

# A census of baryons in the Universe from localized fast radio bursts

<https://doi.org/10.1038/s41586-020-2300-2>

Received: 27 September 2019

Accepted: 13 March 2020

Q1



Check for updates

J.-P. Macquart<sup>1✉</sup>, J. X. Prochaska<sup>2,3✉</sup>, M. McQuinn<sup>4</sup>, K. W. Bannister<sup>5</sup>, S. Bhandari<sup>5</sup>, C. K. Day<sup>5,6</sup>, A. T. Deller<sup>6</sup>, R. D. Ekers<sup>1,5</sup>, C. W. James<sup>1</sup>, L. Marnoch<sup>5,7</sup>, S. Osłowski<sup>6</sup>, C. Phillips<sup>5</sup>, S. D. Ryder<sup>7</sup>, D. R. Scott<sup>1</sup>, R. M. Shannon<sup>6</sup> & N. Tejos<sup>8</sup>

More than three-quarters of the baryonic content of the Universe resides in a highly diffuse state that is difficult to detect, with only a small fraction directly observed in galaxies and galaxy clusters<sup>1,2</sup>. Censuses of the nearby Universe have used absorption line spectroscopy<sup>3,4</sup> to observe the ‘invisible’ baryons, but these measurements rely on large and uncertain corrections and are insensitive to most of the Universe’s volume and probably most of its mass. In particular, quasar spectroscopy is sensitive either to the very small amounts of hydrogen that exist in the atomic state, or to highly ionized and enriched gas<sup>4–6</sup> in denser regions near galaxies<sup>7</sup>. Other techniques to observe these invisible baryons also have limitations; Sunyaev–Zel’dovich analyses<sup>8,9</sup> can provide evidence from gas within filamentary structures, and studies of X-ray emission are most sensitive to gas near galaxy clusters<sup>9,10</sup>. Here we report a measurement of the baryon content of the Universe using the dispersion of a sample of localized fast radio bursts; this technique determines the electron column density along each line of sight and accounts for every ionized baryon<sup>11–13</sup>. We augment the sample of reported arcsecond-localized<sup>14–18</sup> fast radio bursts with four new localizations in host galaxies that have measured redshifts of 0.291, 0.118, 0.378 and 0.522. This completes a sample sufficiently large to account for dispersion variations along the lines of sight and in the host-galaxy environments<sup>11</sup>, and we derive a cosmic baryon density of  $\Omega_b = 0.051^{+0.021}_{-0.025} (H_0/70 \text{ km s}^{-1} \text{ Mpc}^{-1})^{-1} (\pm 95 \text{ per cent confidence; } H_0 \text{ is Hubble’s constant})$ . This independent measurement is consistent with values derived from the cosmic microwave background and from Big Bang nucleosynthesis<sup>19,20</sup>.

Q2

The Commensal Real-time ASKAP Fast Transients (CRAFT) survey on the Australian Square Kilometre Array Pathfinder (ASKAP) has commissioned a mode capable of localizing fast radio bursts (FRBs) with subarcsecond accuracy, thus enabling identification of their host galaxies and measurement of their redshifts  $z$ . ASKAP consists of 36 antennas equipped with phased array feeds, able to view 30 degrees<sup>2</sup> on the sky. Bursts are detected by incoherently summing the total power signal of individual beams from each of the antennas. Bursts detected in the incoherent pipeline are subsequently localized interferometrically by triggering a download of voltage data from a 3.1-s-duration ring buffer that is correlated and imaged at high time resolution to provide the localizations<sup>15,16</sup>. The 6-km baselines of the array yield statistical position errors of approximately  $10''(S/N)^{-1}$ , where the final coherent signal-to-noise ratio of the burst,  $S/N$ , exceeds 50 for any burst whose signal-to-noise ratio in the incoherent pipeline is greater than 9. The resulting statistical (thermal) uncertainties are smaller than  $0.2''$ . Systematic

errors in these positions are typically smaller than  $0.5''$ . At  $z = 0.5$ ,  $1''$  corresponds to 5 kpc which is approximately the precision needed to associate an FRB with its host galaxy while reducing the chance coincidence probability to  $<1\%$  (ref. <sup>16</sup>).

We report the detection of four localized ASKAP bursts. Table 1 lists the burst properties, sky positions and host galaxy offsets, while Fig. 1 shows the host galaxy identifications (see also ref. <sup>21</sup> and Methods). Their dispersion measures (DMs) are well in excess of the  $30\text{--}100 \text{ pc cm}^{-3}$  contributions expected from the disk and halo of the Milky Way at high Galactic latitudes<sup>13,22</sup>, with the large excesses attributable to the intergalactic medium (IGM) and gas within each burst host galaxy. Two other ASKAP-detected bursts and their host galaxies were reported previously<sup>15,16</sup> in addition to three other host-galaxy identifications<sup>14,17,18</sup>.

The precise localization of a set of FRBs to their host galaxies provides the first ensemble of  $\text{DM}_{\text{FRB}}$  and  $z_{\text{FRB}}$  measurements. The  $\text{DM}_{\text{FRB}}$  measurement represents the electron density weighted by  $(1+z)^{-1}$  integrating

Q3

Q4

Q5

Q6

<sup>1</sup>International Centre for Radio Astronomy Research, Curtin Institute of Radio Astronomy, Curtin University, Perth, Western Australia, Australia. <sup>2</sup>University of California Observatories–Lick Observatory, University of California, Santa Cruz, CA, USA. <sup>3</sup>Kavli Institute for the Physics and Mathematics of the Universe, Kashiwa, Japan. <sup>4</sup>Astronomy Department, University of Washington, Seattle, WA, USA. <sup>5</sup>Commonwealth Science and Industrial Research Organisation, Australia Telescope National Facility, Epping, New South Wales, Australia. <sup>6</sup>Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria, Australia. <sup>7</sup>Department of Physics and Astronomy, Macquarie University, North Ryde, New South Wales, Australia. <sup>8</sup>Instituto de Física, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile. ✉e-mail: J.Macquart@curtin.edu.au; xavier@ucolick.org

**Table 1 | Properties of FRBs interferometrically localized with ASKAP**

FRB	Time (UTC <sup>a</sup> )	DM (pc cm <sup>-3</sup> )	Rotation measure (rad m <sup>-2</sup> )	Fluence (Jy ms) <sup>b</sup>	RA (h:min:s) <sup>c</sup>	Dec. (degree:arcmin:arcsec) <sup>c</sup>	Host redshift	Offset from host nucleus (kpc)
180924	16:23:12.6265	361.42(6)	14(1)	16(1)	21:44:25.255 ± 0.006 ± 0.008	-40:54:00.10 ± 0.07 ± 0.09	0.3214	3.5 ± 0.9
181112	17:31:15.48365	589.27(3)	10.9(9)	26(3)	21:49:23.63 ± 0.05 ± 0.24	-52:58:15.4 ± 0.3 ± 1.4	0.4755	3.1 <sup>+15.7</sup> <sub>-3.1</sub>
190102	05:38:43.49184	363.6(3)	110	14(1)	21:29:39.76 ± 0.06 ± 0.16	-79:28:32.5 ± 0.2 ± 0.5	0.291	1.5 <sup>+3.4</sup> <sub>-1.5</sub>
190608	22:48:12.88391	338.7(5)	–	26(4)	22:16:04.75 ± 0.02 ± 0.02	-07:53:53.6 ± 0.3 ± 0.3	0.1178	7.0 ± 1.3
190611	5:45:43.29937	321.4(2)	–	10(2)	21:22:58.91 ± 0.11 ± 0.23	-79:23:51.3 ± 0.3 ± 0.6	0.378	17.2 ± 4.9
190711	01:53:41.09338	593.1(4)	–	34(3)	21:57:40.68 ± 0.051 ± 0.15	-80:21:28.8 ± 0.08 ± 0.3	0.522	1.5 <sup>+3.6</sup> <sub>-1.5</sub>

The FRB detection pipeline makes use of the DBSCAN algorithm<sup>25</sup>, as implemented by ref. <sup>26</sup>, to mitigate RFI and reduce the frequency of false-positive FRB triggers.

<sup>a</sup>Burst arrival time referenced to a frequency of 1,152 MHz.

<sup>b</sup>Quoted errors on the last significant digit of the fluence represent a 90% confidence limit.

<sup>c</sup>Errors listed after the burst position represent the statistical and systematic uncertainties respectively, and are combined in quadrature for a final absolute positional uncertainty.

over all physical distance increments  $ds$  to a given FRB:  $DM_{\text{FRB}} = \int n_e ds / (1+z)$ . Physically, we expect  $DM_{\text{FRB}}$  to separate into four primary components:

$$DM_{\text{FRB}}(z) = DM_{\text{MW,ISM}} + DM_{\text{MW,halo}} + DM_{\text{cosmic}}(z) + DM_{\text{host}}(z) \quad (1)$$

with  $DM_{\text{MW,ISM}}$  the contribution from our Galactic ISM,  $DM_{\text{MW,halo}}$  the contribution from our Galactic halo<sup>13</sup>,  $DM_{\text{host}}$  the contribution from the host galaxy including its halo and any gas local to the event, and  $DM_{\text{cosmic}}$  the contribution from all other extragalactic gas. Only  $DM_{\text{cosmic}}$ , determined by its path length through the IGM and the increase in baryon density with look-back time, is expected to have a strong redshift dependence, although  $DM_{\text{host}}$  is weighted by  $(1+z_{\text{FRB}})^{-1}$  and may correlate with age, for example, if host galaxies have systematically lower mass at earlier times.

Adopting our cosmological paradigm of a flat universe with matter and dark energy, the average value of  $DM_{\text{cosmic}}$  to redshift  $z_{\text{FRB}}$  is:

$$\langle DM_{\text{cosmic}} \rangle = \int_0^{z_{\text{FRB}}} \frac{c \bar{n}_e(z) dz}{H_0(1+z)^2 \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}} \quad (2)$$

with mean density  $\bar{n}_e = f_d \rho_b(z) m_p^{-1} (1 - Y_{\text{He}}/2)$ , where  $m_p$  is the proton mass,  $Y_{\text{He}} = 0.25$  is the mass fraction of helium, assumed doubly ionized in this gas,  $f_d(z)$  is the fraction of cosmic baryons in diffuse ionized gas (this accounts for dense baryonic phases, for example, stars, neutral gas; see Methods),  $\rho_b(z) = \Omega_b \rho_{c,0} (1+z)^3$ , and  $\Omega_m$  and  $\Omega_\Lambda$  are the matter and dark energy densities today in units of  $\rho_{c,0} = 3H_0^2/8\pi G$  where we parameterize Hubble's constant  $H_0$  in terms of the dimensionless  $h_{70} = H_0/(70 \text{ km s}^{-1} \text{ Mpc}^{-1})$ .

The  $DM_{\text{MW,ISM}}$  term (equation (1)) arises primarily from the so-called warm ionized medium of the Galaxy and is estimated from a model of this ISM component<sup>22,23</sup>. At the high Galactic latitudes ( $|b| > 33^\circ$ ) of the ASKAP sample, the value is  $DM_{\text{MW,ISM}} \approx 30 \text{ pc cm}^{-3}$ . The  $DM_{\text{MW,halo}}$  term is not well constrained<sup>13</sup>, but is expected to be in the range of approximately 50–100  $\text{pc cm}^{-3}$ . Hereafter we assume  $DM_{\text{MW,halo}} = 50 \text{ pc cm}^{-3}$  and emphasize that the sum of its scatter and uncertainty are less than those of  $DM_{\text{cosmic}}$  and  $DM_{\text{host}}$  which we discuss below.

Figure 2 shows the theoretical curve for  $\langle DM_{\text{cosmic}} \rangle$  versus  $z_{\text{FRB}}$  for the Planck15 cosmology<sup>20</sup> and a model estimate of the scatter (90% interval) due to statistical variations in foreground cosmic structure (see Methods). Overplotted on the model are the estimated  $DM_{\text{cosmic}}$  and measured  $z_{\text{FRB}}$  values for all arcsecond-localized FRBs. We have estimated  $DM_{\text{cosmic}}$  by subtracting the following from the measured  $DM_{\text{FRB}}$  value:  $DM_{\text{MW,ISM}}$  from the Galactic ISM model; our assumed  $DM_{\text{MW,halo}}$  contribution; and an ansatz of  $DM_{\text{host}} = 50/(1+z) \text{ pc cm}^{-3}$  estimated from theoretical work and informed from the analysis below. We ignore

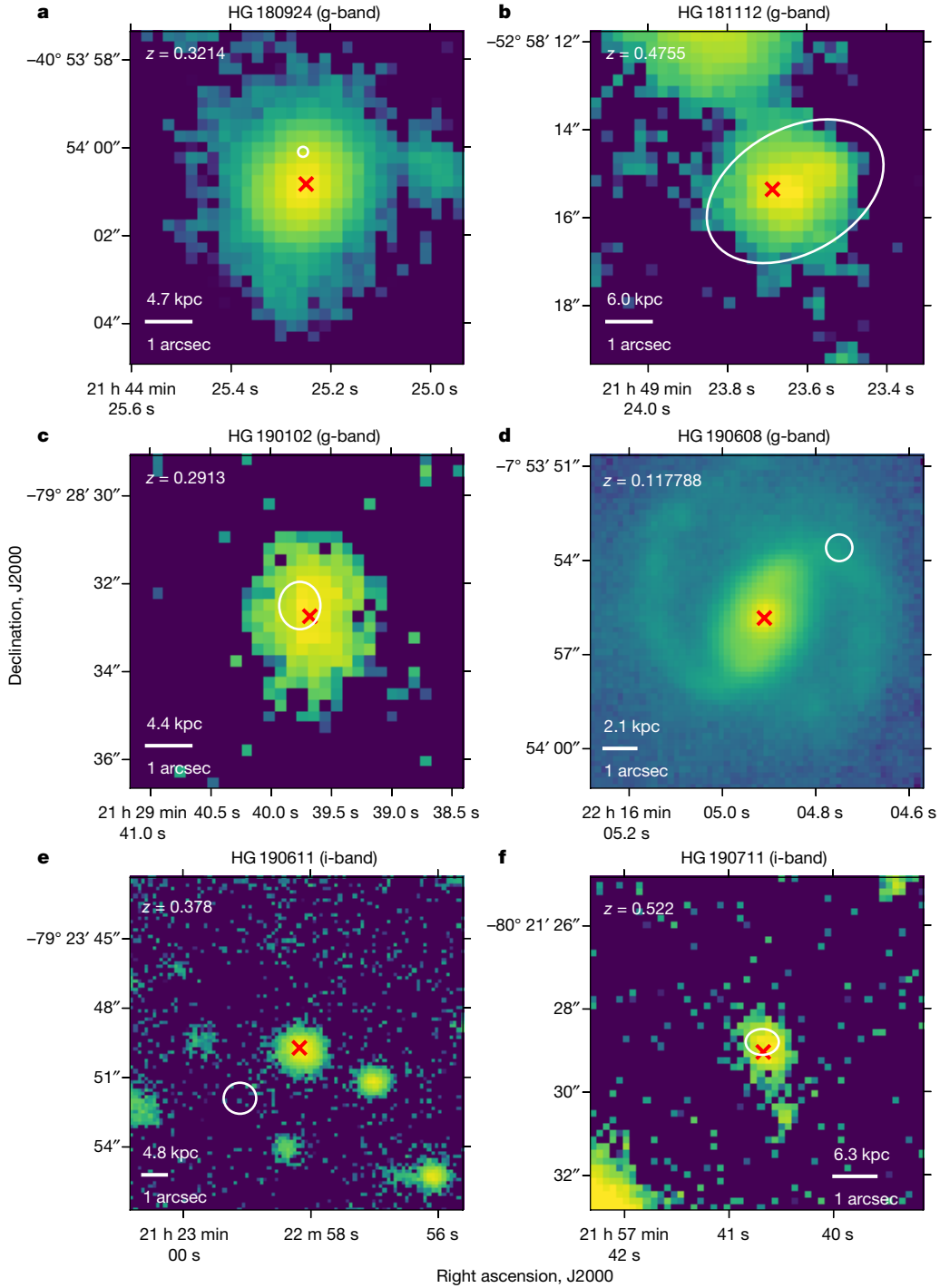
FRB 121102 and FRB 190523 in most of the analysis that follows because of selection bias in their discovery: FRB 180916 due to its low Galactic latitude (see Methods), and FRB 190611 due to its tentative association with a host galaxy (see Methods). The five ASKAP FRBs that remain comprise what we term the gold-standard sample. The agreement between model and data is striking. Effectively, the FRB measurements confirm the presence of baryons with the density estimated from the cosmic microwave background (CMB) and Big Bang nucleosynthesis (BBN), and these five measurements are consistent with all the missing baryons being present in the ionized IGM.

This result motivated us to quantitatively test for the consistency of  $\Omega_b h_{70}$  with CMB and BBN measurements, simultaneously determining the uncertain host-galaxy contributions to  $DM_{\text{FRB}}$  as well as the sightline-to-sightline variance in dispersion owing to the IGM. We do this by analysing the joint likelihood of our sample of five (seven, including FRBs 190523 and 190611)  $DM_{\text{FRB}}, z_{\text{FRB}}$  measurements against a four parameter model: one parameter for the large-scale structure scatter in  $DM_{\text{cosmic}}$ , two parameters for  $DM_{\text{host}}$  (a mean and a scatter), and  $\Omega_b h_{70}$ . Our model for the contributions to  $DM_{\text{FRB}}$  starts with equations (1) and (2), and we develop parametric models for  $DM_{\text{host}}$  and the intrinsic scatter in  $DM_{\text{cosmic}}$ . We again fix  $DM_{\text{MW,halo}} = 50 \text{ pc cm}^{-3}$  and adopt the Galactic ISM model<sup>22</sup> for  $DM_{\text{MW,ISM}}$ . Uncertainty in these values can be absorbed into our model for  $DM_{\text{host}}$ .

For  $DM_{\text{cosmic}}$ , our model accounts for scatter in the electron column from foreground structures, which is largely caused by random variation in the number of haloes a given sightline intersects. Cosmological simulations show that this variation is sensitive to the extent to which galactic feedback redistributes baryons around galactic haloes<sup>11,16</sup> and that the fractional standard deviation of the cosmic DM equals approximately<sup>11</sup>  $Fz^{-1/2}$  for  $z < 1$ , where the parameter  $F$  quantifies the strength of the baryon feedback (0.1 being strong feedback and 0.4 being weak). Stronger feedback corresponds to situations in which feedback processes expel baryons to larger radii from their host galaxies or where more massive haloes are evacuated by such feedback. The formalism incorporates the effect of large-scale structure associated with voids and the intersection of sightlines with clusters. We find that a one-parameter model based on a physically motivated shape for the probability distribution of  $DM_{\text{cosmic}}$  given  $F$  provides a successful description of a wide range of cosmological simulations (see Methods). Our form for the distribution is strongly asymmetric towards lower redshifts, admitting large  $DM_{\text{cosmic}}$  values that we find are important in the estimation of  $\Omega_b h_{70}$ .

We chose our model for  $DM_{\text{host}}$  to follow a log-normal distribution, characterized by a median  $\exp(\mu)$  and logarithmic width parameter  $\sigma_{\text{host}}$  such that the standard deviation of the distribution is  $\exp(\mu) e^{\sigma_{\text{host}}^2/2} (e^{\sigma_{\text{host}}^2} - 1)^{1/2}$ . We do not attempt to incorporate redshift-dependent evolution in the host-galaxy dispersion contribution, but we do scale the distribution of  $DM_{\text{host}}$  by the factor  $(1+z_{\text{host}})^{-1}$  applicable

Q7



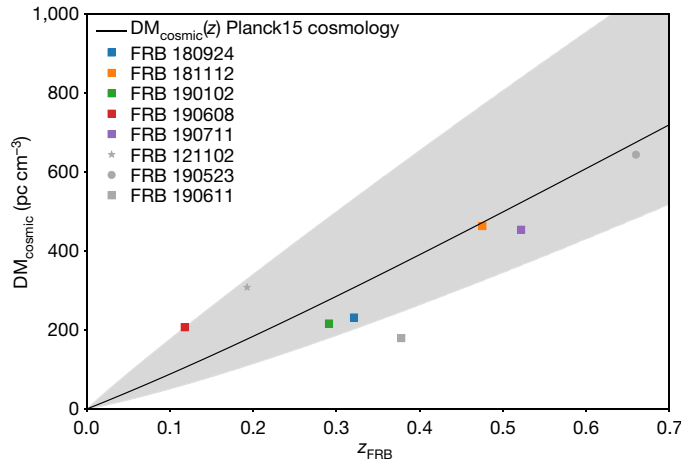
**Fig. 1 | Locations of FRBs relative to their host galaxies.** **a–f**, Optical images of the host galaxies of six FRBs localized by ASKAP, including the four new bursts reported here. **a**, FRB 180924; **b**, FRB 181112; **c**, FRB 190102; **d**, FRB 190608; **e**, FRB 190611; and **f**, FRB 190711. HG, host galaxy. White ellipses denote the 90%

confidence region of each burst position, including statistical uncertainty and phase referencing errors, while the red crosses mark the measured centroids of each host galaxy. The identification of the host galaxy of FRB 190611 is tentative. **a–d**, Deep VLT g-band images; **e, f**, deep GMOS i-band images.

to a parcel of plasma at redshift  $z_{\text{host}}$  so that  $\text{DM}_{\text{host}}$  is interpreted as the dispersion in the rest frame of the host galaxy. Our choice of a log-normal distribution is conservative in that it allows for a tail extending to large positive values, which may not be present in our sample given our selection criteria (see Methods) and the burst locations relative to the host stellar surface density. We explore  $\text{DM}_{\text{host}}$  distributions with median values in the range  $\exp(\mu) = 20\text{--}200 \text{ pc cm}^{-3}$  and  $\sigma_{\text{host}}$  in the range 0.2–2.0.

Our final analysis compares the relative likelihood of models in a four-parameter space ( $\Omega_b h_{70}$ ,  $F$ ,  $\sigma_{\text{host}}$ ,  $\mu$ ); see Methods for a Bayesian

approach that yields similar constraints. Marginalizing the other parameters over ranges restricted by other physical constraints, we derive the constraints on  $\Omega_b h_{70}$  shown in Fig. 3 using our five-FRB gold-standard sample. The results are fully consistent with the joint CMB + BBN estimations and with only five (seven) burst redshifts the experiment yields a precision of  $\sigma(\Omega_b h_{70})/\Omega_b h_{70} = 0.31$  (0.28) at the 68% confidence level, with  $F$  marginalized over the range [0.09, 0.32] (see Methods). This quantitative result for  $\Omega_b h_{70}$  substantiates our inference from the DM– $z$  relation in Fig. 2 that the FRB ensemble has resolved the missing

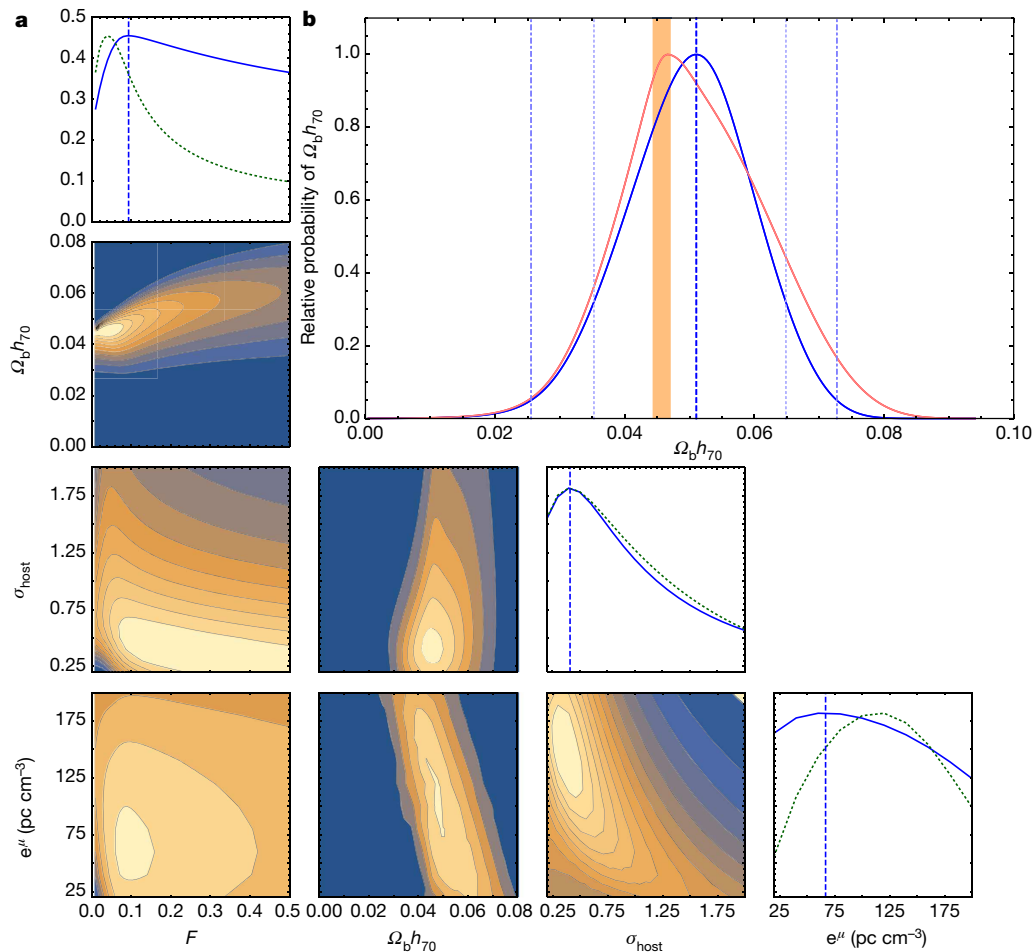


**Fig. 2 | The DM–redshift relation for localized FRBs.** Data points are estimations of the cosmic dispersion measure ( $DM_{\text{cosmic}}$ ) versus FRB redshift ( $z_{\text{FRB}}$ ) for all current arcsecond- and subarcsecond-localized FRBs. The  $DM_{\text{cosmic}}$  values are derived by correcting the observed dispersion measure  $DM_{\text{FRB}}$  for the estimated contributions from our Galaxy and the FRB host galaxy (the latter assumed here to be  $50(1+z)^{-1} \text{ pc cm}^{-3}$ ; see text for details). Coloured points represent the gold-standard sample on which our primary analysis is based. The solid line denotes the expected relation between  $DM_{\text{cosmic}}$  and redshift for a universe based on the Planck15 cosmology (that is,  $\Omega_b = 0.0486$  and  $H_0 = 67.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ). The shaded region encompasses 90% of the  $DM_{\text{cosmic}}$  values from a model for ejective feedback in Galactic haloes that is motivated by some simulations (with  $F = 0.2$  in equation (4) in Methods), illustrating that the observed scatter is largely consistent with the scatter from the IGM.

baryons problem. The ratio of our estimated  $\Omega_b$  to that from CMB and BBN measurements is  $1.1^{+0.5}_{-0.6} h_{70}$ . Formally, we exclude  $\Omega_b h_{70} < 0.02$  (0.01) at the 98.6% (99.8%) confidence level. This constraint should

improve considerably in the near term as ASKAP and other facilities acquire a larger sample of bursts with redshifts.

Additionally, analysis of our gold-standard sample mildly favours a median host galaxy contribution of about  $100 \text{ pc cm}^{-3}$  with a factor of two dispersion around this value ( $\sigma_{\text{host}} \approx 1$ ). This quantifies our



**Fig. 3 | The density of cosmic baryons derived from the FRB sample.** The constraints on the IGM parameters  $\Omega_b h_{70}$  and  $F$ , and the host galaxy parameters  $\exp(\mu)$  and  $\sigma_{\text{host}}$ , for a log-normal DM distribution are derived using the five gold-standard bursts (as described in the text and Methods). **a**, Corner plots displaying the probability of a given value of  $F$ ,  $\exp(\mu)$  or  $\sigma_{\text{host}}$  relative to its most likely value, and marginalized over the other parameters: heavy dashed lines represent the most likely values in each case. The green, dotted lines in the corner plots of  $F$ ,  $\exp(\mu)$  and  $\sigma_{\text{host}}$  denote the relative likelihood of these parameters when  $\Omega_b h_{70}$  is constrained to the value set by the CMB + BBN measurements.

The contours displayed are in increments of 10% of the peak value. **b**, Magnified view of the corner plot for  $\Omega_b h_{70}$  where the orange shaded region denotes the range to which  $\Omega_b h_{70}$  is confined by CMB + BBN measurements. The dotted and dot-dashed lines represent the 68% and 95% confidence intervals of each parameter, respectively. The distribution of  $\Omega_b$  is alternatively marginalized over the range of  $F$  indicated by cosmological simulations,  $[0.09, 0.32]$  (blue curve; see Methods), and over the entire range of  $F$   $[0, 0.5]$  investigated here (red curve).

result that the host contributions are sufficiently small to not compromise the use of FRBs for cosmology and IGM science. Even with our current small sample, we are beginning to constrain viable models for the redistribution of the cosmic baryons by galactic feedback. If we adopt a prior on  $\Omega_b h_{70}$  from the CMB, BBN and supernovae surveys<sup>24</sup>, we find  $F = 0.04^{+0.26}_{-0.04}$  (68% confidence), and if we further include FRB 190523 and FRB 190611 we find  $F = 0.23^{+0.27}_{-0.12}$  (see Methods). A factor of two smaller error would start to differentiate between viable feedback scenarios (as discussed further in Methods), suggesting that FRBs have not only revealed that all the baryons are present but—with modestly larger samples—could constrain where they lie.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2300-2>.

1. Fukugita, M., Hogan, C. J. & Peebles, P. J. E. The cosmic baryon budget. *Astrophys. J.* **503**, 518–530 (1998).
2. Cen, R. & Ostriker, J. P. Where are the baryons? II. Feedback effects. *Astrophys. J.* **650**, 560–572 (2006).
3. Shull, J. M., Smith, B. D. & Danforth, C. W. The baryon census in a multiphase intergalactic medium: 30% of the baryons may still be missing. *Astrophys. J.* **759**, 23 (2012).
4. Nicastro, F. et al. Observations of the missing baryons in the warm-hot intergalactic medium. *Nature* **558**, 406–409 (2018).
5. Tripp, T. M. et al. The heavy-element enrichment of Ly $\alpha$  clouds in the Virgo supercluster. *Astrophys. J.* **575**, 697–711 (2002).
6. Tumlinson, J. et al. The large, oxygen-rich halos of star-forming galaxies are a major reservoir of galactic metals. *Science* **334**, 948–952 (2011).
7. Prochaska, J. X., Weiner, B., Chen, H. W., Mulchaey, J. & Cooksey, K. Probing the intergalactic medium/galaxy connection. V. On the origin of Ly $\alpha$  and O VI absorption at  $z < 0.2$ . *Astrophys. J.* **740**, 91 (2011).
8. Hojjati, A. et al. Cross-correlating Planck tSZ with RCSLenS weak lensing: implications for cosmology and AGN feedback. *Mon. Not. R. Astron. Soc.* **471**, 1565–1580 (2017).
9. de Graaff, A., Cai, Y.-C., Heymans, C. & Peacock, J. A. Probing the missing baryons with the Sunyaev-Zel'dovich effect from filaments. *Astron. Astrophys.* **624**, A48 (2019).
10. Eckert, D. et al. Warm-hot baryons comprise 5–10 per cent of filaments in the cosmic web. *Nature* **528**, 105–107 (2015).
11. McQuinn, M. Locating the “missing” baryons with extragalactic dispersion measure estimates. *Astrophys. J.* **780**, L33 (2014).
12. Macquart, J. P. et al. Fast transients at cosmological distances with the SKA. In *Proc. Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (eds Bourke, T. L. et al.) 55 (Proceedings of Science, 2015).
13. Prochaska, J. X. & Zheng, Y. Probing Galactic haloes with fast radio bursts. *Mon. Not. R. Astron. Soc.* **485**, 648–665 (2019).
14. Chatterjee, S. et al. A direct localization of a fast radio burst and its host. *Nature* **541**, 58–61 (2017).
15. Bannister, K. W. et al. A single fast radio burst localized to a massive galaxy at cosmological distance. *Science* **365**, 565–570 (2019).
16. Prochaska, J. X. et al. The low density and magnetization of a massive galaxy halo exposed by a fast radio burst. *Science* **366**, 231–234 (2019).
17. Ravi, V. et al. A fast radio burst localized to a massive galaxy. *Nature* **572**, 352–354 (2019).
18. Marcote, B. et al. A repeating fast radio burst source localized to a nearby spiral galaxy. *Nature* **577**, 190–194 (2020).
19. Cooke, R. J., Pettini, M. & Steidel, C. C. One percent determination of the primordial deuterium abundance. *Astrophys. J.* **855**, 102 (2018).
20. Planck Collaboration. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* **594**, A13 (2016).
21. Bhandari, S. et al. The host galaxies and progenitors of fast radio bursts localized with the Australian Square Kilometre Array Pathfinder. *Astrophys. J.* (in the press).
22. Cordes, J. M. & Lazio, T. J. W. NE2001.I. A new model for the galactic distribution of free electrons and its fluctuations. Preprint at <https://arxiv.org/abs/astro-ph/0207156> (2002).
23. Gaensler, B. M., Madsen, G. J., Chatterjee, S. & Mao, S. A. The vertical structure of warm ionised gas in the Milky Way. *Publ. Astron. Soc. Aust.* **25**, 184–200 (2008).
24. Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. Preprint at <http://arXiv.org/abs/1807.06209> (2018).
25. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. Second Int. Conf. on Knowledge Discovery and Data Mining (KDD-96)* 226–231 (AAAI Press, 1996).
26. Novikov, A. PyClustering: data mining library. *J. Open Source Softw.* **4**, 1230 <https://doi.org/10.21105/joss.01230> (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Q8

Q9

## Methods

## Sample selection

Analogously to cosmological studies of the distance ladder with supernovae, we wish to establish a strict set of criteria for the FRB sample to minimize biases while maximizing statistical power. On the latter point, we wish to construct the largest sample while avoiding events whose DM is dominated by non-cosmological effects (that is, host or Galactic gas). Regarding bias, the greatest concern is the association of a host galaxy with a given FRB on the basis of the DM– $z$  relation, that is, adopting this relation as a prior to establish the host identity. This is a valuable practice when one aims to resolve the underlying host-galaxy population<sup>17,27</sup> but would bias any cosmological study. Last, one must also be cognizant of biases related to triggering on FRB events. This practice is a complex function of the FRB fluence, its DM, and the pulse properties<sup>28</sup>.

With these issues in mind, we propose the following set of criteria (1–4) to generate a ‘gold-standard sample’ of FRBs for cosmological study:

1. To make a confident host-galaxy association, we require the probability of mis-identifications to be <1% without invoking the DM– $z$  relation since to do so would introduce a bias. For this we require the 95% localization area to encompass one and only one galaxy unless multiple galaxies have a common redshift. By ‘encompass’ we include light from any part of the galaxy. In practice, this will require a localization to <1" for  $z > 1$ , but becoming less stringent for less distant hosts. We propose an initial set of specific criteria as follows. (i) Define 95% areas for the localization and for each galaxy in the region. Call these  $L$  and  $G_1, G_2$ , and so on. (ii) Demand one and only one galaxy overlap  $L$ . The only exception is if  $z_1 = z_2$ . (iii) For the overlapping  $L$  and  $G$ , require that >50% of the smaller area lies within the larger. (iv) Do this for galaxies as faint as  $R = 25$  (anything fainter is generally too difficult for a spectroscopic redshift anyhow).
2. The finite temporal and spectral resolution of the FRB survey causes a decline in sensitivity with increasing DM to the point that telescope resolution causes an effective threshold at  $DM_{\text{cutoff}}$  at which point a burst would no longer have been detectable. A conservative approach would omit any burst with  $DM_{\text{cutoff}}$  sufficiently low that it excludes a large ( $\geq 30\%$ ) fraction of the total probability of  $p(DM_{\text{total}}|z)$ , on the grounds that it presents a biased probe of  $p(DM_{\text{total}}|z)$ . Although application of this criterion presupposes a DM– $z$  relation and its probability density function (PDF), it does so only weakly. This point is addressed in detail below in the subsection on biases in the probability distribution.
- An event detected near to the sensitivity threshold is biased in the sense that the instrumental decline in sensitivity with increasing DM dictates that any burst detected near this threshold would not have been detectable at higher DMs. Thus, for a given redshift we are biased to finding events with DMs that are under-representative of the entire DM distribution at that redshift. Thus, only more luminous bursts, whose detection  $S/N$  is sufficiently high that  $DM_{\text{cutoff}}$  exceeds the plausible range of  $DM_{\text{total}}$  at that redshift are devoid of this bias.
3. FRB events with extreme properties (for example, high RM, large temporal broadening) will be excluded to minimize the impact of host galaxy and Galactic gas.
4. A cutoff is imposed on FRBs whose expected contribution from the disk component of the Milky Way ISM is large, to avoid large uncertainties in the subtraction of the Galactic ISM DM contribution. Models of the Galactic plasma distribution<sup>22</sup> typically produce errors in known pulsar distances of the order of several tens of per cent (and much higher in some cases)<sup>29,30</sup>. To avoid DM errors in excess of about  $100 \text{ pc cm}^{-3}$  we restrict our sample to those bursts whose predicted  $DM_{\text{ISM}}$  values are less than  $100 \text{ pc cm}^{-3}$ . A conservative application of this criterion restricts FRB detections to sight lines at high Galactic latitude,  $|b| \geq 20^\circ$ .

We acknowledge that all criteria 1–4 are subject to refinement as we learn more about FRB progenitors and their host galaxies.

Regarding criterion 3, a dominant contributor to the DM variance is the circumburst environment and the ISM of the host galaxy. Although it is not possible to make a precise estimate of this component, the burst RM, the amount of Faraday rotation exhibited by linearly polarized emission caused by its propagation through a magnetized plasma, presents a means of identifying those bursts whose radiation has probably propagated through a substantial ( $>100 \text{ pc cm}^{-3}$ ) amount of matter in the host galaxy. For each burst the Milky Way contribution to RM for  $|b| > 10^\circ$  is small ( $<250 \text{ rad m}^{-2}$ ) and measurable<sup>31</sup> and the IGM contribution is estimated<sup>32</sup> to be about  $1 \text{ rad m}^{-2}$ . Galactic haloes, similarly, have been inferred to make contributions of several tens of  $\text{rad m}^{-2}$  to the RM<sup>33</sup> from radio-loud quasar observations, but our first analysis with an FRB<sup>16</sup> yields  $RM < 10 \text{ rad m}^{-2}$ . A suitable cutoff due to host-galaxy ISM contamination is suggested by assuming the host-galaxy magnetic field strength is comparable to that of our Galaxy. Measurements of Faraday rotation and dispersion from pulsars in the Milky Way (see figure 3 in ref. <sup>34</sup>) exhibit a mean trend  $DM = 1.55|RM|^{0.95} \equiv f_{\text{DM}}(RM)$ , where RM and DM are measured in their usual units of  $\text{rad m}^{-2}$  and  $\text{pc cm}^{-3}$  respectively. We find that the root-mean-square (r.m.s.) deviation of the actual DM values from their values predicted on the basis of this trend using  $|RM|$  are 69% of the DM (that is, the r.m.s. errors are 69% of the mean DM value:  $\langle [DM - f_{\text{DM}}(RM)]^2 / DM^2 \rangle^{1/2} = 0.69$ ). We further find that there is an 85% probability that the actual DM deviates from its predicted value by less than 0.9 times the actual DM value, and a 96% probability that the predicted value differs by less than 2.0 times the DM value. We therefore suggest that a cutoff criterion  $|RM - RM_{\text{MW}}|_{\text{observed}} < 100(1+z)^2 \text{ rad m}^{-2}$  bounds the dispersion measure to  $DM \lesssim 250(1+z) \text{ pc cm}^{-3}$  with 85% confidence.

A similar trend observed between the DM and the temporal smearing of Galactic pulsars caused by scattering<sup>35,36</sup> can also be used to place upper bounds on the host contribution. Recent updates to this relation<sup>36</sup> indicate that, on average, a pulse smearing time,  $\tau$ , less than 33 ms (2 ms) limits DM to  $<200 \text{ pc cm}^{-3}$  ( $300 \text{ pc cm}^{-3}$ ) at 0.327 GHz (1 GHz). However the DM– $\tau$  relation exhibits  $\sim 0.8$  dex variation about the trend (as discussed in the context of FRBs in the supplementary material in ref. <sup>15</sup>), thus requiring  $\tau < 5 \text{ ms}$  to ensure a reasonable ( $\sim 70\%$ ) confidence that the DM contribution is less than  $200 \text{ pc cm}^{-3}$ . We caution that the use of  $\tau$  as an indicator of the host-galaxy DM contribution is subject to considerable uncertainty, since neither the distances to the scattering material from the bursts, nor even the nature of the turbulence responsible for the temporal smearing observed in FRBs, is well established. The estimates presented here would be invalid, for instance, if the scattering were associated with the direct burst environment rather than the ISM of the host galaxy.

Adopting all the above criteria to the current set of FRBs with redshift estimates based on their association with galaxies (Table 1), we eliminate the following sources from cosmological analysis.

**FRB 121102.** We exclude the repeating FRB 121102<sup>14</sup> from our analysis for two reasons: (a) the rotation measure of this burst is anomalously high<sup>37</sup>, being three orders of magnitude higher than other FRBs in this sample and indicating that this burst DM is likely contaminated by an abnormally high circumburst or host galaxy contribution, and (b) its location within  $2^\circ$  of the Galactic plane imparts a larger and probably less well constrained DM contribution from the Milky Way relative to the high Galactic latitude bursts detected by ASKAP.

**FRB 190523.** We have conducted the analysis both with and without FRB 190523. The host galaxy identification<sup>17</sup> from the larger,  $3'' \times 8''$  localization region, is more uncertain than the ASKAP FRB detections and was partially based on an assumed DM–redshift relation which presents a potential source of bias in our analysis.



**FRB 171020.** The identification of the host galaxy associated with FRB 171020 is predicated on a search volume confined to a specific distance based on an assumed DM–redshift relation<sup>27</sup>, and is therefore excluded from the sample. Moreover, it is difficult to ascribe a numerical value to the likelihood of a correct association in this instance.

**FRB 190611.** Our follow-up observations for FRB 190611 identify a galaxy at J212258.0-792350 with redshift  $z = 0.378$  offset by  $\sim 2''$  from the current estimate of the FRB localization. The large offset ( $\sim 10$  kpc at that redshift) and large systematic uncertainty in the FRB localization and the presence of a closer, faint source revealed by deep GMOS i-band imaging preclude a secure association at present. As with FRB 190523, we conduct our analysis both with and without this burst in our sample.

### FRB imaging and astrometry

The procedure for characterizing the position and positional uncertainty of FRBs 190102, 190608, 190611 and 190711 followed that described in the supplementary material of refs. <sup>15,16</sup>. For the purposes of extracting these observables, we use only the total intensity data.

For each FRB, raw voltage data for a suitable calibrator source was captured via the CRAFT pipeline in the hours following the burst detection. For FRB 190102 and FRB 190608, the source PKS 1934-638 was used, while for FRB 190611 and FRB 190711, it was PKS 0407-658. From these calibrator data and the FRB data, visibility data sets were produced using the DiFX correlator<sup>38</sup>. An initial coarse search for the FRB position used the DM, pulse duration, and approximate position from the incoherently summed FRB detection data, and after detection in the interferometric data a re-correlation was performed with revised position, DM, and pulse time/duration. Radio frequency interference (RFI) was mitigated for the FRB data set by subtracting visibilities from an adjacent time range surrounding the burst itself. Additionally, for each FRB a visibility data set and image was generated using the entire 3.1 s of raw voltage data, to identify background radio continuum sources whose positions could be compared to catalogue values and verify the astrometric accuracy.

Per-station frequency-dependent complex gain calibration was derived from the calibrator data set using the ParselTongue<sup>39</sup> based pipeline described in ref. <sup>15</sup> and transferred to the FRB data sets, before imaging in the Common Astronomy Software Applications (CASA) package. Best-fit positions and uncertainties were the extracted for each source using the task JMFIT in the Astronomical Image Processing System (AIPS)<sup>40</sup>.

Statistical uncertainties on the FRB positions were less than  $0.5''$  in all cases. However, as discussed in refs. <sup>15,16</sup>, the phase referenced FRB images will be subject to a systematic positional shift resulting from the spatial and temporal extrapolation of calibration solutions. The magnitude of this systematic shift can be estimated by comparing the positions of continuum sources in the field surrounding the FRBs to their catalogue values. The accuracy to which this can be performed depends on the number of continuum sources visible in the ASKAP continuum image and their brightness, as well as the degree to which their intrinsic source structure can be modelled (or neglected). For any given continuum source, the presence of unmodelled structure will act to shift the position of the source centroid and results in a measured offset between the ASKAP and reference positions, which perturbs the actual systematic positional shift. However, the direction of such a shift depends on the source structure, and hence should not be correlated between different continuum sources. For FRB 190102, FRB 190611 and FRB 190711, observations made with the Australia Telescope Compact Array at a comparable frequency and angular resolution to the ASKAP image minimize the impact of source structure, but for FRB 190608, we made use of the Faint Images of the Radio Sky at Twenty centimetres (FIRST) survey<sup>41</sup>, which has angular resolution roughly twice that of the ASKAP images.

Assuming the phase referencing errors result in a simple translation of the FRB field image, we estimate the magnitude of this offset and its uncertainty with a weighted mean of the measured offsets for each of the continuum sources in the FRB field, after discarding any sources that were resolved in either the ASKAP image or the reference image. The magnitude of the offset ranged between 0 and 1.7 arcsec, with uncertainties ranging from 0.3 to 0.6 arcsec.

### Host identification and spectroscopy

The optical spectroscopy and redshift determinations for FRB 180924 and FRB 181112 have been outlined previously<sup>15,16</sup>. Spectroscopy of the host galaxies of FRB 190102 and FRB 190611 was conducted using the FOcal Reducer and low dispersion Spectrograph 2<sup>42</sup> (FORS2) on the European Southern Observatory's Very Large Telescope (VLT) on Cerro Paranal. FORS2 was configured with the GRIS\_300I grism, an OG590 blocking filter, and a  $1.3''$  wide slit, yielding a resolution  $R_{\text{FWHM}} \approx 550$ . For FRB 190102  $2 \times 600$  s exposures were obtained on 2019 March 25 UT, while for FRB 190611  $2 \times 1,350$  s exposures were taken on 2019 July 12 UT. These and associated calibration images were processed with the Pypelt software package<sup>43</sup> to derive flux and wavelength calibrated spectra.

For the host galaxy of FRB 190608, the optical spectrum from the seventh data release (DR7) of the Sloan Digital Sky Survey<sup>44</sup> (SDSS) was retrieved from the IGMSPEC database<sup>45</sup>.

Imaging of the host galaxies of FRB 180924, FRB 181112 and FRB 190102 was undertaken using FORS2 on the VLT, while the FRB 190608 and FRB 190711 hosts were imaged with VLT/X-shooter<sup>46</sup>. Imaging of the host galaxies of FRB 190611 and FRB 190711 was undertaken using GMOS on Gemini-South<sup>47</sup>, from sets of 44 and 12 images of 100 s each in the i-band, respectively.

The FORS2 images were first reduced with ESO Reflex<sup>48</sup>, further processed in Python, and then co-added using a median combine in Montage<sup>49</sup>. The WCS solutions were updated with Astrometry.net<sup>50</sup>, with further adjustments performed by comparison with Gaia<sup>51</sup> or Dark Energy Survey<sup>52</sup> positions. The X-shooter images were reduced using a custom Python pipeline making use of the package CCDPROC<sup>53</sup>, including measures to cope with prominent fringe patterns in i-band; the images were then co-added and the astrometry adjusted with the same method as above. The GMOS images were reduced and co-added with PYRAF using standard procedures; the astrometry was adjusted with the same method stated above. Projected distances were estimated using the Javascript Cosmology Calculator<sup>54</sup>.

Two of the FRBs in the gold-standard sample, FRB 190608 and FRB 190711, have offsets larger than 1 arcsec from the galaxy light centroid. FRB 190608, however, is a  $z = 0.11$  galaxy (that is, nearby) and the chance projection is even less than 0.3%. Regarding FRB 190711 we estimate a  $6.1 \times 10^{-3}$  probability that an unrelated galaxy is within a region out to a distance between the galaxy centroid and the outermost edge of the FRB error circle (for the measured  $R(\text{AB}) = 23.7 \pm 0.2$  mag as calibrated against the SkyMapper survey), and we estimate a probability  $p = 1.9 \times 10^{-3}$  for an unrelated galaxy to be within the FRB error circle but below the detection limit of  $r = 25.5$  mag. The remainder of the host-galaxy associations for each FRB have a probability  $P < 10^{-3}$  of a chance occurrence<sup>16</sup>.

The radio burst dynamic spectra and host-galaxy optical spectra are shown in Extended Data Fig. 1.

### Estimating $\langle \text{DM}_{\text{cosmic}} \rangle$

Central to the analysis is an estimate of the average  $\text{DM}_{\text{cosmic}}$  value as a function of redshift and for a given cosmology, as defined in equation (2). Previous formulations<sup>55,56</sup> have adopted similar definitions but with less precise considerations for  $f_d(z)$ , the fraction of cosmic baryons in diffuse ionized gas. Our formulation considers the redshift evolution of three dense baryonic components that will not contribute to  $n_e$ : (1) stars; (2) stellar remnants (for example, white dwarfs, neutron stars); and (3) the neutral ISM of galaxies. For (1), we interpolate

the empirically estimated stellar mass density estimates<sup>57</sup>. For (2), we adopt the estimation of ref. <sup>58</sup> which is 30% of the stellar mass. For (3), we assume the mass ratio of the ISM to stars is constant from  $z = 0-1$  and adopt the present-day estimate<sup>58</sup> of  $M_{\text{ISM}}/M_* = 0.38$ . The model also allows for the partial ionization of helium but this is not relevant for the FRBs considered here. All of these calculations are encoded in Python in the public FRB repository (<https://github.com/FRBs/FRB>). Censuses of the gas and star evolution of baryons in  $z < 1$  systems constrain the error in the fraction of neutral and non-diffuse baryons (that is,  $1 - f_d(z)$ ) to about 30% at present. Thus, with this component constituting ~15% of the total baryon budget at  $z \approx 0$ , the correction to  $\Omega_b h_{70}$  is uncertain at a level below 6%, well below the level of precision that investigation of the current FRB sample permits. We refer the reader to ref. <sup>59</sup> for a discussion of the constraints on  $f_d(z)$  possible in future with a larger sample of FRBs.

## Cosmological parameter estimation

The ASKAP FRB measurements and localizations afford a new opportunity to constrain our cosmological paradigm through estimations of  $\text{DM}_{\text{cosmic}}$  and  $z_{\text{FRB}}$ . The cosmic DM is governed primarily by the baryonic density  $\Omega_b$  and the expansion rate of the Universe,  $H_0$ , and the fraction of baryons in the diffuse phase,  $f_d(z)$ . In the following, we will assume a flat cosmology with  $\Omega_\Lambda = 0.691$  (Planck15). The expansion rate is dominated by this dark energy term for  $z < 0.7$ , so cosmological analysis of the ASKAP FRBs is not sensitive to the precise value of  $\Omega_m$  and, therefore, to a close approximation,  $\langle \text{DM}_{\text{cosmic}} \rangle \propto \Omega_b H_0$ . We therefore proceed to place a constraint on this product.

To construct a likelihood function  $\mathcal{L}$  from our FRB measurements, we build a model for  $\text{DM}_{\text{cosmic}}$  and its uncertainty. The model is based primarily on the cosmological parameters, but it must also allow for a nuisance parameter which accounts for the DM of our Galactic halo and that of the host galaxy:  $\text{DM}_{\text{MW,halo}} + \text{DM}_{\text{host}}$ . For the former term, theoretical models informed by observation suggest  $\text{DM}_{\text{MW,halo}} \approx 50 \text{ pc cm}^{-3}$  with a small dispersion<sup>13,56</sup>, but we acknowledge that the mean value is poorly constrained. We expect the variance in these terms to be driven by  $\text{DM}_{\text{host}}$ , which follows from the large range in DM values observed for the ISM of our Galaxy<sup>22</sup>, even if one ignores whether FRBs occur in ‘special’ locations within a galaxy. Furthermore, the very high RM and the (probably related) large DM excess of FRB 121102 above  $\text{DM}_{\text{cosmic}}$  implies at least one FRB with a large  $\text{DM}_{\text{host}}$  value<sup>37</sup>.

The PDF for  $\text{DM}_{\text{host}}$  has limited theoretical motivation. In the following, we assume a log-normal distribution which has two salient features: (1) it is positive definite; (2) it exhibits an asymmetric tail to large values. The latter property allows for high  $\text{DM}_{\text{host}}$  values that might arise from gas local to the FRB, for example, an H II region or circumstellar medium. Formally, we adopt a log-normal distribution:

$$p_{\text{host}}(\text{DM}_{\text{host}} | \mu, \sigma_{\text{host}}) = \frac{1}{(2\pi)^{1/2} \text{DM}_{\text{host}} \sigma_{\text{host}}} \exp \left[ -\frac{(\log \text{DM}_{\text{host}} - \mu)^2}{2\sigma_{\text{host}}^2} \right] \quad (3)$$

This distribution has a median value of  $e^\mu$  and variance  $e^{\mu+2\sigma_{\text{host}}^2/2} (e^{\sigma_{\text{host}}^2} - 1)^{1/2}$ . We consider distributions with  $e^\mu$  in the range  $20-200 \text{ pc cm}^{-3}$  and  $\sigma_{\text{host}}$  in the range  $0.2-2.0$ . An illustrative set of these probability distribution functions for  $\text{DM}_{\text{host}}$  is shown in Extended Data Fig. 2. For consistency of interpretation of  $\text{DM}_{\text{host}}$  values from bursts at disparate redshifts, the probability distribution function is referenced to the rest frame of the host galaxy, so a correction  $\text{DM}_{\text{host}} \rightarrow \text{DM}_{\text{host}}(1+z_{\text{FRB}})^{-1}$  is applied and the distribution normalized accordingly, however, in practice this redshift correction factor varies only over the range 0.7 to 0.9 in the gold-standard sample. The inferred dispersion in  $\text{DM}_{\text{host}}$  is consistent with the expected range of host DMs given the galaxy type, morphology and orientation on the sky and the distance of the FRB from the galaxy centre. However, we are unable

to state more than this at present, and remark that present estimates of the  $\text{DM}_{\text{host}}$  contributions towards specific localized FRBs<sup>15,18</sup> with two quite different host galaxies are, respectively, in the range  $30-81 \text{ pc cm}^{-3}$  and  $<70 \text{ pc cm}^{-3}$ . This suggests that any correction on this basis could be small for our sample. A further interesting aspect of our measurements is that it is beginning to place limits on these corrections.

Altogether,  $\text{DM}_{\text{FRB}}^{\text{model}} = \text{DM}_{\text{cosmic}}(z) + \text{DM}_{\text{host}} + \text{DM}_{\text{MW,ISM}}$  with the last quantity estimated from NE2001 based on the FRB coordinates; given the high Galactic latitudes of the present ASKAP sample we adopt a value  $\text{DM}_{\text{MW,ISM}} = 30 \text{ pc cm}^{-3}$  for these bursts. The mechanics of our treatment of the  $\text{DM}_{\text{host}}$ ,  $\text{DM}_{\text{MW,halo}}$  and  $\text{DM}_{\text{MW,ISM}}$  terms is described in greater detail in equation (6).

The model probability distribution for  $\text{DM}_{\text{cosmic}}$  is derived from theoretical treatments of the IGM and galaxy haloes<sup>11,13</sup> with  $\sigma_{\text{DM}}$  dominated by the physical variance in  $\text{DM}_{\text{cosmic}}$ . Extended Data Fig. 3 shows that comparison against the analytic form (as used in other IGM-related contexts<sup>60</sup>)

$$p_{\text{cosmic}}(\Delta) = A \Delta^{-\beta} \exp \left[ -\frac{(\Delta^{-\alpha} - C_0)^2}{2\alpha^2 \sigma_{\text{DM}}^2} \right], \Delta > 0 \quad (4)$$

provides an excellent match to the  $\text{DM}_{\text{cosmic}}$  distributions observed in our semi-analytic models and in a hydrodynamic simulation, where  $\Delta \equiv \text{DM}_{\text{cosmic}} / \langle \text{DM}_{\text{cosmic}} \rangle$ . The motivation for this form is that in the limit of small  $\sigma_{\text{DM}}$ , the distribution of DM should approach a Gaussian owing to the Gaussianity of structure on large scales (a non-negligible component of the variance of  $p_{\text{cosmic}}(\Delta)$  comes from tens of megaparsec structures) and in the low- $\sigma_{\text{DM}}$  limit the halo gas is more diffuse and so the PDF approaches a Gaussian owing to the intersection of the line of sight with more structures. Conversely, when the variance is large, this PDF captures the large skew that results from a few large structures that contribute to the DM of many sightlines. The sharp low-DM cutoff in the distribution reflects the fact that a large component of the IGM is highly diffuse, and displays much less variance than the halo-related component, thus imposing a strict lower limit to the DM. The parameter  $\beta$  is related to the inner density profile of gas in haloes. If the 3D density profile scales as  $\rho \propto r^{-\alpha}$ ,  $\beta = (\alpha + 1)/(\alpha - 1)$  such that an isothermal profile with  $\alpha = 2$  has  $\beta = 3$  and an inner slope of  $\alpha = 1.5$  has  $\beta = 5$ . Such slopes are consistent with those found in numerical simulations of intrahalo gas<sup>61,62</sup>. The indices  $\alpha = 3$  and  $\beta = 3$  provide the best match to our models (although we find that  $p_{\text{cosmic}}(\Delta)$  is weakly sensitive to order unity changes in these parameters, with  $\beta = 3$  having the most flexibility for our  $z = 0.11$  measurement relative to  $\beta = 4$ ). We use the parameter  $\sigma_{\text{DM}}$  in  $p_{\text{cosmic}}(\Delta)$  as an effective standard deviation even though formally the standard deviation with  $\beta = 3$  diverges logarithmically. We find that  $\sigma_{\text{DM}}$  is closely tied to the true standard deviation when imposing motivated maximum cutoffs for  $\Delta$  on the distribution. The mean of the distribution requires that  $\langle \Delta \rangle = 1$ , which fixes the remaining parameter  $C_0$  in  $p_{\text{cosmic}}(\Delta)$ .

Extended Data Figure 3 shows models that use equation (4) for  $p_{\text{cosmic}}$  relative to numerical calculations at redshifts that span the considered range. The solid curves are the previously described semi-analytic models<sup>11</sup>, which assume that haloes below the specified mass have been evacuated of gas, and the ‘swinds’ simulation of ref. <sup>63</sup>. The dashed curves show the function evaluated for the best-fit  $\sigma_{\text{DM}}$ , and the dot-dashed curves adopt the parameterization  $\sigma_{\text{DM}} = Fz^{-0.5}$  and scale off the  $z = 0.5$  best fit value for  $\sigma_{\text{DM}}$  yielding  $F$  of 0.09, 0.15 and 0.32 in our semi-analytic models in which haloes of  $10^{14}$ ,  $10^{13}$  and  $10^{11}$  solar masses ( $M_\odot$ ) are evacuated of their gas. The agreement of the dot-dashed curves with the solid numerical model curves demonstrates that  $\sigma_{\text{DM}} = Fz^{-0.5}$  approximates the evolution over the range of our measurements. This scaling is further motivated in the Euclidean limit, applicable for  $z \ll 1$ , where  $\langle \text{DM}_{\text{cosmic}} \rangle = n_e c z / H$  and  $\sigma_{\text{DM}} \approx \text{DM}_{\text{halo}} \sqrt{N} / \langle \text{DM}_{\text{cosmic}} \rangle$ , where  $n_e$  is the mean electron density and



$N$  is the number of haloes intersected, which is proportional to the path length probed or  $cz/H$ .

While our analytic parameterization describes the distribution of  $DM_{\text{cosmic}}$  both in semi-analytic models and numerical simulations, we use the more flexible semi-analytic models to set the marginalization range in  $F$  that is used for some constraints on  $\Omega_b h_{70}$ . Here we argue that the considered semi-analytic models shown in Extended Data Fig. 3 span the likely range of possible feedback scenarios. These models approximate haloes as retaining their gas in a manner that traces the dark matter above some mass threshold. This approximates the picture in many simulations<sup>61,62</sup> and analytic models<sup>64,65</sup> in which the fraction of halo gas retained is a strongly increasing function of halo mass before saturating at unity. Furthermore, gas that is outside haloes is less effective at contributing variance: take the example where gas is distributed out to a distance  $R$  around a halo. The probability a sight-line intersects this gas scales as  $R^2$ , leading to less shot noise for larger  $R$ , while the contribution of each individual system scales as  $R^{-2}$ , leading to a smaller contribution for larger  $R$ . This picture motivates the semi-analytic model's approximation that ejected gas diffusely traces large-scale structure<sup>11</sup>.

Simulations and models generally find that haloes below threshold masses in the range of about  $(10^{11} - 10^{13})M_\odot$  are evacuated of gas<sup>61,62,64,65</sup>, although some implementations of stellar quasar feedback can result in different predictions<sup>66</sup>. Halo gas in  $M > 10^{14}M_\odot$  haloes is constrained by X-ray observations to mostly reside within such haloes<sup>67</sup>. Our strongest feedback model, in which  $F = 0.09$ , pushes up against this observational limit. Our model with the weakest feedback assumes that dwarf-galaxy-sized haloes with  $10^{11}M_\odot$  retain their gas and yields  $F = 0.32$  (and we find that  $F$  is just marginally larger if the  $10^{10}M_\odot$  haloes of the smallest dwarf galaxies retain their gas, haloes that would be just massive enough to overcome the pressure of the IGM and retain their gas<sup>68</sup>). Thus our models span the range of likely feedback scenarios.

Given this semi-analytic formalism, we proceed to estimate the model likelihood by computing the joint likelihoods of all FRBs:

$$\mathcal{L} = \prod_{i=1}^{N_{\text{FRBs}}} P_i(DM'_{\text{FRB}}|z_i) \quad (5)$$

where  $P_i(DM'_{\text{FRB}}|z_i)$  is the probability of the total observed  $DM_{\text{FRB}}$  corrected for the Galaxy:

$$DM'_{\text{FRB}} \equiv DM_{\text{FRB}} - DM_{\text{MW,ISM}} - DM_{\text{MW,halo}} = DM_{\text{host}} + DM_{\text{cosmic}} \quad (6)$$

For a burst at a given  $z_i$  and the model parameters we have:

$$P_i(DM'_{\text{FRB},i}|z_i) = \int_0^{DM'_{\text{FRB}}} p_{\text{host}}(DM_{\text{host}}|\mu, \sigma_{\text{host}}) p_{\text{cosmic}}(DM'_{\text{FRB},i} - DM_{\text{host}}, z_i) dDM_{\text{host}} \quad (7)$$

and  $p_{\text{host}}(DM_{\text{host}}|\mu, \sigma_{\text{host}})$  the PDF for  $DM_{\text{host}}$ . With the likelihood function defined we construct a grid of  $\Omega_b h_{70}$ ,  $F$ ,  $\mu$  and  $\sigma_{\text{host}}$  values and marginalize over the last three to obtain the constraint on  $\Omega_b h_{70}$ . These results are presented in Fig. 3 in the main text for the gold-standard sample. Extended Data Fig. 4 presents the results of the same analysis when FRBs 190523 and 190611 are included in the data set.

To place confidence intervals on  $\Omega_b h_{70}$  and  $\sigma_{\text{DM}}$ , we use the likelihood ratio test statistic  $\mathcal{D}$ :

$$\mathcal{D}(\Omega_b h_{70}, F, \mu, \sigma_{\text{host}}) = 2 \log \mathcal{L}_{\text{max}} - 2 \log \mathcal{L}(\Omega_b h_{70}, F, \mu, \sigma_{\text{host}}) \quad (8)$$

where  $\mathcal{L}_{\text{max}}$  is the maximum value of  $\mathcal{L}$ , that is, for parameters maximizing the likelihood. According to Wilks' theorem, for a sufficiently large number of FRBs,  $\mathcal{D}$  will be distributed according to a  $\chi^2_n$  distribution

with  $n = 4$  degrees of freedom<sup>69</sup>. If the cumulative distribution function of the  $\chi^2_4(x)$  distribution is  $\text{CDF}(x)$ , solving  $\text{CDF}(x) = p$  constrains the  $\Omega_b h_{70}, F, \mu, \sigma_{\text{host}}$  parameter space to the region  $\mathcal{D} \leq x$  at confidence level (C.L.)  $p$ .

Uncertainties in these confidence estimates are probably dominated by systematic effects in the sample selection, and small number statistics. To test both, we extend the gold-standard sample of five bursts to include FRBs 190523 and 190611. The resulting analysis is shown in Extended Data Fig. 4. Compared with Fig. 3, the inclusion of two further bursts shifts the maximum-likelihood estimate for  $\Omega_b h_{70}$  at 68% C.L. from  $0.051^{+0.014}_{-0.015}$  to  $0.042^{+0.011}_{-0.012}$ , that is, consistent with the original uncertainties. This does not mean that there is no systematic bias, nor that Wilks' theorem holds precisely for our sample, but rather that any such effects are minor compared to the inherent uncertainties from our small sample size of localized bursts.

### Accounting for biases in $P(DM, z)$

The cosmological evolution of the FRB population, and its intrinsic luminosity function, can strongly influence the observed/expected distribution of FRBs in redshift-DM space<sup>70</sup>,  $P(DM, z)$ . We therefore perform our likelihood maximization over  $P(DM|z)$  only. This discards the information contained in the redshifts of our detected FRBs, but makes the procedure more robust against factors influencing the redshift distribution.

The remaining bias comes from changing sensitivity as a function of DM. This can be either direct, through DM-smearing within frequency channels, or indirect, through increased scatter broadening associated with the same gaseous structures causing the observed DM.

We wish to compute the dispersion measure limit,  $DM_{\text{cutoff}}$ , at which a given FRB would have been undetectable. The  $S/N$  of a detected burst depends on its intrinsic (or scatter-broadened) width,  $w$ , the time resolution of the detection system,  $t_{\text{res}}$ , and the amount of dispersion measure time smearing between adjacent 1 MHz spectral channels,  $t_{\text{smear}}(DM)$ . The resulting width of the pulse is:

$$\Delta t_{\text{obs}}(DM) = \sqrt{w^2 + t_{\text{res}}^2 + t_{\text{smear}}^2(DM)} \quad (9)$$

We compute  $DM_{\text{cutoff}}$  such that the burst, detected by our system with a signal-to-noise ratio of  $s_0$  at a  $DM_{\text{obs}}$  would have fallen below our detection threshold of  $s_d = 9.0\sigma$ . For each burst we thus solve

$$s_d = s_0 \sqrt{\frac{\Delta t_{\text{obs}}(DM_{\text{obs}})}{\Delta t_{\text{obs}}(DM_{\text{cutoff}})}} \quad (10)$$

Extended Data Table 1 lists the DM, widths, time resolution, detection  $S/N$  values and derived  $DM_{\text{cutoff}}$  values for each of the bursts in our sample.

### MCMC analysis

To complement the likelihood analysis presented in the main text, we have performed Bayesian inference of a model constructed to describe the DM and redshift measurements of the FRBs. The model consists of four parameters describing two PDFs for distinct components of the dispersion measure: (i)  $DM_{\text{cosmic}}$ , which describes the extragalactic dispersion measure including both the diffuse IGM and the gas associated with intervening galactic haloes; and (ii)  $DM_{\text{host}}$ , which describes ionized gas associated with the host galaxy (we assume a fixed  $DM_{\text{MW,halo}}$  value of  $50 \text{ pc cm}^{-3}$  for the Galactic halo). We parameterize the former PDF with equation (4), that is,  $p_{\text{cosmic}}(\Delta)$  with  $\Delta \equiv DM_{\text{cosmic}}/\langle DM_{\text{cosmic}} \rangle$  and  $\langle DM_{\text{cosmic}} \rangle$  the average value for the assumed cosmology (equation (2)). The foregoing subsection on cosmology and host-galaxy parameter estimation describes theoretical treatments that motivate one to adopt  $\alpha = 3$  and  $\beta = 3$  in equation (4) and to adopt the functional form of  $\sigma_{\text{DM}} = F/z^{1/2}$  for its dispersion parameter. For  $\langle DM_{\text{cosmic}} \rangle$ , we modulate

its amplitude via the product  $\Omega_b h_{70}$ . Therefore,  $p_{\text{cosmic}}(\Delta)$  is governed by two free parameters:  $F$  and  $\Omega_b h_{70}$ .

We adopt the same  $p_{\text{host}}(\text{DM}_{\text{host}}, \mu, \sigma_{\text{host}})$  PDF described earlier, with free parameters  $\exp(\mu)$  and  $\sigma_{\text{host}}$ . From these two PDFs we construct a likelihood function for the set of observed FRBs using equations (5) and (7). Note that measurement uncertainty in  $\text{DM}_{\text{FRB}}$  does not enter into the evaluation of  $\mathcal{L}$  because the dispersion from  $\text{DM}_{\text{cosmic}}$  and  $\text{DM}_{\text{host}}$  are much greater. Put another way, our model is constructed to describe the observed distribution of  $\text{DM}_{\text{FRB}}$  values with an anticipated dispersion substantially exceeding the uncertainty in individual  $\text{DM}_{\text{FRB}}$  measurements (typically  $<1 \text{ pc cm}^{-3}$ ).

Effectively, two of the parameters ( $\Omega_b h_{70}, \mu$ ) set the amplitude of the  $\text{DM}-z$  relation and two describe its dispersion ( $F, \sigma_{\text{host}}$ ). We anticipate a degeneracy between each set although if  $\text{DM}_{\text{host}}$  is approximately independent of redshift then this apparent degeneracy may be resolved. Only the dispersion in  $\text{DM}_{\text{cosmic}}$ , parameterized by  $F$ , allows for large negative excursions from the mean relation. Lastly, we introduce priors for the four parameters based on a combination of experimentation, physical expectation and scientific motivation. For  $\Omega_b h_{70}$ , the scientific focus of this manuscript, we adopt a uniform prior ranging from 0.015–0.095 which easily spans the Planck15 estimate. For  $F$ , we adopt a uniform prior in the interval (0.01, 0.5), a larger range than anticipated by our models in the frequentist analysis presented above. Regarding  $\exp(\mu)$ , we adopt a uniform prior in the interval [20, 200]  $\text{pc cm}^{-3}$ . We consider lower values for the mean to be non-physical and we will find that larger values are disfavoured by the observations. Lastly, we assume a uniform prior for  $\sigma_{\text{host}}$  in the interval [0.2, 2]. The larger  $\sigma_{\text{host}}$  values give non-negligible probability for  $\text{DM}_{\text{host}}$  values in excess of 1,000  $\text{pc cm}^{-3}$ . Future observations, especially an ensemble of FRBs at low redshift, will better inform these priors on  $\mu$  and  $\sigma_{\text{host}}$ .

Adopting the above likelihood and priors, we performed a Bayesian inference of the four parameters using the gold-standard sample of FRB measurements and standard MCMC techniques. These were performed with the PYMC3 software package using slice sampling and four independent chains of 40,000 samples after a tuning period of 2,000 samples. Extended Data Fig. 5 presents a corner plot of the combined samples. A principal result is that the data yield a  $\Omega_b h_{70}$  distribution fully consistent with the independent estimates from the CMB, BBN and supernovae. Quantitatively, the  $\Omega_b h_{70}$  samples have a median value of 0.056 and a 68% confidence interval spanning [0.046, 0.066] (see Extended Data Table 2). Taken strictly, at 95% confidence these FRB measurements require a universe with at least 70% of the baryons inferred from BBN and CMB analysis. These results hold despite the weak priors placed on the PDF for  $\text{DM}_{\text{host}}$ , but we warn that they are dependent on the value assumed for  $\text{DM}_{\text{MW,halo}}$ .

Extended Data Figure 5 also reveals the anticipated anti-correlations between  $\mu$  and  $\Omega_b h_{70}$  and (to a lesser extent)  $F$  and  $\sigma_{\text{host}}$ . We expect these to weaken as the FRB sample grows in size and redshift range. Lastly, we note that the  $F$  and  $\mu$  parameters have maximal probability at one edge of their assumed prior intervals. Values of  $F$  that are on the higher side of the considered range (a range that spans the possible model space) are modestly favoured. For  $\mu$ , we consider 20  $\text{pc cm}^{-3}$  to be the lowest sensible mean contribution from the host galaxy (which could also mean a lower value for the Galactic halo than adopted here).

The frequentist analysis in the main text and this Bayesian MCMC analysis agree very well on the gold-standard sample. The most notable differences are that the MCMC analysis prefers a distribution for  $\exp(\mu)$  that is more peaked to smaller  $\exp(\mu)$  values and one for  $F$  that peaks towards larger values, although with no value for  $\exp(\mu)$  or  $F$  strongly preferred by either analysis. When the parameters are not well constrained one would not expect perfect agreement between the methods, as, for example, the Bayesian analysis is sensitive to our prior on  $\exp(\mu)$  when this parameter is not well constrained. It is expected that the differences between the two methods will become smaller with more data. Already for the seven-burst sample (Extended Data Fig. 4),

the distribution for  $F$  in the frequentist analysis is more similar to the MCMC analysis of the gold-standard sample.

## Data availability

The datasets generated during and/or analysed during this study are available at <https://data-portal.hpc.swin.edu.au/dataset/observations-of-four-localised-fast-radio-bursts-and-their-host-galaxies>.

## Code availability

Custom code is available at <https://github.com/FRBs/FRB>.

- Mahony, E. K. et al. A search for the host galaxy of FRB 171020. *Astrophys. J.* **867**, L10 (2018).
- Connor, L. Interpreting the distributions of FRB observables. *Mon. Not. R. Astron. Soc.* **487**, 5753–5763 (2019).
- Schnitzler, D. H. F. M. Modelling the Galactic distribution of free electrons. *Mon. Not. R. Astron. Soc.* **427**, 664–678 (2012).
- Yao, J. M., Manchester, R. N. & Wang, N. A new electron-density model for estimation of pulsar and FRB distances. *Astrophys. J.* **835**, 29 (2017).
- Oppermann, N. et al. An improved map of the Galactic Faraday sky. *Astron. Astrophys.* **542**, A93 (2012).
- Akabori, T. & Ryu, D. Faraday rotation measure due to the intergalactic magnetic field. *Astrophys. J.* **723**, 476–481 (2010).
- Bernet, M. L., Miniati, F., Lilly, S. J., Kronberg, P. P. & Dessauges-Zavadsky, M. Strong magnetic fields in normal galaxies at high redshift. *Nature* **454**, 302–304 (2008).
- Ravi, V. et al. The magnetic field and turbulence of the cosmic web measured using a brilliant fast radio burst. *Science* **354**, 1249–1252 (2016).
- Bhat, N. D. R., Cordes, J. M., Camilo, F., Nice, D. J. & Lorimer, D. R. Multifrequency observations of radio pulse broadening and constraints on interstellar electron density microstructure. *Astrophys. J.* **605**, 759–783 (2004).
- Krishnakumar, M. A., Mitra, D., Naidu, A., Joshi, B. C. & Manoharan, P. K. Scatter broadening measurements of 124 pulsars at 327 MHz. *Astrophys. J.* **804**, 23 (2015).
- Michilli, D. et al. An extreme magneto-ionic environment associated with the fast radio burst source FRB 121102. *Nature* **553**, 182–185 (2018).
- Deller, A. T. et al. DiFX-2: a more flexible, efficient, robust, and powerful software correlator. *Publ. Astron. Soc. Pacif.* **123**, 275–287 (2011).
- Kettenis, M., van Langevelde, H. J., Reynolds, C. & Cotton, B. ParselTongue: AIPS talking Python. In *Astronomical Data Analysis Software and Systems XV* (eds Gabriel, C., Arviset, C., Ponz, D. & Enrique, S.) 497–500 (ASP Conf. Ser. **Vol. 351**, Astronomical Society of the Pacific, 2006).
- Greisen, E. W. AIPS, the VLA, and the VLBA. In *Information Handling in Astronomy — Historical Vistas* (ed. Heck, A.) 109–125 (Astrophysics and Space Science Library Vol. 285, Kluwer Academic Publishers, 2003).
- Becker, R. H., White, R. L. & Helfand, D. J. The FIRST survey: faint images of the radio sky at twenty centimeters. *Astrophys. J.* **450**, 559 (1995).
- Appenzeller, I. et al. Successful commissioning of FORS1 — the first optical instrument on the VLT. *Messenger* **94**, 1–6 (1998).
- PyPelt Documentation Release 1.0.1. [https://pypelit.readthedocs.io/\\_/downloads/en/latest/pdf/](https://pypelit.readthedocs.io/_/downloads/en/latest/pdf/) (2020).
- Abazajian, K. N. et al. The seventh data release of the Sloan Digital Sky Survey. *Astrophys. J. Suppl. Ser.* **182**, 543–558 (2009).
- Prochaska, J. X. The igmspec database of public spectra probing the intergalactic medium. *Astron. Comput.* **19**, 27–33 (2017).
- Vernet, J. et al. X-shooter, the new wide band intermediate resolution spectrograph at the ESO Very Large Telescope. *Astron. Astrophys.* **536**, A105 (2011).
- Gimeno, G. et al. On-sky commissioning of Hamamatsu CCDs in GMOS-S. *Proc. SPIE* **9908**, 99082S (2016).
- Freudling, W. et al. Automated data reduction workflows for astronomy. The ESO Reflex environment. *Astron. Astrophys.* **559**, A96 (2013).
- Berriman, G. B. & Good, J. C. The application of the Montage Image Mosaic Engine to the visualization of astronomical images. *Publ. Astron. Soc. Pacif.* **129**, 058006 (2017).
- Lang, D., Hogg, D. W., Mierle, K., Blanton, M. & Roweis, S. Astrometry.net: blind astrometric calibration of arbitrary astronomical images. *Astron. J.* **139**, 1782–1800 (2010).
- Lindgren, L. et al. Gaia Data Release 2: the astrometric solution. *Astron. Astrophys.* **616**, A2 (2018).
- Abbott, T. M. C. et al. The Dark Energy Survey data release 1. *Astrophys. J. Suppl. Ser.* **239**, 25 (2018).
- Craig, M. et al. astropy/ccdproc: v1.3.0, post1 <https://doi.org/10.5281/zenodo.1069648> (2017).
- Wright, E. A cosmology calculator for the World Wide Web. *Publ. Astron. Soc. Pacif.* **118**, 1711–1715 (2006).
- Inoue, S. Probing the cosmic reionization history and local environment of gamma-ray bursts through radio dispersion. *Mon. Not. R. Astron. Soc.* **348**, 999–1008 (2004).
- Deng, W. & Zhang, B. Cosmological implications of fast radio burst/gamma-ray burst associations. *Astrophys. J.* **783**, L35 (2014).
- Madau, P. & Dickinson, M. Cosmic star-formation history. *Annu. Rev. Astron. Astrophys.* **52**, 415–486 (2014).
- Fukugita, M. & Peebles, P. J. E. The cosmic energy inventory. *Astrophys. J.* **616**, 643–668 (2004).

59. Walters, A., Ma, Y.-Z., Sievers, J. & Weltman, A. Probing diffuse gas with fast radio bursts. *Phys. Rev. D* **100**, 103519 (2019).
60. Miralda-Escudé, J., Haehnelt, M. & Rees, M. J. Reionization of the inhomogeneous universe. *Astrophys. J.* **530**, 1–16 (2000).
61. Hafen, Z. et al. The origins of the circumgalactic medium in the FIRE simulations. *Mon. Not. R. Astron. Soc.* **488**, 1248–1272 (2019).
62. Fielding, D., Quataert, E., McCourt, M. & Thompson, T. A. The impact of star formation feedback on the circumgalactic medium. *Mon. Not. R. Astron. Soc.* **466**, 3810–3826 (2017).
63. Faucher-Giguère, C.-A., Kereš, D. & Ma, C.-P. The baryonic assembly of dark matter haloes. *Mon. Not. R. Astron. Soc.* **417**, 2982–2999 (2011).
64. Sharma, P., McCourt, M., Parrish, I. J. & Quataert, E. On the structure of hot gas in haloes: implications for the  $L_X$ - $T_X$  relation and missing baryons. *Mon. Not. R. Astron. Soc.* **427**, 1219–1228 (2012).
65. Voit, G. M. Ambient column densities of highly ionized oxygen in precipitation-limited circumgalactic media. *Astrophys. J.* **880**, 139 (2019).
66. Jaroszynski, M. Fast radio bursts and cosmological tests. *Mon. Not. R. Astron. Soc.* **484**, 1637–1644 (2019).
67. Debackere, S. N. B., Schaye, J. & Hoekstra, H. The impact of the observed baryon distribution in haloes on the total matter power spectrum. *Mon. Not. R. Astron. Soc.* **492**, 2285–2307 (2020).
68. Noh, Y. & McQuinn, M. A physical understanding of how reionization suppresses accretion on to dwarf haloes. *Mon. Not. R. Astron. Soc.* **444**, 503–514 (2014).
69. Wilks, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60–62 (1938). <https://doi.org/10.1214/aoms/1177732360>.
70. Macquart, J. P. & Ekers, R. FRB event rate counts — II. Fluence, redshift, and dispersion measure distributions. *Mon. Not. R. Astron. Soc.* **480**, 4211–4230 (2018).

**Acknowledgements** We thank H. Yang and L. Infante for providing IMACS imaging around FRB 190611 that informed the further follow-up observations on VLT and Gemini-S presented here. We are grateful to Australia Telescope National Facility (ATNF) operations staff and to Murchison Radio-astronomy Observatory staff for supporting our ASKAP operations, and the ATNF steering committee for allocating time for these observations. K.W.B., J.-P.M. and R.M.S. acknowledge support by Australian Research Council (ARC) grant DP180100857. A.T.D. and R.M.S. are the recipients of ARC Future Fellowships FT150100415 and FT190100155 respectively. S.O. and R.M.S. acknowledge support through ARC grant FL150100148. R.M.S. also acknowledges support through ARC grant CE170100004. N.T. acknowledges support from FONDECYT grant 11191217 and PUCV/VRIEA project 039.395/2019. The Australian Square Kilometre Array Pathfinder and Australia Telescope Compact Array are part of the Australia

Telescope National Facility which is managed by CSIRO. Operation of ASKAP is funded by the Australian Government with support from the National Collaborative Research Infrastructure Strategy. ASKAP uses the resources of the Pawsey Supercomputing Centre. Establishment of ASKAP, the Murchison Radio-astronomy Observatory and the Pawsey Supercomputing Centre are initiatives of the Australian Government, with support from the Government of Western Australia and the Science and Industry Endowment Fund. Part of this work was performed on the OzSTAR national facility at Swinburne University of Technology. OzSTAR is funded by Swinburne University of Technology and the National Collaborative Research Infrastructure Strategy (NCRIS). We acknowledge the Wajarri Yamatji as the traditional owners of the Murchison Radio-astronomy Observatory site. This work includes observations collected at the European Southern Observatory under ESO programmes 0102.A-0450(A), 0103.A-0101(A) and 0103.A-1010(B). This work includes data obtained from programme GS-2019B-Q-132 at the Gemini Observatory, acquired through the Gemini Observatory Archive and processed using the Gemini PYRAF package. Gemini Observatory is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science Foundation (US), the National Research Council (Canada), CONICYT (Chile), Ministerio de Ciencia, Tecnología e Innovación Productiva (Argentina), Ministério da Ciência, Tecnologia e Inovação (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea). PYRAF is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

**Author contributions** J.X.P., M.M. and J.P.M. framed the analysis approach and drafted the manuscript, with contributions from A.T.D., R.D.E. and R.M.S. K.W.B. developed the detection and localization pipelines, and with A.T.D. and C.P. jointly developed the correlation code and interferometry processing pipeline. D.R.S. made additional improvements to the performance of the detection pipeline. R.M.S. and S.B. detected the FRBs and performed follow-up astrometry. A.T.D. and C.K.D. derived the FRB positions from the CRAFT voltage data, and S.R., J.X.P., N.T. and L.M. obtained and reduced the optical data to derive the burst host-galaxy identifications and redshifts. M.M., J.X.B. and J.-P.M. developed the IGM model and analysis code, and S.O. adapted the code to run on the Swinburne supercomputer and collated the results. C.W.J. contributed to the maximum likelihood analysis and framed the approach to estimate parameter uncertainties.

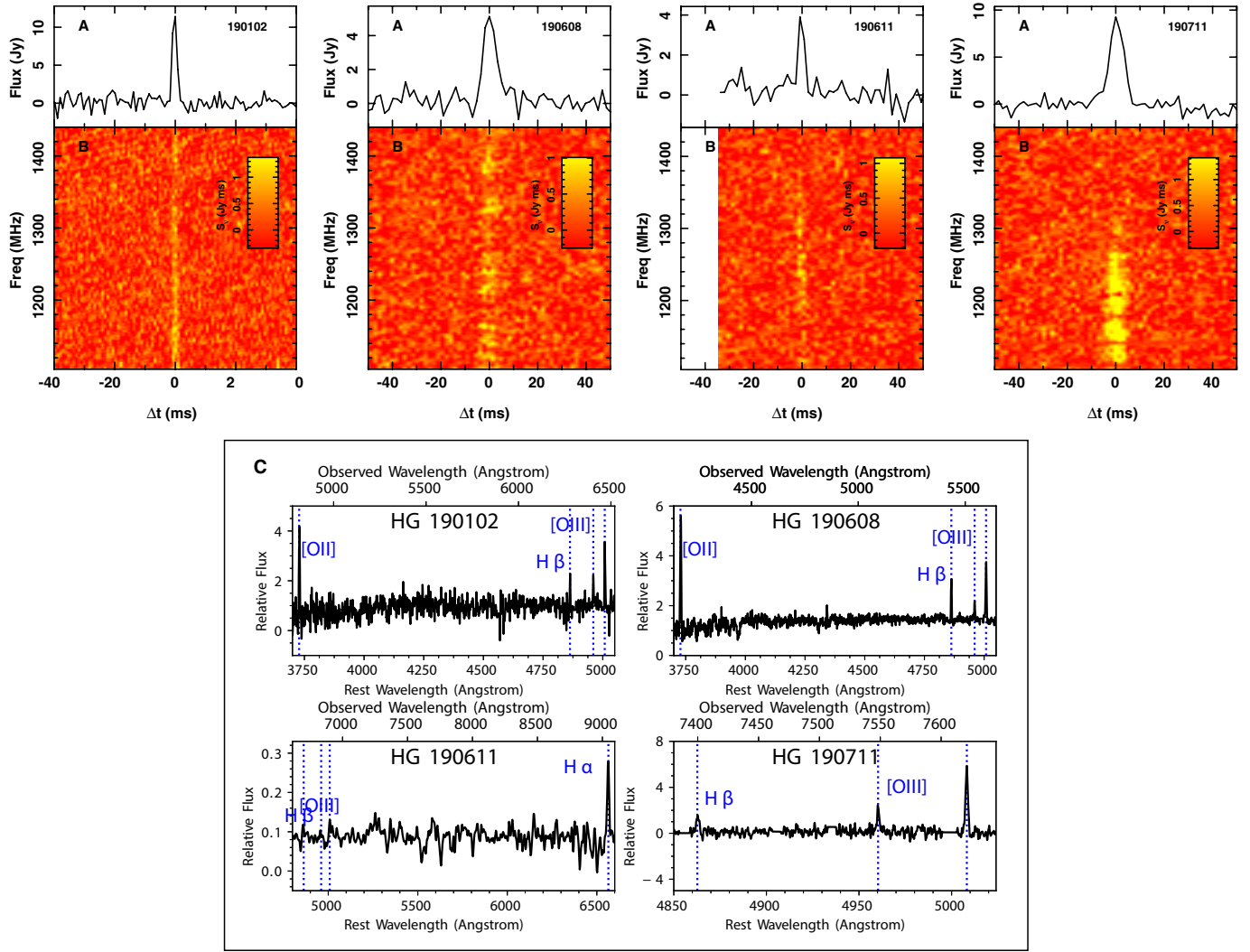
**Competing interests** The authors declare no competing interests.

#### Additional information

**Correspondence** and requests for materials should be addressed to J.-P.M. or J.X.P.

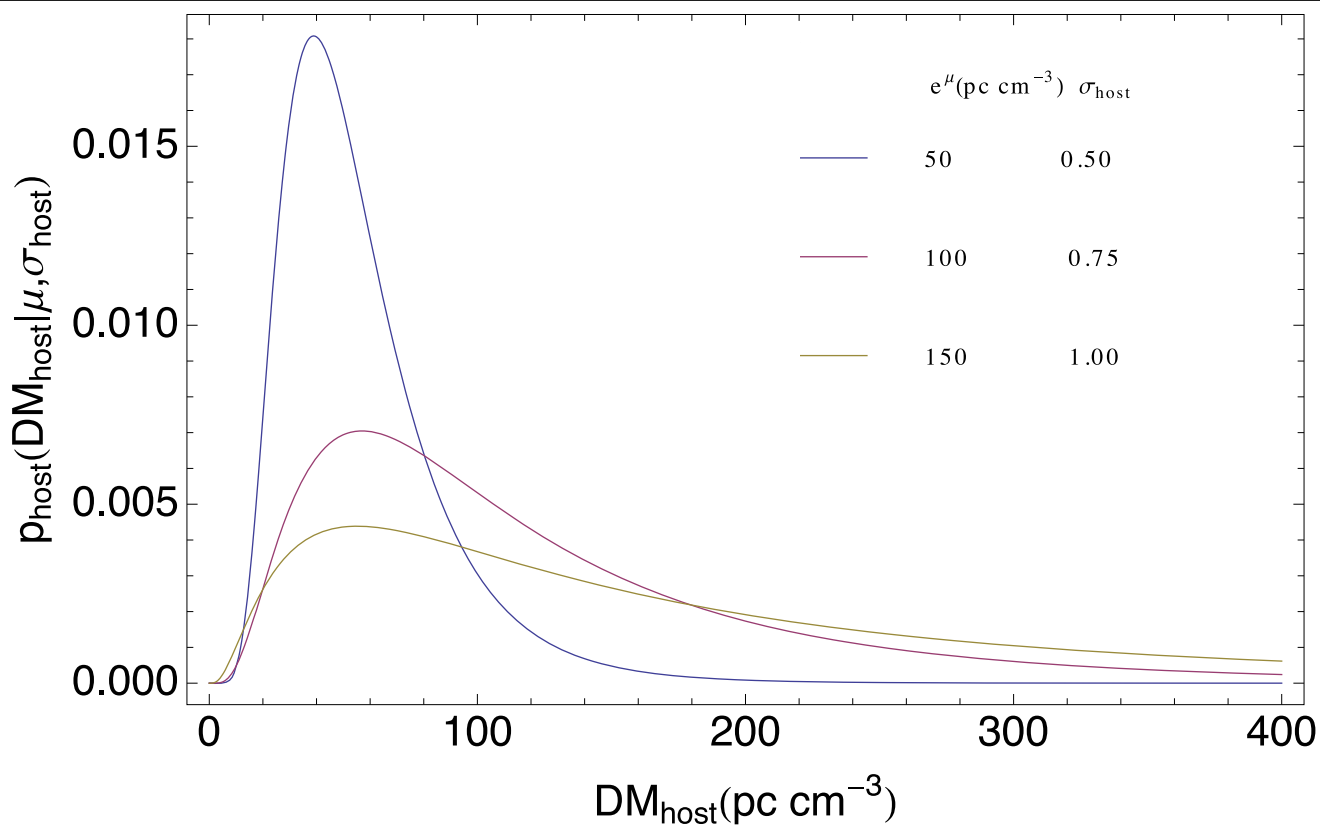
**Peer review information** *Nature* thanks Kiyoshi Masu, Fabrizio Nicastro and Anthony Walters for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

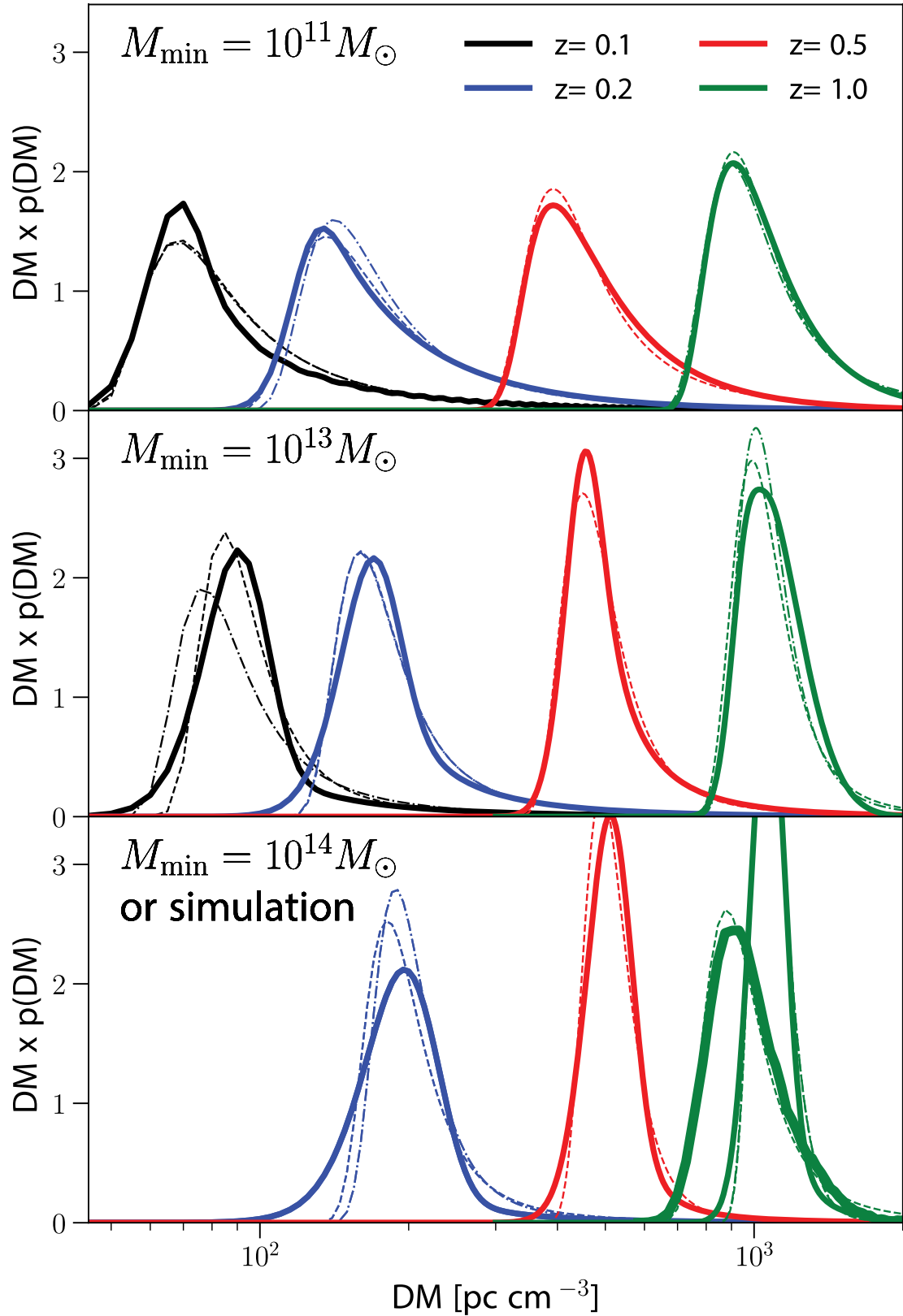


**Extended Data Fig. 1 | The pulse profiles and host galaxy spectra of the four new FRBs presented here.** Top row, ASKAP data. The pulse profiles (upper subpanels, labelled A) and the radio dynamic spectra (lower subpanels, labelled B) show the detections by the ASKAP incoherent capture system (ICS) of FRB 190102 with a time resolution of 0.864 ms, and of FRBs 190608, 190611

and 190711 with a resolution of 1.728 ms. The spectral resolution is 1 MHz across the 336-MHz bandwidth. Bottom row, the SDSS (HG 190608) and VLT/FORS2 (HG 190102, HG 190611 and HG 190711) optical spectra of the host galaxies located at the respective FRB positions (see Table 1), and the spectral lines from which their redshifts are deduced.



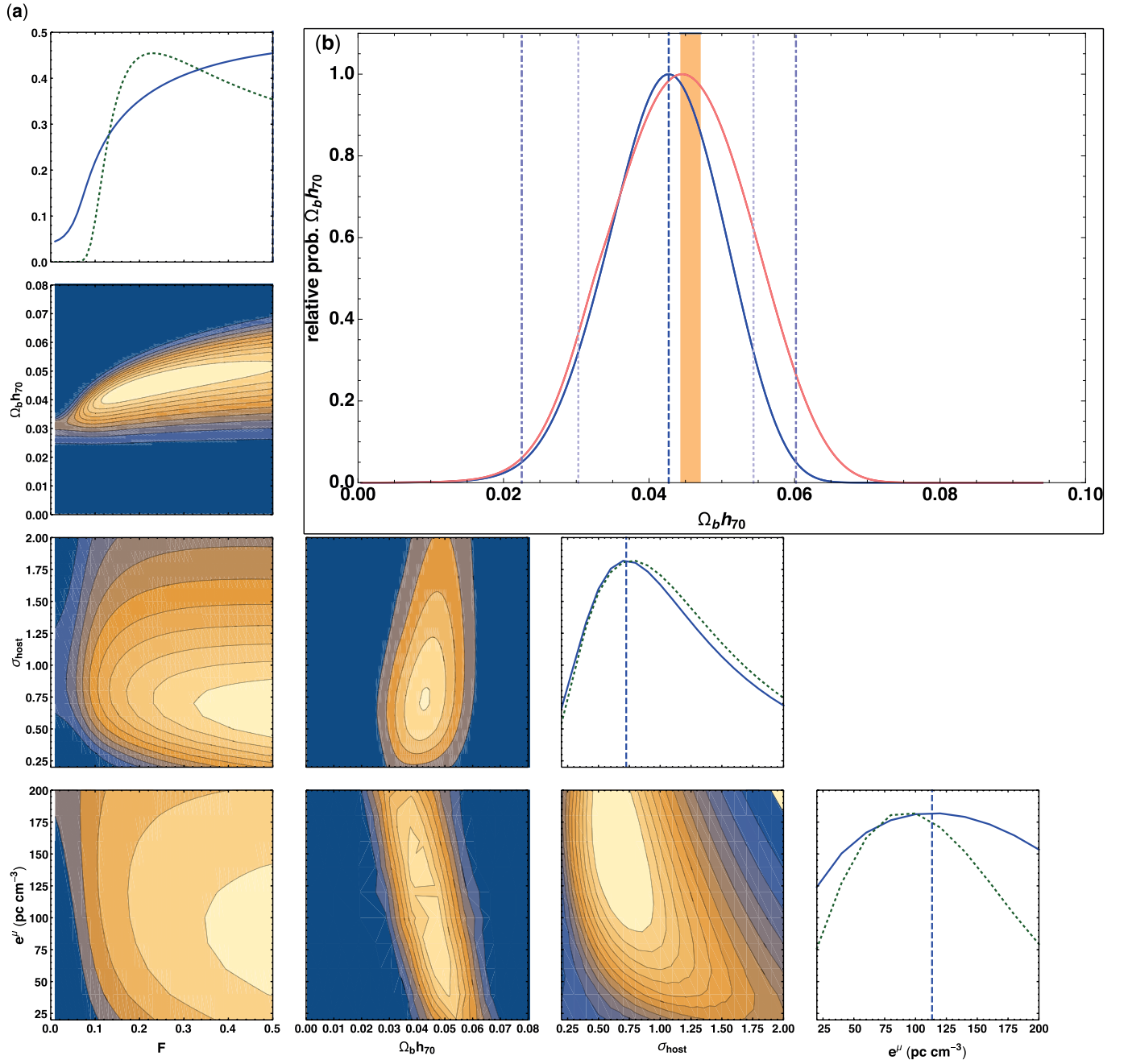
**Extended Data Fig. 2 | The shape of the distribution used to model the host-galaxy dispersion measure,  $\text{DM}_{\text{host}}$ .** The behaviour of the probability distribution  $p_{\text{host}}(\text{DM}_{\text{host}}|\mu, \sigma_{\text{host}})$  is shown for an illustrative set of parameters spanning the range of plausible values for  $\mu$  and  $\sigma_{\text{host}}$ .



**Extended Data Fig. 3 | The expected contribution of the cosmic baryons to the dispersion measure.** The probability distribution of  $\text{DM}_{\text{cosmic}}$  due to the cosmic baryons,  $p(\text{DM})$ , in semi-analytic models and simulations, as encoded in black, blue, red and green in order of increasing redshift ( $z$ ; see key), is compared to the analytic form used in our analysis (DM; equation (4)). The thinner solid curves show semi-analytic models<sup>11</sup> in which the minimum halo

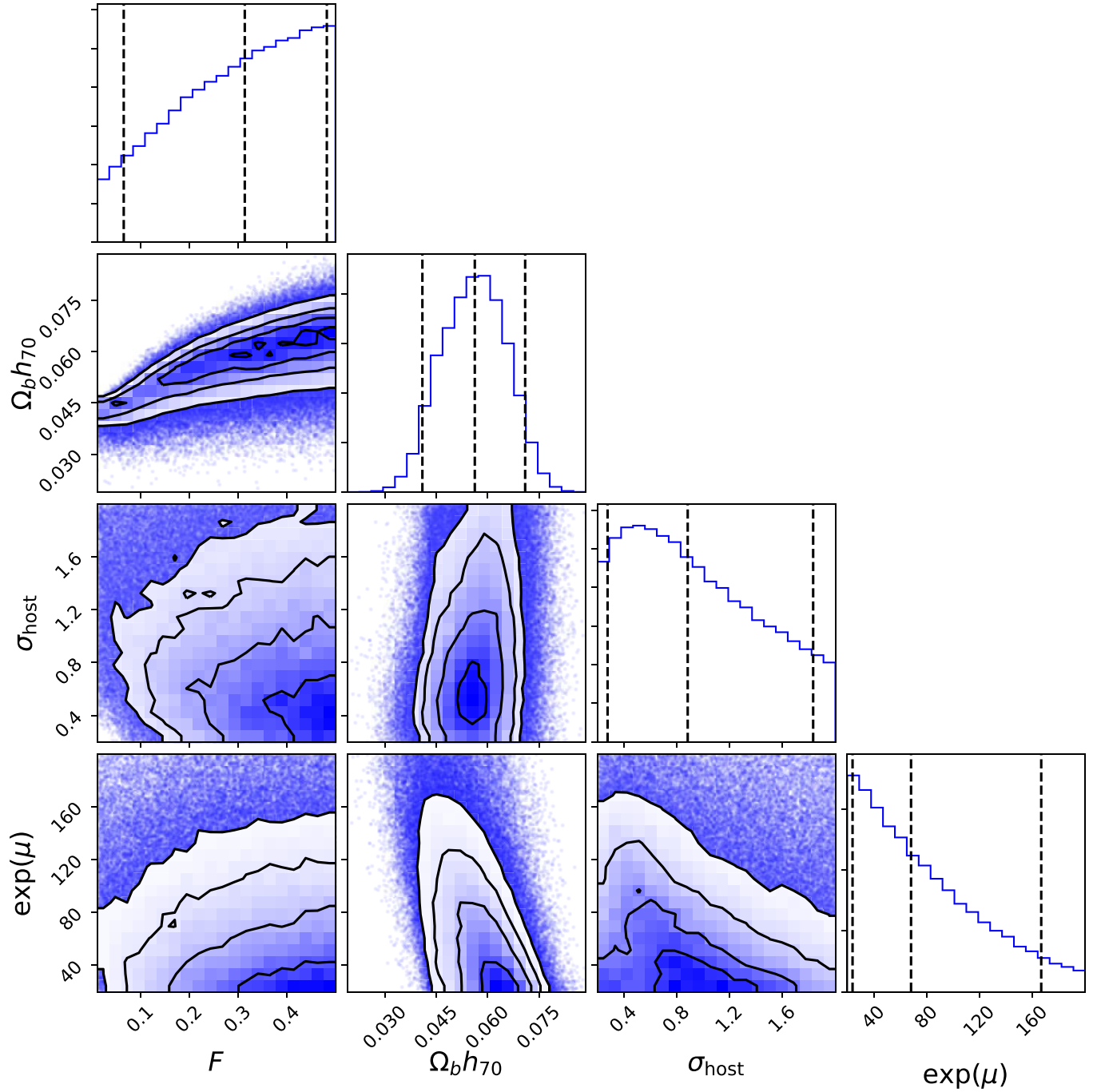
mass that can resist feedback and retain its gas is given by  $M_{\min}$ . The dashed curves are the best-fit analytic function, and the dot-dashed curves assume the  $\sigma_{\text{DM}} = Fz^{-1/2}$  scaling from the  $z = 0.5$  best-fit for which  $F = 0.32, 0.15$  and  $0.09$  for the top, middle and bottom panels, respectively. Because of the success of this Euclidean-space scaling, we adopt it in our analysis. The thicker green solid curve in the bottom panel is calculated from a hydrodynamic simulation<sup>63</sup>.





**Extended Data Fig. 4 | The density of cosmic baryons derived from the extended FRB sample.** The constraints on the IGM parameters  $\Omega_b h_{70}$  and  $F$ , and on the host-galaxy parameters  $\mu$  and  $\sigma_{\text{host}}$  for a log-normal host-galaxy DM

distribution are shown in an identical manner to Fig. 3, but derived using the seven-burst sample (that is, including the five gold-standard bursts as well as FRBs 190523 and 190611).



**Extended Data Fig. 5 | Constraints on the cosmic baryon density and FRB host-galaxy parameters derived using a Bayesian approach.** The results of a Markov Chain Monte Carlo (MCMC) analysis based on our five-FRB gold-standard sample presented in the main text demonstrate broad

agreement with the results of the frequentist analysis presented in Fig. 3. The outermost vertical lines on the histogram denote the confidence region corresponding to that parameter, with the central line indicating the mean value.

Extended Data Table 1 | Detection properties of the ASKAP FRBs

Parameter	Unit	Prior	Median	68%	95%
$F$	None	$\mathcal{U}(0.011,0.5)$	0.31	0.15,0.44	0.04,0.49
$\exp(\mu)$	pc cm <sup>-3</sup>	$\mathcal{U}(20.0,200)$	68.2	33.2,127.8	22.0,181.1
$\sigma_{\text{host}}$	None	$\mathcal{U}(0.2,2)$	0.88	0.43,1.53	0.24,1.91
$\Omega_b h_{70}$	None	$\mathcal{U}(0.015,0.095)$	0.056	0.046,0.066	0.038,0.073

The values of  $\text{DM}_{\text{cutoff}}$  denote the maximum DM at which a burst with those properties listed would have been detectable at an  $S/N$  threshold  $s_i = 9.5$  with the ASKAP telescope back-end at a centre frequency of 1,295 MHz given the burst width  $w$  and search time resolution  $t_{\text{res}}$ , and its 1 MHz spectral resolution.

<sup>a</sup>The detection  $S/N$  value listed is that reported by the incoherent detection pipeline for the telescope beam in which the detection signal was strongest.

<sup>b</sup>The voltage-capture system enables the follow-up of subthreshold events detected in the incoherent pipeline, and subsequent interferometric validation, which would increase the  $S/N$  of a valid event by a factor  $\gtrsim 5$ . The reported  $\text{DM}_{\text{cutoff}}$  is referenced to the threshold  $s_i = 9.0$  relevant to the observing run during which this event was detected.

Extended Data Table 2 | Results of the MCMC analysis

Parameter	Unit	Prior	Median	68%	95%
$F$	None	$\mathcal{U}(0.011,0.5)$	0.31	0.15,0.44	0.04,0.49
$\exp(\mu)$	$\text{pc cm}^{-3}$	$\mathcal{U}(20.0,200)$	68.2	33.2,127.8	22.0,181.1
$\sigma_{\text{host}}$	None	$\mathcal{U}(0.2,2)$	0.88	0.43,1.53	0.24,1.91
$\Omega_b h_{70}$	None	$\mathcal{U}(0.015,0.095)$	0.056	0.046,0.066	0.038,0.073

Estimates of the range of parameters from this analysis at the 68% and 95% confidence levels. These are consistent with the results of the approach described in the main text.

# Author Queries

Journal: **Nature**

Paper: **s41586-020-2300-2**

Title: **A census of baryons in the Universe from localized fast radio bursts**

## AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e-proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query Reference	Reference
Q1	<ul style="list-style-type: none"><li>• Check that there has been no corruption of mathematical symbols.</li><li>• A single sentence summarizing your paper (websum) of no more than 30 words, which will appear online on the table of contents and in e-alerts, has been provided below. Please check this sentence for accuracy and appropriate emphasis. When you receive the eproof link, please check that the display items are as follows (ms no: 2019-09-14302C): Figs 1, 2, 3 (colour); Tables: One; Boxes: None; Extended Data display items: 5 figures, 2 tables; SI: None. The eproof will contain the main-text figures edited by us and (if present) the Extended Data items (unedited except for the legends) and the Supplementary Information (unedited). Please check the edits to all main-text figures (and tables, if any) very carefully, and ensure that any error bars in the figures are defined in the figure legends.</li></ul>
Q2	please check I have correctly made your suggested changes to the penultimate sentence of the first paragraph.
Q3	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.
Q4	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.
Q5	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es).
Q6	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.

# Author Queries

Journal: **Nature**  
Paper: **s41586-020-2300-2**  
Title: **A census of baryons in the Universe from localized fast radio bursts**

**AUTHOR:**

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e-proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query Reference	Reference
Q7	thank you for adding the citations of refs 25 and 26 as footnote d in Table 1. There seemed to be no single place in the table to cite d, as the whole table relates to detected FRBs, so I have put the text under the table in line with house style.
Q8	If ref. 22 (preprint) has now been published in final peer-reviewed form, please update the reference details if appropriate.
Q9	ref. 24 (preprint) has now been published in final peer-reviewed form, please update the reference details if appropriate.