Mon. Not. R. Astron. Soc. **355,** 1339–1347 (2004) doi:10.1111/j.1365-2966.2004.08416.x

Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters

Richard A. Battye,[1] Rod D. Davies[1] and Jochen Weller[2⋆]

[1] *Jodrell Bank Observatory, University of Manchester, Macclesfield, Cheshire SK11 9DL*
[2] *Institute of Astronomy, Madingley Road, Cambridge CB3 OHA*

Accepted 2004 September 16. Received 2004 September 7; in original form 2004 February 19

ABSTRACT

We discuss the possibility of performing blind surveys to detect large-scale features of the Universe using 21-cm emission. Using instruments with ∼5–10 arcmin resolution currently in the planning stage, it should be possible to detect virialized galaxy clusters at intermediate redshifts using the combined emission from their constituent galaxies, as well as less overdense structures, such as protoclusters and the 'cosmic web', at higher redshifts. Using semi-analytic methods, we compute the number of virialized objects and those at turnaround which might be detected by such surveys. We find that there is a surprisingly large number of objects even using small (∼5 per cent) bandwidths and elaborate on some issues pertinent to optimizing the design of the instrument and the survey strategy. The main uncertainty is the fraction of neutral gas relative to the total dark matter within the object. We discuss this issue in the context of the observations which are currently available.

Key words: galaxies: clusters: general – radio lines: general.

1 INTRODUCTION

Emission and absorption of electromagnetic radiation due to transitions between the hyperfine states of neutral hydrogen (H I) (Field 1958) has had a significant impact on our understanding of our Galaxy and our immediate cosmic neighbourhood. At present, however, the highest redshift detection of 21-cm emission is at $z \approx 0.18$ (Zwaan et al. 2001) and only very shallow surveys ($z < 0.04$) of the whole sky have been performed (Zwaan et al. 1997; Kilborn et al. 1999; Ryan-Weber et al. 2002; Lang et al. 2003). In this article we would like to point out that it will soon be possible to perform blind unbiased searches for large objects, both virialized and collapsing, using 21-cm emission as their tracer.

It has long been speculated that it might be possible to detect the onset of the epoch of reionization using redshifted 21-cm emission (Scott & Rees 1990; Madau, Meiksin & Rees 1997) and indeed, spurred on by the claim that the reionization epoch might have been at much higher redshifts ($z \approx 17$) than previously thought (Spergel et al. 2003), there has been much recent work on this subject (Furlanetto, Sokasian & Hernquist 2004; Gnedin & Shaver 2004; Zaldarriga, Furlanetto & Hernquist 2004). Moreover, it has also been pointed out that the evolution of the spin temperature, $T_{\rm S}$, prior to reionization might be such that the large-scale structure can be observed in absorption against the cosmic microwave background (CMB) at very low frequencies (∼30–50 MHz) enabling accurate determination of a variety of important cosmological parameters (Loeb & Zaldarriaga 2004). Our motivation is somewhat different: to map the large-scale structure (LSS) once it has developed sufficiently for the spin temperature to be given by the kinetic temperature of the gas ∼1000 K and $T_{\rm S} \gg T_{\rm CMB}$, that is, using 21-cm emission. This issue was considered by Subramanian & Padmanabhan (1993), Kumar, Padmanabhan & Subramanian (1995) and Bagla, Nath & Padmanabhan (1997).

For various proposed instruments, our aim is to make conservative estimates of the number of virialized clusters that might be found in the range $z < 1$, and also of collapsing objects with lower overdensities at $z > 1$. These higher redshift objects are the protoclusters which would have formed virialized clusters by the present day and in some sense characterize the so-called 'cosmic web' of LSS observed in cosmological *N*-body simulations. They are likely to be H I rich. Naively, virialized objects would not necessarily be the most obvious places to look for neutral gas since the process of virialization and the creation of the intracluster medium involves violent gravitational processes which are likely to strip neutral gas from the constituent galaxies by tidal interactions and ram pressure (Gunn & Gott 1972). However, by virtue of their large overdensity such objects should still contain a substantial neutral component, probably larger at high redshifts than locally, albeit locked up in galaxies. The fiducial detection threshold that we will discuss in this paper will be $\approx 10^{11}\,{\rm M}_\odot$ of H I, which we will show might be possible on large patches of the sky; it is worth pointing out that this would only require a cluster to contain 20 gas-rich spirals with H I masses $\sim 5 \times 10^9\,{\rm M}_\odot$. This is likely to be the case for many clusters.

Most importantly, in the context of this discussion, all the objects we will consider, both virialized and collapsing, will have angular

⋆E-mail: jweller@fnal.gov

diameters ∼5–10 arcmin which will match the resolution of the telescopes and arrays with ∼100-m baselines currently being planned to detect redshifted 21-cm emission. This will make them ideal survey targets since they would fill the beam, in contrast to ordinary galaxies at the same redshift whose signal will be substantially diluted and will most likely be below the confusion limit.

Blind surveys for galaxies and galaxy clusters have been performed in the optical waveband for many years. The most recent galaxy redshift surveys (2dFGRS and SDSS) have yielded significant constraints on cosmological parameters via measurements of the power spectrum and redshift-space distortions, as well as a wealth of understanding of the properties of the individual galaxies (Peacock 2003; Tegmark et al. 2004). If deep redshift surveys using H I could be performed they would open up a new window on the Universe which could have some technical advantages (and inevitably some disadvantages) over optical approaches. In particular, using the flexible spectral resolution of radio receivers, objects can be selected using spectroscopic methods rather than, for example, photometrically from optical plates, making them flux-limited over the specific redshift range probed. The structure of the selection function should, therefore, be simple to define. It should also be possible to deduce the redshift and accurate linewidths for individual objects allowing one to probe the underlying mass distribution directly and establish the precise selection criteria a posteriori. This could overcome some of the issues of bias inherent in optical surveys. Moreover, potential confusion with other spectral lines is unlikely in the radio waveband.

A standard line of argument often put forward is that the optical surveys are sensitive to the sum of all the starlight from the galaxies and could be significantly biased by the non-linearities involved in the star formation process, while the neutral component is of primordial origin and hence is unbiased. This is unlikely to be completely true since the neutral component is also involved in these processes. However, observing the H I component is more than just another view of the same object. The H I haloes of nearby galaxies extend out very much further than their optical counterparts, suggesting that there is extra information in the dynamics of the H I.

As with galaxies, the study of clusters has focused on the central region where diffuse hot gas is situated. This emits very strongly in the X-ray waveband and can also lead to fluctuations in the CMB via the Sunyaev–Zeldovich (SZ) effect. However, typically the core radius of the cluster is only about 1/10th of the linear size of the object and 1/1000th of its volume. It is likely that the study of these objects using H I as the tracer will lead to new insights in an analogous way to galaxies. In particular many of the H I rich spiral galaxies are likely to be situated away from the core and there is also a possibility of a component of diffuse H I similar to that seen around individual galaxies.

At present radio telescopes which are sufficiently sensitive to map the sky quickly enough to be competitive with optical surveys do not exist. However, the next few years could see the start of development of instruments which are many times more powerful in terms of survey speed than those presently available. The Square Kilometre Array (SKA) (see http://www.skatelescope.org for details) is the ultimate goal of much technology development currently underway. Although its precise design is still evolving, its projected specification involves a collecting area of $\sim 10^6$ m^2, subarcsecond resolution, a large instantaneous field of view and significant bandwidth in the region where redshifted H I will be detected. Such an instrument will have the ability to perform significant galaxy redshift surveys tracing H I over a wide range of redshifts.

Prototypes are being planned which might be a few per cent of the full SKA in area. Searches for galaxies at low redshift, extending those currently being performed, would be an obvious possibility for such instruments. Our aim, here, is to discuss issues related to the design of such an instrument in the context of a blind survey for large objects at intermediate and high redshifts such as virialized galaxy clusters and collapsing protoclusters. As we have already discussed, they would be ideal targets for these surveys in terms of their size, our conclusion being that the planned instruments will be sufficiently sensitive to perform significant surveys.

First, we will discuss different proposed concepts for such instruments and perform a simple sensitivity calculation for the limiting H I mass ($M_{\rm HI}^{\rm lim}$) of a survey assuming a fixed integration time.[1] We will then discuss the possibility of detecting virialized clusters. In order to make the link with theoretical predictions of the number of dark matter haloes, we need to make some assumption as to the H I content of individual dark matter haloes. This is a very complicated issue and since we are, at this stage, only trying to make a rough estimate of the number of clusters one might find, we will assume that there exists a universal ratio $f_{\rm HI}(M,z) = M_{\rm HI}/M$ which relates the dark matter mass of a halo, M, to its neutral gas component at a given redshift. This should be a good approximation in the mass range relevant to clusters, but is unlikely to be so of smaller masses. We will make various arguments in order to estimate this fraction. At each stage we will attempt to make conservative assumptions and therefore our estimate of the number of objects found in a given survey should be a lower bound. Under these assumptions we can then deduce the number of objects that a survey might find using the results of N-body simulations. We then discuss some issues related to the design of the instrument, the survey strategy and confusion due to H I in the field. Finally, we will adapt our calculations to compute the number of objects at turnaround and discuss issues relating to their detection

Throughout this paper, we will assume a cosmological model which is flat, with the densities of matter and the cosmological constant relative to critical given by $\Omega_{\rm m} = 0.3$ and $\Omega_\Lambda = 0.7$. The Hubble constant used is 72 km s^{-1} Mpc^{-1} and the power spectrum normalization is assumed to be $\sigma_8 = 0.9$. These values are compatible with a range of observations (Contaldi, Hoekstra & Lewis 2003; Spergel et al. 2003).

2 PROPOSED INSTRUMENTS

A number of different concepts are under discussion which could make observations in the frequency range $400\,{\rm MHz} < f_{\rm obs} < 1200$ MHz. The proposed instruments will, in fact, be sensitive to a much wider range of frequencies (up to 1700 MHz), but this is the most sensible range to discuss the detection of large objects since they will be close to, or completely, unresolved. The design features which are of interest to us here are the overall collecting area, A, the aperture efficiency, η, the instantaneous bandwidth, $\Delta f_{\rm inst}$, the instantaneous field of view (FOV), $\Omega_{\rm inst}$, the longest baseline, d, and the system noise temperature of the receivers, $T_{\rm sys}^{\rm rec}$, which also includes the contribution from the atmosphere. The overall system temperature, $T_{\rm sys}$, at frequency f can be conservatively estimated by

[1] We note that all integration times discussed in this paper are 'on-source' integration times and do not take into account any practical difficulties required to make the observations. These could vary significantly for different experimental configurations.

taking into account the CMB and galactic background

$$T_{\rm sys} \approx T_{\rm sys}^{\rm rec} + 2.7\ {\rm K} + 10\ {\rm K}\left(\frac{f}{408\ {\rm MHz}}\right)^{-3}, \quad (1)$$

in a cold part of the sky.

We will focus on two different concepts which differ in the way they form the instantaneous field of view. First, we will consider a conventional interferometer similar to the Allen Telescope Array (ATA), which comprises $n_{\rm d}$ dishes of diameter D. In this case $\Omega_{\rm inst} \propto (\lambda/D)^2$ and $A = \pi\, n_{\rm d} D^2/4$ where $\lambda \propto 1/f_{\rm obs}$. The current design of the ATA has $n_{\rm d} \approx 350$, $D = 6.1$ m and it is possible that $\eta \approx 0.6$, $d \approx 400$ m and $\Delta f_{\rm inst}/f_{\rm obs} \approx 0.05$. The FOV is set by the size of the individual telescopes and the synthesized beamsize is set by the longest baseline. Such a configuration would have a synthesized beam with $\theta_{\rm FWHM} = 3.2$ arcmin at $f_{\rm obs} = 1000$ MHz and a corresponding instantaneous FOV of 13.4 deg^2. Note that throughout we assume a tapered aperture, that is, $\theta_{\rm FWHM} = 1.2\lambda/d$, which should be a reasonable approximation.

The other concept which we will consider is that of a phased array which could potentially have a much larger FOV, the idea being to use very low-cost antennae, arranged in $n_{\rm p}$ patches of size P, and form $n_{\rm b}$ individual beams, perhaps using off-the-shelf computer hardware. Hence, $\Omega_{\rm inst} \propto n_{\rm b}(\lambda/d)^2 \leqslant (\lambda/P)^2$ and $A = \pi\, n_{\rm p} P^2/4$ assuming that the patches are circular, from which we can deduce that $d \geqslant \sqrt{n_{\rm b}}P$. Such a concept could have more than one instantaneous FOV, possibly allowing for more continuous usage for H I studies. Three possible scenarios for such an instrument are presented in Table 1. These range from conservative, set-up I, to optimistic, set-up III. We should note that for $P \approx 4$ m, none of the above configurations come close to saturating the inequality on P. The value of $n_{\rm b}$ should be taken as being a guide figure since extra computing power could increase this close to the maximum FOV set by P.

The standard formula (Roberts 1975) relating the H I mass, $M_{\rm H\,I}$, and the observed flux density, S, integrated over a velocity width, Δv, in an Friedmann–Robertson–Walker (FRW) universe is given by

$$\frac{M_{\rm H\,I}}{{\rm M}_\odot} = \frac{2.35\times10^5}{1+z}\left(\frac{d_{\rm L}}{\rm Mpc}\right)^2\left(\frac{S\Delta v}{\rm Jy\ km\ s^{-1}}\right), \quad (2)$$

where $d_{\rm L}(z)$ is the luminosity distance to redshift z. For a fiducial velocity width of $\Delta v \approx 800\,{\rm km\,s^{-1}}$, assumed at this stage to be independent of the mass, an object containing $10^{11}\,{\rm M}_\odot$ of H I will have $S_{800} \approx 4, 14, 42, 150, 840\ \mu$Jy at $f_{\rm obs} = 400$, 600, 800, 1000, 1200 MHz corresponding to $z = 2.55$, 1.37, 0.78, 0.42, 0.18, respectively.

Table 1. Three possible sets of design parameters for a phased array. Included also are $\theta_{\rm FWHM}$ and $\Omega_{\rm inst}$ at $f_{\rm obs} = 1000$ MHz, and the figure of merit M relative to the configuration described in the text. In each case we assume that $\eta \approx 0.8$. This value, which is slightly larger than that conventionally assumed for parabolic dishes.

	I	II	III
$A/{\rm m}^2$	5000	10 000	15 000
$\Delta f_{\rm inst}/f_{\rm obs}$	0.02	0.05	0.10
$d/{\rm m}$	150	250	350
$n_{\rm b}$	1000	2000	4000
$T_{\rm sys}^{\rm rec}/{\rm K}$	40	30	20
$\theta_{\rm FWHM}$	8.4′	5.0′	3.6′
$\Omega_{\rm inst}/{\rm deg}^2$	22.2	15.9	16.3
M	≈5	≈60	≈550

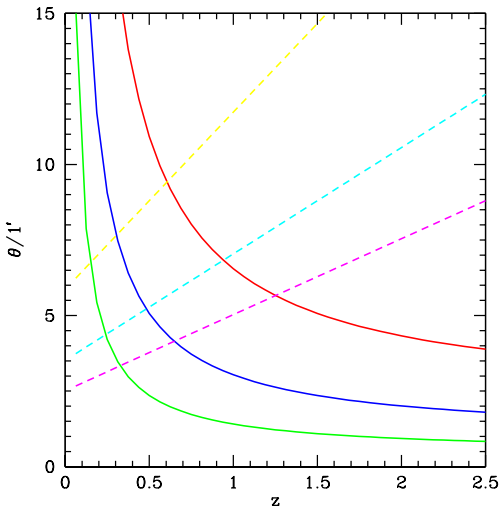

Figure 1. The angular diameter size (twice the virial radius) of virialized objects with $M_{\rm vir} = 10^{15}$, 10^{14} and $10^{13}\,{\rm M}_\odot$ (solid lines, top to bottom, respectively). Included also is the estimated beamsize for the three different phased array configurations (dashed lines, I top, II middle, III bottom). Only the closest and largest objects will be resolved by the instruments discussed in this paper.

Assuming that the clusters are unresolved, which will be a good approximation for the configurations under consideration here at intermediate and high redshift (see Fig. 1 for virialized objects), then the rate at which one can survey the sky over the instantaneous bandwidth is quantified by the instantaneous survey sensitivity

$$\mathcal{R}_{800} = \frac{2T_{\rm sys}}{\eta A\sqrt{\Delta f_{\rm obj}\Omega_{\rm inst}}}, \quad (3)$$

where $\Delta f_{\rm obj}$ is the frequency interval corresponding to Δv. The noise, S_σ, attainable for a survey with angular coverage, $\Delta\Omega$, in integration time, t, is given by $S_\sigma = \mathcal{R}_{800}\sqrt{\Delta\Omega/t}$. This line of argument should also yield important information even when the object is resolved. Projected values for $\mathcal{R}_{800}$ are presented in Table 2 for the ATA and the different phased array set-ups using $f_{\rm obs} = 400$–1200 MHz along with the integration time required to achieve a 5σ detection ($S_{800} > 5S_\sigma$) of $M_{\rm H\,I}^{\rm lim} \approx 10^{11}\,{\rm M}_\odot$ of H I in $\Delta v \approx 800\ {\rm km\,s^{-1}}$ over an area of 100 deg^2. Note that in reality the angular coverage of a given survey is likely to be an integer multiple of the instantaneous FOV. However, this approach to presenting the sensitivity makes it possible to compare experiments with different FOVs in a consistent way.

3 DETECTING GALAXY CLUSTERS

3.1 H I content of galaxy clusters

It is possible to make accurate predictions as to the clustering of dark matter, but the evolution of the H I is much more difficult to predict. Therefore, we will introduce the empirical quantity $f_{\rm H\,I}(M, z)$ which we will define to be the fraction of the H I in a dark matter halo of mass M at redshift z. We anticipate that this quantity is likely to have a substantive scatter $\sigma_{\rm H\,I}^2 = \langle f_{\rm H\,I}^2\rangle - f_{\rm H\,I}^2$, but we shall assume this is zero since estimating it would be futile due to the small number of observations available at present. Moreover, any scatter is likely to increase the number of objects that are found in a given survey, assuming that the mean value is correct. This is because there are typically many more objects just below the mass limit than above it.

Table 2. Approximate instantaneous survey sensitivity R_{800} for $f_{\rm obs} = 400$–1200 MHz in units of mJy sec$^{1/2}$ deg^{-1} for the ATA and the phased array set-ups I, II and III. Also presented is the integration time t_{800} in days required to achieve a 5σ detection threshold of $M_{\rm HI}^{\rm lim} = 10^{11}\,{\rm M}_\odot$ over an area of 100 deg^2 and $\Delta v = 800$ km s^{-1}.

	$f_{\rm obs}$/MHz	ATA	I	II	III
	400	1.80	2.98	1.43	0.73
	500	1.74	3.01	1.41	0.69
	600	1.78	3.14	1.45	0.69
	700	1.85	3.31	1.51	0.71
R_{800}	800	1.93	3.48	1.59	0.74
	900	2.02	3.66	1.67	0.78
	1000	2.11	3.84	1.74	0.81
	1100	2.20	4.01	1.82	0.84
	1200	2.29	4.18	1.89	0.88
	400	4480	12 000	2830	730
	500	1335	4000	875	200
	600	460	1445	300	70
	700	165	535	110	25
t_{800}	800	60	190	40	8.8
	900	20	64	13	2.8
	1000	5.7	19	3.9	0.9
	1100	1.4	4.5	0.9	0.2
	1200	0.2	0.7	0.15	0.03

We will first present a simple order-of-magnitude estimate for $f_{\rm HI}$ and then go on to discuss the current status of observations.

One can estimate the fraction of H I in clusters, $f_{\rm HI}^{\rm clust}$, from that in a typical galaxy, by taking into account different mass-to-luminosity ratios, M/L, of typical clusters and galaxies, and the fact that only spiral galaxies are H I rich. If $f_{\rm spiral}$ is the fraction of spiral galaxies in a typical cluster then

$$f_{\rm HI}^{\rm clust} \sim f_{\rm spiral}\frac{(M/L)^{\rm gal}}{(M/L)^{\rm clust}} f_{\rm HI}^{\rm gal}. \tag{4}$$

By considering the average properties of galaxies (Roberts & Haynes 1994), the value of $f_{\rm HI}^{\rm gal}$ is observed to be $\sim$0.06 for a typical spiral galaxy at $z \approx 0$. The value M/L of such an object is likely to be $\sim$8, whereas that for a typical cluster is $\sim$240. It is believed that $f_{\rm spiral} \sim 0.5$, making $f_{\rm HI}^{\rm clust} \sim 10^{-3}$.

Another approach is to compute the value of $f_{\rm HI}$ for a cluster by summing up the measured values of $M_{\rm HI}$ for the constituent galaxies and computing their line-of-sight velocity dispersion, σ, and hence their virial mass $M_{\rm vir} = 3\pi\sigma^2 R_{\rm H}/(2G)$, where $R_{\rm H}$ is the mean harmonic radius and G is the gravitational constant. We should note that this is likely to underestimate $f_{\rm HI}$ since it ignores the many gas-rich dwarf galaxies which are below the detection threshold of any survey. Moreover, it also ignores, as does the order of magnitude estimate above, the possibility that there exists a diffuse component of H I such as the tidal plumes of H I observed in galaxy groups (Appleton, Davies & Stephenson 1981).

The H I content of the Virgo galaxy cluster is the best studied (Hutchmeier & Richter 1986), but it is well known that this object is not completely relaxed, it being comprised of three interacting subclusters. Hence, the computed value of $f_{\rm HI}$ is likely to be an unreliable statistical indicator of the global value. Studies of the H I content of other clusters are much less detailed. The Abell cluster A3128 has been surveyed for H I (Chengalur et al. 2001). By co-adding the H I at the position of galaxies brighter than an r-magnitude of 16.2, a total H I mass of $1.3 \times 10^{11}\,{\rm M}_\odot$ was detected. $M_{\rm vir}$ for this cluster is $\approx 1.5 \times 10^{14}\,{\rm M}_\odot$ estimated from the velocity dispersion and the angular distribution of the galaxies at an assumed distance of 240 Mpc. For the detected galaxies $M_{\rm HI}/M_{\rm vir} \approx 0.9 \times 10^{-3}$, compatible with our earlier order-of-magnitude estimate. A recent sensitive multibeam study of the Centaurus A group has been made (Banks et al. 1999) and this survey was able to detect H I in galaxies over a wide range of luminosities and, therefore, includes at least part of the dwarf galaxy contribution. For this group $M_{\rm HI}/M_{\rm vir} \approx 1.1 \times 10^{-3}$.

Many different arguments suggest that $f_{\rm HI}$ should increase with redshift. H I is the basic fuel for star formation and it is believed that for an individual object $\dot{M}_\star \propto M_{\rm HI}$. Since the global star formation rate is known to grow approximately linearly with redshift to a maximum about five to ten times the local value at $z \approx$ 1–2, it seems likely that $f_{\rm HI}^{\rm gal}(M, z\sim 1) \approx (5\text{–}10) \times f_{\rm HI}^{\rm gal}(M, z\sim 0)$, an effect which may be even stronger within clusters. A related, but logically distinct fact is the observation that high-redshift clusters contain a larger fraction of blue galaxies when compared to local ones (Butcher & Oemler 1984) and also that they contain a larger fraction of spirals and a reduced number of S0s (Dressler et al. 1997). Both these observed facts suggest that $f_{\rm spiral}$ also increases with z. Hence, using the estimate (4) it seems sensible to deduce that using the locally observed $f_{\rm HI}^{\rm clust}$ is likely to be an underestimate by a factor of a few.

Various estimates of $\Omega_{\rm HI}$ appear in the literature and these can be used to gain some further insight into the issues under consideration here. By making observations of damped Lyman-α systems in an average redshift range of $\langle z\rangle \approx 2.4$ the CORALS project (Ellison et al. 2001) estimate that $\Omega_{\rm HI} \approx 2.6 \times 10^{-3}$. Using the fiducial value of $\Omega_{\rm m}$, one can deduce a value of $f_{\rm HI} \approx 8 \times 10^{-3}$ for the field at this epoch corresponding to a density of $\rho_{\rm HI} \approx 3.6 \times 10^8\,{\rm M}_\odot\,{\rm Mpc}^{-3}$. Using the recently published data from the HIPASS survey it has been deduced (Zwaan et al. 2003) that for $z < 0.04$ the value of $\Omega_{\rm HI}$ is much lower, $\Omega_{\rm HI} \approx 3.8 \times 10^{-4}$, from which one can deduce that $f_{\rm HI} \approx 1.2 \times 10^{-3}$ and $\rho_{\rm HI} \approx 5.3 \times 10^7\,{\rm M}_\odot\,{\rm Mpc}^{-3}$. Remarkably this is approximately a factor 6–7 smaller than the estimated value at $z = 2.4$ and is clearly compatible with our earlier remarks concerning the global star formation rate. Moreover, the estimated value at $z \approx 0$ is very close to the value we have discussed in the context of clusters. This would tend to suggest that the value of $f_{\rm HI}$ is only marginally lower in clusters (say 20–30 per cent) than in the field.

3.2 Number of virialized dark matter haloes

Since we are only trying to compute a lower bound on the number of objects found in a given survey, it is reasonable to assume a value of $f_{\rm HI} \approx 10^{-3}$ constant over the range of masses we are considering and independent of redshift. If $M_{\rm HI}^{\rm lim} \approx 10^{11}\,{\rm M}_\odot$ as discussed earlier for the various planned surveys, then $M_{\rm lim} \approx 10^{14}\,{\rm M}_\odot$ for the dark matter and hence computing the number of objects greater than this limit will give an estimate of the number of objects likely to be found in a survey.

Now, let us compute the number of objects with mass greater than some limiting mass $M_{\rm lim}$ between $z - \Delta z/2$ and $z + \Delta z/2$ in a solid angle $\Delta\Omega$:

$$N(M > M_{\rm lim}) = \Delta z \Delta\Omega \frac{{\rm d}V}{{\rm d}z\,{\rm d}\Omega}\int_{M_{\rm lim}}^{\infty}\frac{{\rm d}n}{{\rm d}M}{\rm d}M, \tag{5}$$

where

$$\Delta z = (1+z)\Delta v = (1+z)\Delta f_{\rm inst}/f_{\rm obs}, \tag{6}$$

is assumed to be small, ${\rm d}V/({\rm d}z\,{\rm d}\Omega)$ is the comoving volume element and ${\rm d}n/{\rm d}M$ is the comoving number density of objects.

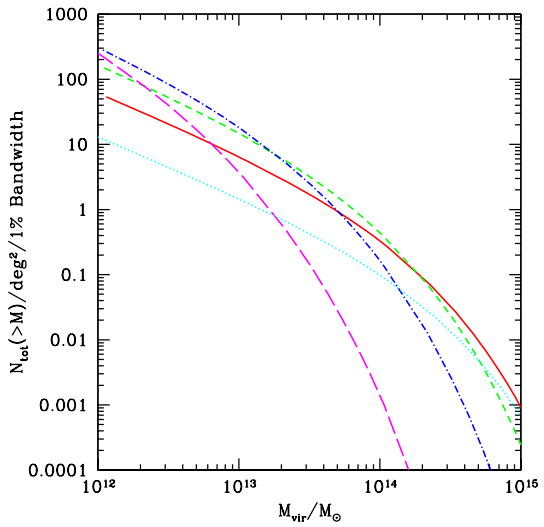

Figure 2. The number of dark matter haloes with mass greater than $M_{\rm vir}$ per deg^2 per 1 per cent bandwidth. The dotted line is for $f_{\rm obs} = 1200$ MHz, the solid line is for 1000 MHz, the short dashed line for 800 MHz, the dot-dashed line for 600 MHz and the long dashed line for 400 MHz. Note that there are between 1 and 0.1 haloes with $M_{\rm vir} \approx 10^{14}\,{\rm M}_\odot$ for $f_{\rm obs} \geqslant 600$ MHz.

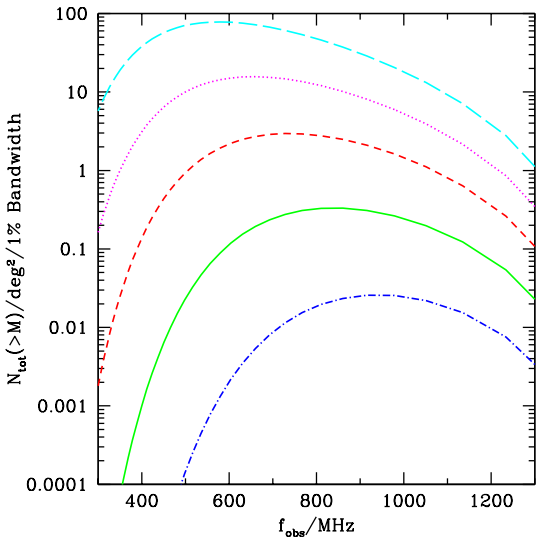

Figure 3. The same quantity as plotted in Fig. 2 but this time against $f_{\rm obs}$ for $M_{\rm vir} = 3 \times 10^{12}\,{\rm M}_\odot$ (long-dashed line), $10^{13}\,{\rm M}_\odot$ (dotted line), $3 \times 10^{13}\,{\rm M}_\odot$ (short-dashed line), $10^{14}\,{\rm M}_\odot$ (solid line) and $3 \times 10^{14}\,{\rm M}_\odot$ (dot-dashed line).

We will use an expression for the comoving number density which is derived from numerical simulations carried out by the VIRGO consortium (Evrard et al. 2002),

$$\frac{{\rm d}n}{{\rm d}M}(z, M) = -0.22\frac{\rho_{\rm m}(t_0)}{M\sigma_M}\frac{{\rm d}\sigma_{\rm M}}{{\rm d}M}\exp[-|A(z, m)|^{3.86}], \quad (7)$$

where $A(z, M) = 0.73 - \log[D(z)\sigma_M]$, $\rho_{\rm m}(t_0)$ is the matter density at the present day, σ_M is the overdensity in a virialized region of mass M and $D(z)$ is the normalized growth factor. Here, the mass is defined to be M_{200}, that inside a region with an approximately spherical overdensity $\Delta_{\rm c} = 200$. This can be related to the virial mass, $M_{\rm vir}$, if we assume an NFW function for the cluster profile (Navarro, Frenk & White 1997). In all the subsequent discussion we have made this correction using a concentration parameter of $c = 5$, in which case that from M_{200} to $M_{\rm vir}$ is typically very small.

In Fig. 2, we plot $N(M > M_{\rm lim})$ per deg^2 per 1 per cent bandwidth against $M_{\rm vir}$ for five different redshifts ($z = 2.55$, 1.37, 0.78, 0.42, 0.18) corresponding to frequencies $f_{\rm obs} = 400$, 600, 800, 1000 and 1200 MHz, respectively. We see that there are ∼0.5 dark matter haloes with $M_{\rm vir} \approx 10^{14}\,{\rm M}_\odot$ per deg^2 per 1 per cent bandwidth for 600 MHz $< f_{\rm obs} <$ 1200 MHz. There are very few objects with $M_{\rm vir} > 2 \times 10^{13}\,{\rm M}_\odot$ for $f_{\rm obs} = 400$ MHz. Larger objects with $M_{\rm vir} \approx 3 \times 10^{14}\,{\rm M}_\odot$ are much rarer for $f_{\rm obs} = 600$ MHz, but for higher frequencies there would still be many objects above this limiting mass. Using the fiducial value of $M_{\rm lim} = 10^{14}\,{\rm M}_\odot$ at $f_{\rm obs} = 1000$ MHz, by referring back to Table 2, we see that it would be possible to cover 10 000 deg^2 in ≈400 d using set-up II. Such a survey would find ≈15 000 objects above this mass limit since there are ≈0.3 objects per deg^2 per 1 per cent bandwidth. We note that a similar survey would take only 90 days with set-up III and there would be twice as many objects found since it has a larger bandwidth. For the same integration time one could cover only about 1000 deg^2 at $f_{\rm obs} = 800$ MHz to the same mass limit and would find ≈1000 objects with set-up II. At $f_{\rm obs} = 1200$ MHz one would be able to cover more area, but there are fewer objects since the comoving volume is less at low redshift, and things are likely to be somewhat more complicated since $10^{14}\,{\rm M}_\odot$ objects would be resolved at this frequency. Much less area could be covered to the same depth at $f_{\rm obs} = 600$ MHz and as we will discuss it might be that such objects are difficult to detect against the background since the beamsize increases with decreasing frequency. Clearly, there is no point in searching for virialized objects at $f_{\rm obs} = 400$ MHz.

We have also computed the same quantity as a function of $f_{\rm obs}$ in Fig. 3 for different values of $M_{\rm lim}$. We see that there are very few virialized objects accessible to observations with $f_{\rm obs} < 600$ MHz, again suggesting that surveys which are intending to search for such objects are unlikely to find anything significant for $z > 1.4$. However, we also see that for the fiducial value of $M_{\rm lim} \approx 10^{14}\,{\rm M}_\odot$ which we have been using, there is a wide range of frequencies for which one will find more than one object per 10 deg^2 in a 1 per cent bandwidth. Many smaller objects [$M_{\rm vir} \approx (3 \times 10^{12})$–$10^{13}\,{\rm M}_\odot$] are accessible at lower frequencies due to the evolution of structure.

3.3 Optimal design and survey strategy

One can attempt to gain some understanding of the optimal design of an instrument and survey strategy by substituting into (5) for $\Delta\Omega$ from (3), for Δz from (6), and using $S_\sigma \Delta v \propto M_{\rm H\,I}/d_{\rm L}^2$. One finds that $N \propto t\mathcal{M}(M_{\rm H\,I}^2/\Delta v)$, where

$$\mathcal{M} \propto \frac{\eta^2 A^2 \Omega_{\rm inst}\Delta f_{\rm inst}}{T_{\rm sys}^2} \quad (8)$$

can be thought of as the instrumental figure of merit which clearly has sensible dependencies on the relevant parameters. This effectively quantifies how many objects one would find in a survey ignoring any potential systematics. Alternatively, the integration time required to find a fixed number of objects is proportional to $\mathcal{M}^{-1}$. It is clear that similar arguments can be made for a galaxy redshift survey and this would, therefore, also apply to the SKA and other similar instruments (although the area, A, would be that corresponding to baselines shorter than the characteristic size of the objects in question, for example, galaxies). This formula should also yield important information when applied to surveys for other kinds of objects.

The value of $\mathcal{M} \approx 27$ for the ATA configuration discussed earlier and its value is presented in Table 1 for the phased arrays. This figure of merit has been normalized to the theoretical sensitivity of the Parkes Telscope Multibeam (PM) receiver with 13 beams (Barnes et al. 2001), although we note that this instrument can only observe in a narrow frequency range around 21 cm. We have also computed this quantity for the Lovell Telescope multibeam receiver with four beams and the Green Bank Telescope (single beam); they are $\mathcal{M} \approx 0.34$ and 0.21, respectively. Note that we have used the frequency-independent $T_{\rm sys}^{\rm rec}$ to compute this quantity.

We see that set-up I is about a factor of 5 more powerful than the PM, while both the presented ATA configuration and set-up II improve on this substantially. Set-up III is more than 500 times more powerful than the PM. For a phased array configuration $\Omega_{\rm inst} \propto n_{\rm b} d^{-2}$ and hence $\mathcal{M} \propto \eta_{\rm F} = A/d^2$ which represents the filling factor of the array. It is clear that, at least for unresolved detections and ignoring the issue of confusion, which we will discuss below, a totally filled array ($\eta_{\rm F} = 1$) will perform the best in terms of having the largest number of objects above the detection threshold. This is because an array with $\eta_{\rm F} \ll 1$ has a much lower sensitivity to surface brightness fluctuations than one with $\eta_{\rm F} \sim 1$. Such an array would be likely to have poor resolution particularly at low frequencies and would, therefore, make it difficult to make any subselection within the sample to cut down on systematics. It is possible that low-resolution arrays could also suffer from confusion-related issues (see below).

Assuming a virialized halo $\Delta v \propto \sigma \propto M_{\rm vir}^{1/3}$ (σ is the velocity dispersion of the object) and if $f_{\rm H\,I}$ is independent of $M_{\rm vir}$, then the multiplicative factor is $M_{\rm H\,I}^2/\Delta v \propto M_{\rm vir}^{5/3}$. Since the integral in (5) is a decaying function of $M_{\rm vir}$ with negative power law $n \equiv n(M_{\rm vir})$, the optimal observing strategy would be to set the noise so that $M_{\rm lim}$ is that for which $n = 5/3$, that is $M_{\rm lim} \approx 2 \times 10^{14}$, 10^{14}, 4×10^{13}, 1.5×10^{13} and $10^{12}\,{\rm M}_\odot$ for $f_{\rm obs} = 1200$, 1000, 800, 600 and 400 MHz, respectively, this being a simple consequence of the evolution of structure. We have presented this power law $n = -{\rm d}(\log\ N)/{\rm d}(\log\ M)$ as a function of $M_{\rm vir}$ in Fig. 4. We see that under these assumptions $M_{\rm lim} \approx 10^{14}\,{\rm M}_\odot$ should be close to the optimal mass limit for $f_{\rm obs} = 1000$ MHz and that lower limits, requiring deeper surveys and hence less angular coverage for a fixed integration time, are needed to be optimal at lower frequencies. We should caution that this is heavily predicated on the assumption that $f_{\rm H\,I}$ is independent of M which we have already argued is unlikely to be the case and this would have to be taken into account before relying on such a calculation to set the depth of an actual observational strategy. If $f_{\rm H\,I}$ decreases with M then this would mean one should perform a deeper survey than one would if it were independent of M. Suffice to say, if one has some idea as to the dependence of $f_{\rm H\,I}$ on M, the method would remain the same, but one would use a different value of n to select the optimum.

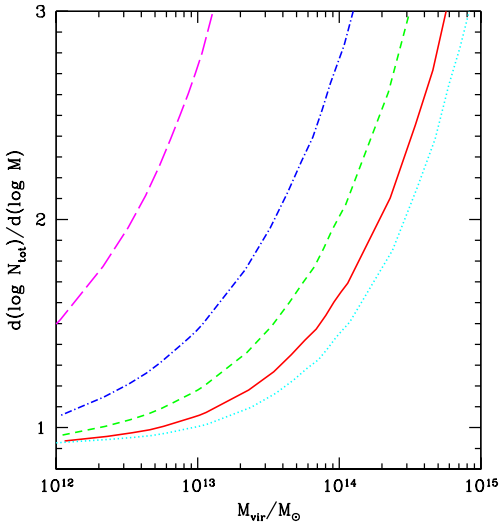

Figure 4. The power law n such that $(N/1\ {\rm deg}^2) \propto M_{\rm vir}^{-n}$. Under the assumption that $f_{\rm H\,I}$ is independent of $M_{\rm vir}$ the optimal strategy is when $n = 5/3$. The lines are labelled as in Fig. 2.

3.4 Confusion noise

Observations using spectral lines are not typically affected by issues of confusion as can often be the case for continuum sources. However, in the current situation we are dealing with very large amounts of H I and very large beams. One has, therefore, to be careful to avoid making the beamsize so large that a typical volume enclosed by the beam and a fiducial velocity width contains an amount of H I comparable to the detection threshold. This becomes more and more important as one makes deeper and deeper surveys, particularly at high redshift. As we shall see this is very sensitive to the small-scale distribution of H I, but in order to make a simple estimate let us note that the comoving volume enclosed by a beam of 10 arcmin and a velocity width of 800 km s^{-1} at $z \approx 1$ is $\sim$1000 Mpc^3. Since the comoving matter density is $\approx 4 \times 10^{10}\,{\rm M}_\odot\,{\rm Mpc}^{-3}$, this means that a typical beam volume contains about $4 \times 10^{13}\,{\rm M}_\odot$ of dark matter and, using our estimate for $f_{\rm H\,I}$, about $4 \times 10^{10}\,{\rm M}_\odot$ of H I. This is greater than the thermal noise for a 5σ detection of $10^{11}\,{\rm M}_\odot$. Clearly this mitigates against our earlier assertion that $\eta_{\rm F} \approx 1$ would be the best situation for unresolved detections. Making $\theta_{\rm FWHM}$ sufficiently small so as to avoid this issue is clearly another important design criterion.

An interferometer or phased array would not, in fact, be sensitive to this smooth mass distribution due to the inherent differencing, but rather to the fluctuations in it, which are typically smaller. But it none the less makes the point that confusion could be significant in this case. This problem is more severe at high redshift since the telescope resolution degrades as redshift increases. Hence, if a telescope were to be ideally suited to detection of clusters at $z \approx 0.5$ then by $z \approx 1$ this would lead to a dilution of the overdensity by factor of $\sim$4.

An alternative way of understanding this is to consider the overdensity in H I expected for a typical cluster. Clusters have overdensities of $\sim$100 in the dark matter and we have already pointed out that, so far as current observations can tell, the value of $f_{\rm H\,I}$ in the field is no more than a factor of $\sim$2 larger than that in a cluster as, for example, in A3128. Naively this might lead one to the belief that it should be easy to detect clusters in this way, and that confusion would not be an important issue for an appropriately designed instrument. However, one has to take into account the effects of redshift-space distortions which stretch out structures in velocity space due to strong gravitational potential. This is the famous Fingers of God effect first seen in the earliest optical redshift surveys.

Although the radius of a cluster which has a velocity dispersion of $\sim$800 km s^{-1} is only about 1 Mpc, the corresponding radial distance is $\Delta v/H_0 \sim 10$ Mpc. Therefore, the physical volume with which one is making a comparison is 10 times larger and the overdensity in redshift space is diluted by a factor $\sim$10. Hence, we can expect a typical cluster to correspond to an rms fluctuation of $\sim$5 in the H I measured in redshift space. It should be possible to detect this by choosing sensible design parameters.

One can make a quantitative estimate of the rms fluctuation in the mass within a volume defined by the beam area and the velocity width $\Delta v = 800$ km s^{-1} by computing

$$\left\langle M_{\rm H\,I}^2 \right\rangle = \Delta z \Delta\Omega \frac{{\rm d}V}{{\rm d}z\,{\rm d}\Omega} \int_0^{M_{\rm lim}} [f_{\rm H\,I}(M,z)]^2 M^2 \frac{{\rm d}n}{{\rm d}M}\,{\rm d}M, \qquad (9)$$

where $\Delta z = (1+z)\Delta v = (1+z) f_{\rm obj}/f_{\rm obs}$ is the width of a fiducial object in redshift space and $\Delta\Omega \propto \theta_{\rm FWHM}^2$ is the beam area. Here, we have assumed that the dominant contribution to the confusion noise comes from uncorrelated Poisson fluctuations, and hence we have ignored the contribution from correlated structure. This is expected when considering very large objects such as clusters. We see that in order to estimate the confusion noise $\langle M_{\rm H\,I}^2 \rangle^{1/2}$ one needs to know $f_{\rm H\,I}(M, z)$ for all M at the particular redshift in question, that is, we need to extrapolate our argument for $f_{\rm H\,I}$ down to the galactic scale and below, which is beyond the scope of this paper. We present the results for an experimental configuration similar to set-up II, that is, $z \approx 0.42$ ($\theta_{\rm FWHM} \approx 5$ arcmin) and $z \approx 1.37$ ($\theta_{\rm FWHM} \approx 8.4$ arcmin) using the fixed value of $f_{\rm H\,I} \approx 10^{-3}$ in Fig. 5, along with some attempt to interpolate between $f_{\rm H\,I} \approx 10^{-2}$ expected for lower mass objects and that which we have argued applies to larger objects, $f_{\rm H\,I} \approx 10^{-3}$. We do this in an ad hoc way using the function

$$f_{\rm H\,I}(M,z) = \frac{10^{-3}(M/{\rm M}_\odot) + 10^{12}}{(M/{\rm M}_\odot) + 10^{14}}, \qquad (10)$$

to gain some insight into what the possible effects could be. We stress that we are not claiming that this expression has physical origin, apart from the fact that it models the correct kind of behaviour. For $M \gg 10^{12}\,{\rm M}_\odot$ this function yields $f_{\rm H\,I} \approx 10^{-3}$ and for $M \ll 10^{12}\,{\rm M}_\odot$, it yields $f_{\rm H\,I} = 10^{-2}$. We note that in such a model the overall H I content of the Universe is larger than in one with the fixed value of $f_{\rm H\,I} = 10^{-3}$. We see that even when using (10) and set-up II at $f_{\rm obs} = 1000$ MHz, one would have a confusion noise which is much lower than the thermal noise mass limit if $M_{\rm H\,I}^{\rm lim} > 2 \times 10^{10}\,{\rm M}_\odot$. Using the constant value of $f_{\rm H\,I=10^{-3}}$ the restriction

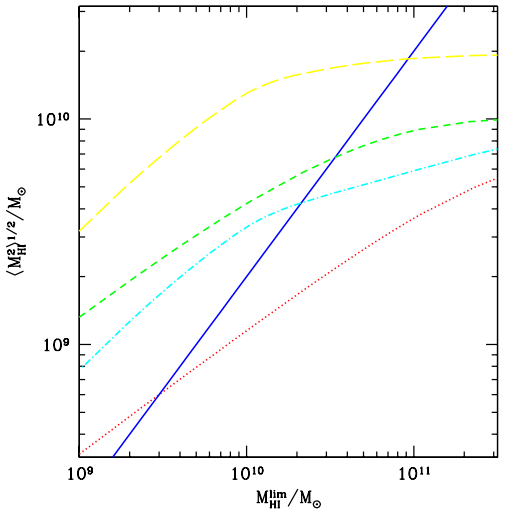

Figure 5. An attempt to estimate the confusion noise as a function of the limiting H I mass of a survey, $M_{\rm H\,I}^{\rm lim}$. The solid line is a representation of the thermal noise: when the other curves are below it the confusion noise is less than the contribution due to thermal noise. The dotted line is for $z = 0.42$ which corresponds to $f_{\rm obs} = 1000$ MHz with $f_{\rm H\,I} = 10^{-3}$ for all M. The equivalent curve for $z = 1.37$ ($f_{\rm obs} = 600$ MHz) is the short dashed line. The dashed-dotted line ($z = 0.42$) and the long dashed line ($z = 1.37$) are those given by using $f_{\rm H\,I}$ from (10). The values of $\theta_{\rm FWHM}$ are those used for set-up II.

is even weaker. Therefore, we can conclude that the confusion noise would be very much lower than the thermal noise value for the fiducial value of $M_{\rm H\,I}^{\rm lim} = 10^{11}\,{\rm M}_\odot$ used throughout this paper. As one might expect things are more restrictive at $f_{\rm obs} = 600$ MHz. For the fixed value of $f_{\rm H\,I}$ one would be restricted to $M_{\rm H\,I}^{\rm lim} > 3 \times 10^{10}\,{\rm M}_\odot$ and if (10) were to be true, the confusion noise would be comparable to the thermal noise at $M_{\rm H\,I}^{\rm lim} = 10^{11}\,{\rm M}_\odot$. In general, one would be more likely to be effected by confusion noise for set-up I due its large beam and less for set-up III and the ATA since they have smaller beams.

4 DETECTING PROTOCLUSTERS

4.1 Number of dark matter haloes at turnaround

The majority of our discussion to date has focused on the possibility of detecting virialized objects. However, we have noted that such objects are likely to have had their H I content depleted relative to the field by at least 20–30 per cent during the process of virialization. It should also be possible to detect objects which are just beginning collapse, at turnaround, when they are just decoupling from the Hubble flow. This could be more efficient at high redshift where there are likely to be very few virialized objects with $M_{\rm vir} > 10^{14}\,{\rm M}_\odot$ and virialized objects could become confused if the beam is too large. At turnaround $\Delta_{\rm c} \approx 5$ and by virtue of the fact that virialization has not yet taken place, it should be possible to use, with some confidence, the value of $f_{\rm H\,I}$ for the field at that time. Assuming a linear rate of star formation between $z = 0$ and $z = 2.4$, one can deduce that

$$f_{\rm H\,I}^{\rm field}(z) \approx (1.2 + 2.8z) \times 10^{-3}, \qquad (11)$$

for the field.

Moreover it is possible that the velocity dispersion of such objects is much less than the fiducial value of $\Delta v = 800$ km s^{-1} used in the earlier parts of the paper for virialized objects. If this is so, the signal-to-noise ratio for a fixed integration time on a source would increase. This is because the signal is proportional to $(\Delta v)^{-1}$ for a fixed H I mass and the noise is proportional to $(\Delta v)^{-1/2}$. This also suppresses the effects of confusion noise discussed in the previous section. The typical value of Δv for such an object is unknown; however, Kumar et al. (1995) have attempted to model this under the assumption that the hydrogen is tracing the dark matter. Formally, the point of turnaround is defined to be that when the average velocity is zero, but the velocity dispersion need not be so. It will be governed by that of the virialized objects within the region which are in the process of merging to build up the cluster. In the subsequent discussion we will allow for Δv to be a modelling parameter, but it is worth discussing various possible values that it might take. If the object at turnaround just comprises two large objects which are collapsing together then the value of Δv will be close to the fiducial value of 800 km s^{-1}. It is possible that such an object contains $10^{14}\,{\rm M}_\odot$ of dark matter and is comprised of $\sim$100 smaller galactic-scale objects in the process of infall with masses of $\sim 10^{12}\,{\rm M}_\odot$. In this case $\Delta v \sim$ 100–200 km s^{-1}. Or it could be that there are many more smaller objects with, say, $\Delta v \sim$ 20–50 km s^{-1}. These possibilities are worth remembering in the subsequent discussion.

We should note that the collapsing objects at $z \approx$ 1–2 are the protoclusters which would have virialized by $z \approx$ 0.5–1. If $R_{\rm ta}$, is the radius of the object at turnaround then the radius of that object once it has virialized is given by $R_{\rm vir} = R_{\rm ta}/2$ assuming a completely matter-dominated universe (Partridge & Peebles 1967; Gunn & Gott 1972). Corrections can be made to include a cosmological constant (Lahav et al. 1991) and also dark energy (Wang & Steinhardt 1998;

Battye & Weller 2003). Similar relations can be derived relating the time at turnaround $t_{\rm ta}$ and that at virialization $t_{\rm vir} = 2t_{\rm ta}$, and the corresponding redshifts, $1 + z_{\rm ta} = 2^{2/3}(1 + z_{\rm vir})$. Liddle & Lyth (2000) give a detailed exposition of this model.

The formula (7) applies to objects with an overdensity $\Delta_{\rm c} = 200$. If one assumes an NFW profile function for the objects with $c = 5$, one can show that $M_{200} \approx M_5/2$, a relation which is approximately true for a wide range of concentration parameters. Taking the fiducial value of $M_{\rm HI}^{\rm lim} = 10^{11}\,{\rm M}_\odot$ and $f_{\rm HI} \approx 5 \times 10^{-3}$ for $f_{\rm obs} = 600$ MHz, we see that $M_5^{\rm lim} \approx 2 \times 10^{13}\,{\rm M}_\odot$ and the corresponding value of M_{200} is approximately $10^{13}\,{\rm M}_\odot$ which can be used in conjunction with Figs 2–4 to deduce the number of collapsing objects that would be found in a given survey, if one ignores the smaller correction from M_{200} to $M_{\rm vir}$. We see that there are ≈20 objects of this size per deg^2 per 1 per cent bandwidth. For $f_{\rm obs} = 400$ MHz the value of $f_{\rm HI}$ is higher and hence there are about the same number of objects; the increase in $f_{\rm HI}$ offsets the smaller number of objects for a given $M_{\rm vir}$. With this large number of objects and the large beams likely at this observing frequency, one might think that one would be close to confusion limited (∼1 object per beam area), but the extra information provided by the velocity information should help avoid this possibility.

Given the larger value of $f_{\rm HI}$ likely for objects which have not virialized, it might be sensible to consider also the possibility of $M_{\rm HI}^{\rm lim} = 10^{12}\,{\rm M}_\odot$, for which the corresponding values of $M_5^{\rm lim}$ and $M_{200}^{\rm lim}$ are 10 times larger. We see that there are still numerous objects (∼0.1 per deg^2 per 1 per cent bandwidth) of this size for $f_{\rm obs} =$ 600 MHz, although there are markedly fewer for lower values of $f_{\rm obs}$ even taking into account the larger value of $f_{\rm HI}$. Achieving this larger limiting mass at 600 MHz should be possible in a fraction of the time required for $10^{11}\,{\rm M}_\odot$.

We should note that at any given redshift the objects at turnaround will be larger than those which are virialized, but have the same overall mass. Fig. 6 shows the angular diameter size of object with a variety of masses at turnaround. We see that these objects are larger than virialized objects of the same mass at the same redshift. At the high redshifts we are considering here, the objects still have an angular diameter less than 5 arcmin and hence they are likely to be unresolved by the instruments under discussion. At lower redshifts such objects would be resolved and it should be more efficient to detect virialized objects.

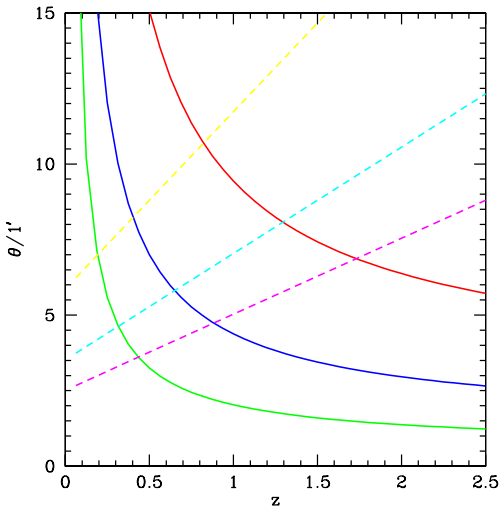

Figure 6. The angular diameter size (twice the radius) of objects which are at turnaround with $M_5 = 10^{14}$, 10^{13} and $10^{12}\,{\rm M}_\odot$ (solid lines, top to bottom, respectively). Included also is the estimated beamsize for the three different phased array configurations (dashed lines, I top, II middle, III bottom). Note that the angular sizes of the collapsing objects are much larger than the objects of the same virial mass at the same redshift.

4.2 Detection-related issues

We can adapt the earlier sensitivity calculation to an arbitrary value of Δv and $M_{\rm HI}^{\rm lim}$. In particular one can compute $\mathcal{R}_{\Delta v}$, $S_{\Delta v}$ and $t_{\Delta v}$ from $\mathcal{R}_{800}$, S_{800} and t_{800}. We see that the instantaneous survey sensitivity is given by

$$\mathcal{R}_{\Delta v} = \mathcal{R}_{800}\left(\frac{800\,{\rm km\,s^{-1}}}{\Delta v}\right)^{1/2}. \tag{12}$$

For an H I mass of $M_{\rm HI}^{\rm lim}$ the required flux density is

$$S_{\Delta v} = S_{800}\left(\frac{800\,{\rm km\,s^{-1}}}{\Delta v}\right)\left(\frac{M_{\rm HI}^{\rm lim}}{10^{11}\,{\rm M}_\odot}\right), \tag{13}$$

and the actual integration time required to achieve a 5σ detection of such an object on an area 100 deg^2 is

$$t_{\Delta v} = t_{800}\left(\frac{\Delta v}{800\,{\rm km\,s^{-1}}}\right)\left(\frac{10^{11}\,{\rm M}_\odot}{M_{\rm HI}^{\rm lim}}\right)^2. \tag{14}$$

If we assume that $\Delta v = 100$ km s^{-1} then we see that it would be possible to cover close to 260 deg^2 in a day of integration $M_{\rm HI}^{\rm lim} = 10^{12}\,{\rm M}_\odot$ using set-up II at $f_{\rm obs} = 600$ MHz. Since there are ≈0.15 objects per deg^2 per 1 per cent bandwidth above the corresponding mass limit then one would hope to find around 200 collapsing objects. It would take around 10 times as long to do a similar survey at $f_{\rm obs} = 400$ MHz and one would find approximately 10 times fewer objects taking into account the larger value of $f_{\rm HI}$ and the very much reduced number of objects with a given mass. None the less, ∼20 objects in 10 d of integration time is definitely worthwhile.

Neither of the above survey parameters are optimal. If we assume that Δv is either weakly dependent on M_5 or not at all, then the optimal mass limit would correspond to the value of M_{200} for which $n = 2$ in contrast to the virialized case. Assuming that $M_{200} \approx M_{\rm vir}$, we see from Fig. 4 that the optimal values of M_{200} are $3 \times 10^{13}\,{\rm M}_\odot$ and $2 \times 10^{12}\,{\rm M}_\odot$ for $f_{\rm obs} = 600$ MHz and 400 MHz, respectively. The corresponding H I mass limits are $M_{\rm HI}^{\rm lim} = 3 \times 10^{11}\,{\rm M}_\odot$ and $3 \times 10^{10}\,{\rm M}_\odot$. For $f_{\rm obs} = 600$ MHz, it would require 4.3 d of integration time to achieve this optimum depth on 100 deg^2. There are ≈2.3 objects per deg^2 per 1 per cent bandwidth and hence such a survey would find ≈1150 objects (which is more than the ≈860 that would be found by mapping 1110 deg^2 in 4.3 d with a limit of $10^{12}\,{\rm M}_\odot$ as suggested above).

5 CONCLUSIONS

To summarize, we have shown that instruments which might be built within the next few years have a realistic chance of detecting large objects, both virialized and collapsing, using H I emission as their tracer, opening a new window on the Universe. If a detection threshold of $M_{\rm HI}^{\rm lim} \approx 10^{11}\,{\rm M}_\odot$ can be achieved at around $z \approx 0.4$ then it should be possible to find a surprisingly large number of virialized objects. Similarly, it should be possible to detect many objects at turnaround with $z > 1$. We have also made comments as to the optimal design of an instrument and the survey strategy for these applications. Clearly, more sophisticated simulations of the large-scale distribution of H I are required, but we believe that the basic picture we have put forward is likely to remain.

It is clear that the detection of the large number of objects, both virialized and collapsing, predicted in this paper could have a significant impact on our understanding of the Universe. In the regime where one can optimally detect virialized clusters ($z < 1$) it should be possible, using the extra velocity information, to compute accurately the dark matter mass of each of the objects which are detected and establish the selection function. Since the number of virialized objects is sensitive to cosmological parameters, accurate estimates of $\Omega_{\rm m}(z)$ and σ_8 should be possible. Such measurements could shed some light on the nature of the dark energy thought to pervade the Universe; the properties of the dark energy may also be accessible to such an analysis. The nature of the collapsing structures for $z > 1$ is also of significant interest since they would enhance our understanding of the early stages of galaxy formation. We have used all the available information to make estimates of the number of objects which would be found. However, we have also noted that these are somewhat uncertain, particularly the velocity structure. Clearly the detection of a large number of objects will have a significant impact on our understanding of the distribution of H I at high redshifts and we believe that we have shown that it is worth taking very seriously.

ACKNOWLEDGMENTS

We would like to thank Martin Rees, Martin Haehnelt, Clive Dickinson, Ian Browne and Peter Wilkinson for extremely helpful comments and suggestions. RAB is funded by PPARC and JW in part by Kings College.

REFERENCES

Appleton P. N., Davies R. D., Stephenson R. J., 1981, MNRAS, 195, 327
Bagla J. S., Nath B., Padmanabhan T., 1997, MNRAS, 289, 671
Banks G. D. et al., 1999, ApJ, 524, 612
Barnes D. G. et al., 2001, MNRAS, 322, 486
Battye R., Weller J., 2003, Phys. Rev. D, 68, 083506
Butcher H., Oemler A., 1984, ApJ, 285, 426
Chengalur J. N., Braun R., Wieringa M., 2001, A&A, 372, 768
Contaldi C. R., Hoekstra H., Lewis A., 2003, Phys. Rev. Lett., 90, 221303
Dressler A. et al., 1997, ApJ, 490, 577
Ellison S. L. et al., 2001, A&A, 379, 393
Evrard A. E. et al., 2002, ApJ, 573, 7
Field G. B., 1958, Proc. Inst. Radio Eng., 46, 240
Furlanetto S. R., Sokasian A., Hernquist L., 2004, MNRAS, 347, 187
Gnedin N. Y., Shaver P. A., 2004, ApJ, 608, 611
Gunn J. E., Gott J. R., 1972, ApJ, 176, 1
Hutchmeier W. K., Richter O. G., 1986, A&AS, 64, 111
Kilborn V., Webster R. L., Stavely-Smith L., 1999, Publ. Astron. Soc. Aust., 16, 8
Kumar A., Padmanabhan T., Subramanian K., 1995, MNRAS, 272, 544
Lahav O., Rees M. J., Lilje. P. B., Primack J. R., 1991, MNRAS, 251, 128
Lang R. et al., 2003, MNRAS, 342, 738
Liddle A., Lyth D., 2000, Cosmological Inflation and Large-Scale Structure. Cambridge Univ. Press, Cambridge
Loeb A., Zaldarriaga M., 2004, Phys. Rev. Lett., 92, 211301
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Partridge R. B., Peebles P. J. E., 1967, ApJ, 147, 868
Peacock J., 2003, Phil. Trans. Roy. Soc. London, A, 361, 2479
Roberts M. S., 1975, in Sandage A., Sandage M., Kristin J., eds, Galaxies in the Universe. Univ. of Chicago Press, Chicago, p. 326
Roberts M. S., Haynes M. P., 1994, ARA&A, 32, 115
Ryan-Weber E. et al., 2002, AJ, 124, 1954
Scott D., Rees M. J., 1990, MNRAS, 247, 510
Spergel D. N. et al., 2003, ApJS, 148, 175
Subramanian K., Padmanabhan T., 1993, MNRAS, 265, 101
Tegmark M. et al., 2004, Phys. Rev. D, 69, 103501
Wang L., Steinhardt P. J., 1998, ApJ, 508, 483
Zaldarriga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622
Zwaan M. A., Briggs F. H., Sprayberry D., Sorar E., 1997, ApJ, 490, 173
Zwaan M. A., van Dokkum P. G., Verheijen M. A. W., 2001, Sci, 293, 1800
Zwaan M. A. et al., 2003, AJ, 125, 2842

This paper has been typeset from a TEX/LATEX file prepared by the author.