Improvement of cosmological constraints with the cross correlation between line-of-sight optical galaxy and FRB dispersion measure

Chenghao Zhu
Physics Department, Fudan University, Shanghai 200433, China

Jiajun Zhang
Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China
(Dated: May 10, 2022)

Fast Radio Bursts (hereafter FRBs) can be used in cosmology by studying the Dispersion Measure (hereafter DM) as a function of redshift. The large scale structure of matter distribution is regarded as a major error budget for such application. Using optical galaxy and dispersion measure mocks built from N-body simulations, we have shown that the galaxy number density can be used as a tracer for large scale electron density and help improve the measurement of DM as a function of redshift. We have shown that, using the line-of-sight galaxy number counts within 1’ around the given localized FRB source can help improve the cosmological parameter constraints by more than 20%.

I. INTRODUCTION

Currently, ΛCDM model is considered as the standard model of cosmology [1]. It is supported by varieties of observations, including cosmic microwave background temperature anisotropy measurement from Planck [2], baryon acoustic oscillations data from SDSS galaxy surveys [3], type Ia supernovae data from the Joint Light-curve Analysis [4], and $H_0$ data from distance ladder observation [5] [6]. The constrains of cosmological parameters are quite precise today, but still lots of attempts are made to verify the consistency from different methods. There is some tension in the measurement of $H_0$ [7][10]. Therefore, we demand new methods probing the cosmic expansion history, such as using standard siren, 21cm intensity mapping, FRBs [11][12].

Fast Radio bursts are short-duration (~ms) radio bursts (~GHz) which was first detected in 2007 [15]. FRBs seems to be very common in our universe [16]. From their isotropic dispersion measures, most FRBs have been confirmed to have extragalactic origins [17]. FRB200428 is the first FRB which is located inside the Milky Way [18]. The astrophysical mechanism of FRBs is still mysterious, but more and more information about their source have been reported [18] [19]. Several good reviews have summarized current theories about FRBs and its physical mechanism [20][22]. Moreover, hundreds of FRBs are gradually observed nowadays. Their information are available online [1]. Many radio telescopes such as CHIME [23], PARKES [24], FAST [25], BINGO [26], SKA [27] have detected considerable FRBs or can be expected to observe much more FRBs in the near future.

Since the first known repeater FRB121102 [19] was detected and located, with or without a repeating source, several FRBs are well located as well [28][29]. With their extragalactic location known, FRBs show great potential to be new independent probes of cosmology [30][31]. The dispersion measure (hereafter DM) of an FRB contains the information about ionized baryons of the path from the source to earth. The expectation of DM contributed by the intergalactic medium from an FRB at redshift $z$ is highly related to the cosmological model and parameters. Hence the FRBs with their DM($z$) can be used to constrain cosmology, although it is still hard to separate DM from its host galaxy’s contribution.

Previously, Macquart et al. [32] has tried to do parameter constrain from known localized FRBs in the universe. Combined with constrains from other observations [2] as their prior, Walters et al. [33] explore the constrains with more FRBs generated by simulation. Futher, Walters et al. [34] discussed the result with a weaker assumption of knowledge about diffuse gas fraction.

From their large dispersion measure, naturally we can associate FRBs with large scale structure. Rafiei-Ravandi et al. [35] has summarized the angular cross-correlation between DM and large-scale structure of galaxy with the results released by CHIME/FRB catalogue [23]. In addition, Shirasaki et al. [36] investigated the angular cross-correlation in detail with hydrodynamical simulation illustrisTNG. Following their inspiration, we attempted to use such a cross-correlation to partly eliminated the variance of DM contributed by the large-scale structure, in order to obtain a better cosmological constraints. In this work, we will verify the improvement of cosmological parameter constraints using FRBs with the knowledge of line-of-sight galaxy number.

In this paper, we will show our results as follows: in section II we will introduce some basic background knowledge, including the definition of DM and how to calculate it in different coordinate; in section III we will introduce the method how we deal with the N-body simulation data; in section IV we will show some noteworthy results from mocks; in section V we will show the concrete MCMC improvement after applying the cross-correlation we acquire; finally in section VI a brief sum-
The observed delay time is \( \Delta t \) for electrons along the path. But in the observer frame, the DM frequency is \( \nu \) traveling through the plasma, which is determined by the dispersion effect of plasma:

\[
\Delta t = \frac{e^2}{2\pi m_e c} \left( \frac{1}{v_1^2} - \frac{1}{v_2^2} \right) \int \frac{n_e}{1+z} \, dz \\
\simeq 4.15 s \left( \frac{\nu_1}{1 \text{GHz}} \right)^{-2} - \left( \frac{\nu_2}{1 \text{GHz}} \right)^{-2} \frac{DM}{10^3 \text{pc} \cdot \text{cm}^{-3}}.
\]

It is noteworthy that, in the rest frame, DM is defined as \( DM' = \int n_e(z) \, dl \) which is the column density of free electrons along the path. But in the observer frame, the observed delay time is \( \Delta t' = \Delta t' \cdot (1+z) \) and the observed frequency is \( \nu' = \nu'/(1+z) \). Consequently, the measured DM by an earth observer is modified as \( DM = \int \frac{n_e(z)}{1+z} \, dl \).

The total DM of an FRB can be separated as

\[
DM = DM_{MW} + DM_{IGM} + DM_{HG} \frac{1}{1+z}.
\]

where we denote Milky Way, intergalactic medium and host galaxy by MW, IGM and HG. In actual observation, contribution of Milky Way can be obtained from research about pulsars, though not very accurately. Usually we cannot get adequate information of host galaxy of an FRB, and the contribution of DM from the host galaxy is very likely unknown. While the contribution of IGM, which is highly associated with large scale structure, can provide information about cosmology.

\section{II. PRELIMINARIES}

In general, DM represents the delay time of a signal traveling through the plasma, which is determined by the dispersion effect of the plasma:

\[
\Delta t = \frac{e^2}{2\pi m_e c} \left( \frac{1}{v_1^2} - \frac{1}{v_2^2} \right) \int \frac{n_e}{1+z} \, dz \\
\simeq 4.15 s \left( \frac{\nu_1}{1 \text{GHz}} \right)^{-2} - \left( \frac{\nu_2}{1 \text{GHz}} \right)^{-2} \frac{DM}{10^3 \text{pc} \cdot \text{cm}^{-3}}.
\]

The mean DM contributed by IGM can simply derived from its definition. In this article, we apply \( \Lambda \)CDM cosmology. Then, in the comoving ordinate, the average number density of electron at redshift \( z \) can be expressed as

\[
\bar{n}_e = \rho_{e,0} (1+z)^3 \Omega_b \times \left( \frac{3}{4} \frac{I_{H}(z)}{m_p} + \frac{1}{4} \frac{I_{He}(z)}{4m_p} \right),
\]

where \( \rho_{e,0} = \frac{3H_0^2}{8\pi G} \) is the critical density of the universe today, \( \Omega_b \) is fraction of baryon, \( I_{H}(z) \) is the average number of ionized electron in the IGM emitted by one hydrogen atom, and \( I_{He} \) denotes the contribution of helium.

The relation between physical distance interval and redshift \( z \) is given by

\[
dl = \frac{1}{1+z} c H(z) dz \\
= \frac{1}{1+z} \frac{c}{H_0} \frac{dz}{\sqrt{(1+z)^3 \Omega_M + \Omega\Lambda}}.
\]

Here we use \( \Lambda \)CDM model again and assumed flat universe, thus \( \Omega\Lambda = 1 - \Omega_M \). Plugging back into the definition of DM, we finally obtain

\[
\overline{DM}_{IGM} = \frac{3c\Omega_b H_0}{8\pi G m_p} \int_0^z (1+z') \frac{I(z')}{E(z')} \, dz',
\]

where \( I(z') = \frac{3}{4} I_{H}(z) + \frac{1}{4} I_{He}(z) \) and \( E(z') = \sqrt{(1+z)^3 \Omega_M + \Omega\Lambda} \).

\section{III. METHOD}

MDR1 is the first MultiDark simulation, performed in 2010\cite{38}. This simulation contains about 8.6 billion particles in a \( (1 \text{Gpc}/h)^3 \) cube, using WMAP5 cosmology. MDR1 provides the matter density field data and the halo catalogue. In our research, we need to calculate the dispersion measure between arbitrary two points in the

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig1.png}
\caption{Without loss of generality, we display a 2-dimensional case for its convenience to illustrate. Here we set \( A_0 \) to be our ‘observer’ and \( A_3 \) to be the FRBs source. \( \text{dens } i \ (i=1,2,3,4) \) are the density given by MDR1 database, \( A_1 \) and \( A_2 \) are intersection points of path with the grids.}
\end{figure}
FIG. 2. From left to right, we plot the mollweide projection images from the observer to z=0.1, 0.5 and 1.4. From up to bottom, we plot the mollweide projection images of DM, galaxy number with "cone" method and galaxy number with "cone+" method. Notice that we take logarithm for illustrating $n_{\text{gal}}$, i.e. the value of each pixel is $\ln(1 + n_{\text{gal}})$ for both 'cone' and 'cone+'.

box and the galaxy number in an arbitrary light-cone (or weighted light-cone which we will introduce later).

The side length of the cosmological cube is 1 Gpc/h, which is still too small to cover the redshift range of FRBs. We have used the periodic boundary condition of the simulation to construct a much larger box to cover all the FRBs we have considered, up to $z = 2$.

MDR1 only gives the data of over-density in each cell (for convenience we will use density instead, and denote it by $\text{dens} = \rho_{\text{background}} \times (1 + \text{overdensity})$). To calculate DM, we need to obtain number density of free electron in each cell. For simplicity, we have set the distribution of baryon to be the same as dark matter. We ignored the difference in metallicity and ionization fraction, assumed that $I_H = 1$ and $I_{He} = 2$. Then expression of DM becomes

$$DM = \int_0^\chi \text{dens}(\chi') \times \frac{7}{8} \frac{\Omega_b}{\Omega_m} d\chi'.$$

(7)

As shown in Fig.1, we have illustrated how to calculated DM between arbitrary two points from the density fields,

$$DM(A_0A_3) = \int_{A_0}^{A_3} n_e(\chi')(1+z) d\chi'$$

$$= \text{dens}_3 \int_{A_0}^{A_1} (1+z) d\chi' + \text{dens}_1 \int_{A_1}^{A_2} (1+z) d\chi'$$

$$+ \text{dens}_2 \int_{A_2}^{A_3} (1+z) d\chi'.$$

(8)

We use halo occupation distribution[39] to transform halo catalogue into galaxy number. We choose Favole’s model[40] to obtain galaxy number of each halo,

$$N_{\text{cen}}(M) = \frac{1}{2} [1 + erf(\log M - \log M_{\text{min}})]\frac{\sigma_{\log M}}{\sigma_{\log M}}$$

$$N_{\text{sat}}(M) = N_{\text{cen}}(M) \left(\frac{M - M_0}{M_1}ight)\alpha,$$

(9)

where $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the error function. In our research, we set up 5 parameters of Halo occupation distribution as follows: $M_{\text{min}} = 10^{12.681}$, $M_0 = 10^{12.296}$, $M_1 = 10^{13.635}$, $\sigma_M = 0.532$ and $\alpha = 0.994$. Then the average galaxy number is $\bar{n}_{\text{gal}} = 0.0012/(\text{Mpc}/h)^3$.

IV. MOCK DATA

Since in the simulation, DM can be directly calculated, we randomly choose a point in the box to be our observer,
and obtain the DM from all angles of the full sky to different redshift. Following the method illustrated above, we get 20 global maps of DM in total from \( z = 0.1 \) to \( z = 2 \) equidistantly with corresponding maps of galaxy number (hereafter \( n_{\text{gal}} \)). With these data, we can give a detailed discussion about the correlation to prepare for further analysis.

### A. Line-of-sight Galaxy Number

We use two different methods to obtain \( n_{\text{gal}} \). In the first way which we denote by 'cone', we sum those galaxies whose halo center is contained in our light-cone with half top angle \( \theta \). In the second way denoted by 'cone+', we count all galaxies whose position vector's projection on the axis pointing from observer to source is positive, but with a factor exponentially decay about its perpendicular distance to the axis, that is, \( \exp[-(|v| \sin \theta)/(v_z \tan \theta)] \), where \( v \) is the position vector, \( v_z \) is its projection to the axis and \( \theta \) is the half top angle of the 'cone+'.

In Fig.??, we have shown the Mollweide projection map of DM, "cone" galaxy number and "cone+" galaxy number from the observer to \( z = 0.1, 0.5 \) and 1.4. We can clearly see that the DM map strongly correlate with the galaxy number distribution, which is what we expected. At high redshift such as \( z = 1.4 \), it is hard to identify correlation between DM and galaxy number from the map. We will show quantitative results of cross correlation between DM and galaxy number in the next subsection.

### B. Cross Correlation

The value of \( n_{\text{gal}} \) of each pixel depend on the line-of-sight angle \( \theta \) we set, and so does the correlation between DM and \( n_{\text{gal}} \). Hence we have tried with different angles at redshift \( z=0.1, 0.5, 1.4 \), which we think can reflect the behavior in the whole range we want to investigate.

The cross correlation coefficient between DM and \( n_{\text{gal}} \) for different \( \theta \) are shown in Fig.?? We are looking for an optimal number of \( \theta \), using which can provide us good cross correlation between DM and \( n_{\text{gal}} \). We can see that considering different redshift, \( \theta = 0.0027 \) is an optimal choice and for simplicity we choose half top-angle for both 'cone' and 'cone+' to be \( \theta = 0.003 \) (radian) which is approximately equal to 1' in degree. The mollweide projection images are shown in the Fig.?? At low redshift, we will encounter many pixels whose value is 0, in the mollview image of cone. Such a problem can be solved by using 'cone+'. Hence the cross-correlation between DM and \( n_{\text{gal}} \) will be better if we use 'cone+' at low redshift. But at high redshift, the little difference between 'cone' and 'cone+' will not make significant influence on the cross-correlation. The comparison is shown in Fig.?? We have compared the cross correlation coefficient between DM and \( n_{\text{gal}} \) using 'cone' and 'cone+', and we have also shown the redshift evolution of the cross correlation using 'cone'. It is clear that at higher redshift, the correlation is better and the difference between 'cone' and 'cone+' is smaller. Because 'cone+' require more complicated calculation for the galaxy distribution and considering that we are more interested in high redshift FRBs and their DM, we will use the 'cone' method to do further analysis about DM.
FIG. 5. The $20 \times 20$ matrix of Pearson product-moment correlation coefficients obtained from the data is shown on the left, and the right one is the matrix we obtained from fitting function, which is smooth.

FIG. 6. Here we show the MCMC result with using 500 FRBs without correction. The blue line marks the true value of parameters. The true value lies in the $1\sigma$ area.

V. MCMC CONSTRAIN

We will follow the formula Eq.6 to do MCMC analysis with several FRBs randomly obtained in the simulation. In our research, we will not assume any prior knowledge about the parameters; hence all the parameters will follow uniform distribution in their well-defined interval ($\Omega_b, \Omega_m \sim u(0,1)$). We randomly set points in the simulation box to be FRBs source and use the same method to calculate their DM. Then we will try to find a method to associate the DM with $n_{gal}$, and finally compare the results.

We have obtained a $20 \times 20$ covariance matrix from the 20 mocks built from the simulations at 20 redshifts, to acquire the covariance matrix of FRBs at arbitrary redshift, We use them to fit a function of the form

$$R(z_i, z_j) = b(z)e^{-\frac{\Delta z}{a(z)}} + (1 - b(z))$$

where $R(z_i, z_j)$ is the $(i-j)^{th}$ term of Pearson product-moment correlation coefficients; $\Delta z = z_j - z_i$, $(z_j > z_i)$ is their redshift difference; $a(z), b(z)$ are our fitting parameters. We then use a third-order polynomial $p(z) = p_3 z^3 + p_2 z^2 + p_1 z + p_0$ to fit $a(z)$ and $b(z)$, the results are shown in table I. In Fig[5] We also plot a smoother image of Pearson product-moment coefficient matrix compared to original one with our fitting function, in order to verify whether it is appropriate. The result shows it is applicable within the range of our research. The MCMC result from 500 FRBs with the location of true value is shown in the Fig.6. The true values well lie in the $1 - \sigma$ area, so that we use the covariance matrix and the related method to do further study.

From Eq[6] we can see that the cosmological parameters that we can put constraints using $DM(z)$ is $\Omega_M$, $\Omega_b$ and $H_0$, under the flat space assumption. However $\Omega_b$ and $H_0$ are degenerate, the only independent cosmological parameters are $\Omega_M$ and $\Omega_b h$. Therefore, it is sufficient to perform MCMC fitting for these two sets of parameters. In Fig[6] we have shown that using the covariance matrix fitting function and the randomly generated FRBs, we can put constraints on the cosmological parameters and the true value well lie in the $1 - \sigma$ range. In the error range contour, it is well known that the large scale structure plays an important role in the error budget of the function $DM(z)$. With the information of line-of-sight galaxy number, we can have a better estimation of the large scale structure of matter distribution, thus provide a smaller error.

**TABLE I. Fit Parameters**

|     | $p_0$ | $p_1$ | $p_2$ | $p_3$ |
|-----|-------|-------|-------|-------|
| $a(z)$ | 0.122 | 1.505 | -1.300 | 0.500 |
| $b(z)$ | 0.857 | -0.494 | 0.192 | -0.014 |
A. Correction Method

According to the value of $n_{\text{gal}}$, we divided the 196608 pixels into 16 groups equally and calculate the standard deviation and the average of each group. Their distribution is shown in Fig.7. Then we figure out that the relation between $n_{\text{gal}}$ and DM is almost linear when $z \geq 0.3$; but if $z$ is too small, there will be too many pixels with its $n_{\text{gal}}$ is 0. With a determined halo occupation distribution model, the mean $n_{\text{gal}}$ is known. We use these 16 groups of points to fit a function of form separately at different redshift,

$$DM(n_{\text{gal}}, z) = k(z)(n_{\text{gal}} - \bar{n}_{\text{gal}}) + b(z)$$  (11)

As we expect, $b(z)$ is equal to $DM_{\text{mean}}(z)$ within 4% error, and the error decreases to 0.4% as $z$ increases to 2.

Strictly speaking, $k(z)$ is not only dependant on the $\theta$ of the light-cone but also related to the halo occupation model we established. Hence here we will only use interpolation method to obtain $k(z)$ at FRBs’ redshift, and leave further research for our future study.

With known $n_{\text{gal}}$ and DM associated with the same FRB located at $z > 0.2$, we can correct our expectation about corresponding $DM_{\text{mean}}$ by using Eq.10.

$$DM_{\text{cor}} = DM_{frb} - k(z)(n_{\text{gal}} - \bar{n}_{\text{gal}})$$  (12)

where $DM_{frb}$ is the primitive DM, and $DM_{cor}$ is what we will use in MCMC analysis after correction.

B. Result Improvement

With the information of line-of-sight galaxy number provided, it is expected that we can have a better estimation of DM, partially get rid of the error introduced by large scale structure. As the results shown in Fig.8 suggest, our correction is effective, the DM of FRBs distributes closer to the theoretical expectation compared with original data. More practically, if we apply our correction method to put constraints on the cosmological parameters given 100 random FRBs or 500 random FRBs, the improvement of constraints can be clearly seen in Fig.9. The blue contours show the parameter uncertainty range after the correction and the red contours show the range before the correction, which is what we normally see in literature. The detail number of the cosmological parameters are summarised in table II. We shall notice that such improvement is free. Since if we can localize the FRBs and get their redshift, these FRBs are localized by matching the radio signal location to the given galaxy catalog. We naturally know the number of galaxies near the host galaxies of the FRBs, which can be easily obtained. The improvement of the precision of cosmological constraints is more than 20%.

VI. DISCUSSION

With more and more FRBs being discovered and localized, FRB cosmology has attracted quite some attention in the cosmology study. Using dispersion measure

| $10^3 \Omega_{\text{h}}$ | $10\Omega_m$ |
|----------------|-------------|
| True | 3.283 | 2.7 |
| 100-before | 3.7606$^{0.6367}_{-0.6004}$ | 4.58$^{+2.79}_{-2.01}$ |
| 100-after | 3.5468$^{0.5506}_{-0.4570}$ | 3.52$^{+2.15}_{-1.36}$ |
| 500-before | 3.2030$^{0.6572}_{-0.5537}$ | 2.19$^{+1.29}_{-1.22}$ |
| 500-after | 3.0746$^{0.4515}_{-0.3644}$ | 2.06$^{+1.36}_{-0.89}$ |

a we quote the uncertainties based on the 16th, 50th, and 84th percentiles of the samples, and in our case, it is a little different from the maximum likelihood point.
FIG. 9. On the top, we show the improvement of cosmological constraints using 100 FRBs. On the bottom, we show the improvement using 500 FRBs. We can see clear improvement in the cosmological parameters constraints for both cases.

to measure the expansion history of the universe is a unique method. If we have better understanding of the error budget in the dispersion measure of FRBs, we will have better chance to put tight constraints on cosmological parameters.

In this paper, we have proven the idea that, using the information of large scale structure traced by galaxies, we may put tighter constraints with FRBs, really works. We have shown that just use the galaxy number within 1’ around the FRBs host galaxies, we can correct the DM(z) expectation value optimally. Afterwards, we may have more than 20% improvement in the cosmological constraining precision.

We have covered almost all the steps for cosmological analysis using FRBs in this paper, but there are quite some details need to be improved so that our method can be really applied to the real observation. For example, how do the DM-$n_{gal}$ relation rely on the galaxy mean number density, what is the parameters of this relation for some real observational data, is the simple model of ionized gas sufficient for our analysis here, how much improvement will we have if we consider other error budgets etc. We will cover these questions in our future study. Beyond cosmological application, the cross correlation between DM and galaxy distribution can also tell us more about the gas environment around galaxies and the properties of IGM. We hope our study here can provide useful hints for the other study in the future.

ACKNOWLEDGMENTS

J.Z acknowledges support from the Ministry of Science and Technology of China (grant Nos. 2020SKA0110102). We thank Chuhan Jiang, Yao Zhang and Fayin Wang for useful discussion. The CosmoSim database used in this paper is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP). The MultiDark database was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064. We use the MDR1[1] database with its FOF[2] halo catalogue as the simulated universe. We use python packages *emcee*[3] for MCMC analysis; use *getdist*[4] and *corner*[5] for plotting.

[1] P. Bull, Y. Akrami, J. Adamek, T. Baker, E. Bellini, J. B. Jimenez, E. Bentivegna, S. Camera, S. Clesse, J. H. Davis, et al., Physics of the Dark Universe **12**, 56 (2016).
[2] P. Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, et al. (2020).
[3] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, Monthly Notices of the Royal Astronomical Society **449**, 835 (2015).
[4] M. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin,
References