.-C., Tareen A., Kinney J. B., 2018, Phys. Rev. Lett., 121, 160605
Cooke R. J., Pettini M., Steidel C. C., 2018, Astrophys. J., 855, 102
Cordes J. M., Chatterjee S., 2019
Cordes J. M., Lazio T. J. W., 2002
Cordes J. M., Lazio T. J. W., 2003
Dai X., Bregman J. N., Kochanek C. S., Rasia E., 2010, The Astrophysical Journal, 719, 119
Deng W., Zhang B., 2014, Astrophys. J., 783, L35
Fang T., Buote D. A., Bullock J. S., Ma R., 2015, Astrophys. J. Suppl., 217, 21
Farah W., et al., 2018, Mon. Not. Roy. Astron. Soc., 478, 1209
Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518
Gajjar V., et al., 2018, Astrophys. J., 863, 2
Gao H., Li Z., Zhang B., 2014, Astrophys. J., 788, 189
Hani M. H., Ellison S. L., Sparre M., Grand R. J. J., Pakmor R., Gomez F. A., Springel V., 2019, Monthly Notices of the Royal Astronomical Society, 488, 135
Heitsch F., Putman M. E., 2009, The Astrophysical Journal, 698, 1485
Hill J. C., Baxter E. J., Lidz A., Greco J. P., Jain B., 2018, Phys. Rev. D, 97, 083501
Jones M., 1993, Stat Comput, 3
Katz J. I., 2016, Mod. Phys. Lett., A31, 1630013
Katz J. I., 2018, Prog. Part. Nucl. Phys., 103, 1
Keeney B. A., et al., 2017, The Astrophysical Journal Supplement Series, 230, 6
Kinney J. B., 2014, Phys. Rev. E, 90, 011301
Kinney J. B., 2015, Phys. Rev. E, 92, 032107
Kumar P., et al., 2019
Lanzetta K. M., Bowen D. V., Tytler D., Webb J. K., 1995, ApJ, 442, 538
Lehner N., O'Meara J. M., Fox A. J., Howk J. C., Prochaska J. X., Burns V., Armstrong A. A., 2014, ApJ, 788, 119
Lehner N., Howk J. C., Wakker B. P., 2015, ApJ, 804, 79
Li Z.-X., Gao H., Ding X.-H., Wang G.-J., Zhang B., 2018, Nature Commun., 9, 3833
Liang C. J., Chen H.-W., 2014, Mon. Not. Roy. Astron. Soc., 445, 2061
Lim S. H., Mo H. J., Li R., Liu Y., Ma Y.-Z., Wang H., Yang X., 2018, The Astrophysical Journal, 854, 181
Lorimer D. R., Bailes M., McLaughlin M. A., Narkevic D. J., Crawford F., 2007, Science, 318, 777
Macquart J. P., et al., 2015. (arXiv:1501.07535)
Masui K., et al., 2015, Nature, 528, 523
Michilli D., et al., 2018, Nature, 553, 182
Miralda-Escude J., Cen R.-y., Ostriker J. P., Rauch M., 1996, Astrophys. J., 471, 582
Muñoz J. B., Kovetz E. D., Dai L., Kamionkowski M., 2016, Phys. Rev. Lett., 117, 091301
Nicastro F., et al., 2018,] 10.1038/s41586-018-0204-1
O'Meara J. M., Tytler D., Kirkman D., Suzuki N., Prochaska J. X., Lubin D., Wolfe A. M., 2001, ApJ, 552, 718
Petroff E., 2017. (arXiv:1709.02189)
Petroff E., et al., 2015, Mon. Not. Roy. Astron. Soc., 447, 246
Petroff E., et al., 2016, PASA, 33, e045
Petroff E., Hessels J. W. T., Lorimer D. R., 2019, Astron. Astrophys. Rev., 27, 4
Planck Collaboration et al., 2013, A&A, 557, A52
Planck Collaboration et al., 2016, A&A, 594, A13
Platts E., Weltman A., Walters A., Tendulkar S. P., Gordin J. E. B., Kandhai S., 2018,] 10.1016/j.physrep.2019.06.003
Popov S. B., Postnov K. A., Pshirkov M. S., 2018, Phys. Usp., 61, 965
Prochaska J. X., Tumlinson J., 2009, in Proceedings, Astrophysics in the Next Decade : The James Webb Space Telescope and Concurrent Facilities: Tucson, Arizona, September 24–27, 2007. pp 419–456 (arXiv:0805.4635), doi:10.1007/978-1-4020-9457-6˙16
Prochaska J. X., Zheng Y., 2019a, Monthly Notices of the Royal Astronomical Society, 485, 648
Prochaska J. X., Zheng Y., 2019b, MNRAS, 485, 648
Prochaska J. X., Weiner B., Chen H.-W., Mulchaey J., Cooksey K., 2011, ApJ, 740, 91
Prochaska J. X., et al., 2017, ApJ, 837, 169
Putman M. E., Peek J. E. G., Joung M. R., 2012, ARA&A, 50, 491
Rauch M., 1998, Ann. Rev. Astron. Astrophys., 36, 267
Ravi V., 2019, Astrophys. J., 872, 88
Ravi V., Shannon R. M., Jameson A., 2015, Astrophys. J., 799, L5
Ravi V., et al., 2016, Science, 354, 1249

Ravi V., et al., 2019, Nature, 572, 352
Riihimaki J., Vehtari A., 2014, Bayesian Anal., 9, 425
Savage B. D., Narayanan A., Lehner N., Wakker B. P., 2011, The Astrophysical Journal, 731, 14
Shull J. M., Danforth C. W., 2018, Astrophys. J., 852, L11
Silverman B. W., 1986, Density Estimation for Statistics and Data Analysis. Chapman & Hall, London
Spergel D. N., et al., 2007, The Astrophysical Journal Supplement Series, 170, 377
Spitler L. G., et al., 2016, Nature, 531, 202
Spitler L. G., et al., 2018, Astrophys. J., 863, 150
Stocke J. T., Keeney B. A., Danforth C. W., Shull J. M., Froning C. S., Green J. C., Penton S. V., Savage B. D., 2013, The Astrophysical Journal, 763, 148
Tendulkar S. P., et al., 2017, Astrophys. J., 834, L7
Tripp T. M., Savage B. D., Jenkins E. B., 2000, ApJ, 534, L1
Tumlinson J., et al., 2013, ApJ, 777, 59
Tumlinson J., Peeples M. S., Werk J. K., 2017, Annual Review of Astronomy and Astrophysics, 55, 389
Walters A., Weltman A., Gaensler B. M., Ma Y.-Z., Witzemann A., 2018, Astrophys. J., 856, 65
Walters A., Ma Y.-Z., Sievers J., Weltman A., 2019
Wang Y. K., Wang F. Y., 2018, Astron. Astrophys., 614, A50
Wei J.-J., Wu X.-F., Gao H., 2018, Astrophys. J., 860, L7
Werk J. K., Prochaska J. X., Thom C., Tumlinson J., Tripp T. M., Meara J. M., Peeples M. S., 2013, The Astrophysical Journal Supplement Series, 204, 17
Werk J. K., et al., 2014, ApJ, 792, 8
White S. D. M., Rees M. J., 1978, Mon. Not. Roy. Astron. Soc., 183, 341
Yang Y.-P., Zhang B., 2016, Astrophys. J., 830, L31
Yu H., Wang F. Y., 2017, Astron. Astrophys., 606, A3
Zheng Z., Ofek E. O., Kulkarni S. R., Neill J. D., Juric M., 2014, Astrophys. J., 797, 71
Zhou B., Li X., Wang T., Fan Y.-Z., Wei D.-M., 2014, Phys. Rev., D89, 107303
Zitrin A., Eichler D., 2018, Astrophys. J., 866, 101

APPENDIX

FRB Distribution

We initially considered the star formation rate density as an approximation for the FRB rate density:

$$\psi(z) = K\left[0.015\frac{(1+z)^{2.7}}{1+[(1+z)/2.9]^{5.6}}\right] \tag{8.1}$$

The event rate density is converted to an event rate per distance bin $\left(\left[\text{events y}^{-1}\ \text{Mpc}^{-3}\right] \rightarrow \left[\text{events y}^{-1}\ \text{Mpc}\right]\right)$. The resultant PDF is compared to a histogram of the 7 known z values (Fig. 8). The star formation rate doesn't appear to reflect the data to date, so we chose to proceed with the PDFs of observed z values.

A Consistency Check

Here KDE and DEFT are used on a subset of the observed data, $n = 60$ out of $n = 90$. For KDE: at 68% confidence $DM_{\rm FRB,obs} = 103.0^{+34.4}_{-25.1}$pc cm^{-3}, and at 95% $DM_{\rm FRB,obs} = 103.0^{+72.9}_{-38.6}$pc cm^{-3}. For DEFT: at 68% confidence $DM_{\rm FRB,obs} = 93.0^{+27.0}_{-31.2}$pc cm^{-3}, and at 95% $DM_{\rm FRB,obs} = 93.0^{+50.0}_{-44.0}$pc cm^{-3}. These results are in good agreement with the $DM_{\rm gap}$ values obtained with the full FRB dataset.

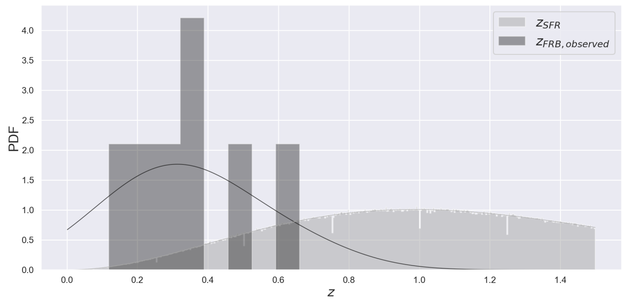


Figure 8. PDFs of FRB redshift distributions simulated using the star formation rate.

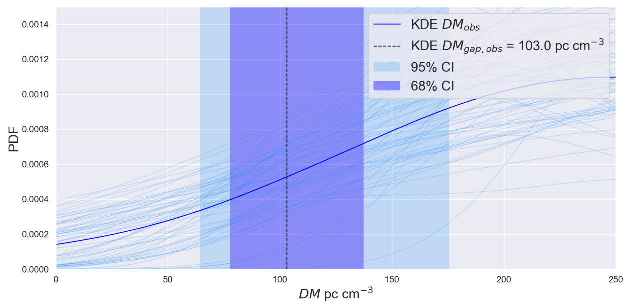


Figure 9. KDE with a subsample of observed data with $n = 60$. $DM_{\rm gap} = 103.0$ cm^{-3}pc with 68% and 95% CIs.

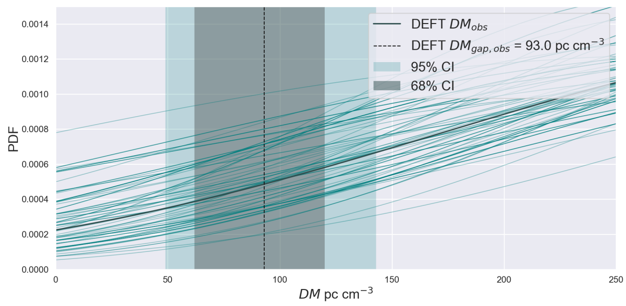


Figure 10. DEFT with a subsample of observed data with $n = 60$. $DM_{\rm gap} = 93.0$ cm^{-3}pc with 68% and 95% CIs.