Fast Fourier transform telescope

Fast Fourier transform telescope

Max Tegmark

Department of Physics & MIT Kavli Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Matias Zaldarriaga

Center for Astrophysics, Harvard University, Cambridge, Massachusetts 02138, USA

We propose an all-digital telescope for 21 cm tomography, which combines key advantages of both single dishes and interferometers. The electric field is digitized by antennas on a rectangular grid, after which a series of fast Fourier transforms recovers simultaneous multifrequency images of up to half the sky. Thanks to Moore's law, the bandwidth up to which this is feasible has now reached about 1 GHz, and will likely continue doubling every couple of years. The main advantages over a single dish telescope are cost and orders of magnitude larger field-of-view, translating into dramatically better sensitivity for large-area surveys. The key advantages over traditional interferometers are cost (the correlator computational cost for an N-element array scales as $N\log_2 N$ rather than N^2) and a compact synthesized beam. We argue that 21 cm tomography could be an ideal first application of a very large fast Fourier transform telescope, which would provide both massive sensitivity improvements per dollar and mitigate the off-beam point source foreground problem with its clean beam. Another potentially interesting application is cosmic microwave background polarization.

I. INTRODUCTION

Since Galileo first pointed his telescope skyward, design innovations have improved attainable sensitivity, resolution and wavelength coverage by many orders of magnitude. Yet we are still far from the ultimate telescope that simultaneously observes light of all wavelengths from all directions, so there is still room for improvement.

From a mathematical point of view, telescopes are Fourier transformers. We want to know individual Fourier modes **k** of the electromagnetic field, as their direction $\hat{\mathbf{k}}$ encodes our image and their magnitude $k = \omega/c = 2\pi/\lambda$ encodes the wavelength, but the field at a given spacetime point $(\mathbf{r}, t)$ tells us only a sum of all these Fourier modes weighted by phase factors $e^{i[\mathbf{k}\cdot\mathbf{r}+\omega t]}$.

Traditional telescopes perform the spatial Fourier transform from **r**-space to **k**-space by approximate analog means using lenses or mirrors, which are accurate across a relatively small field of view, and perform the temporal Fourier transform from t to ω using slits, gratings or bandpass filters. Traditional interferometers used analog means to separate frequencies and measure electromagnetic field correlations between different receivers, then Fourier-transformed to **r**-space digitally, using computers. In the trade off between resolution, sensitivity and cost, single dish telescopes and interferometers are highly complementary, and which is best depends on the science goal at hand.

Thanks to Moore's law, it has very recently become possible to build all-digital interferometers up to about 1 GHz, where the analog signal is digitized right at each antenna and subsequent correlations and Fourier transforms are done by computers. In addition to reducing various systematic errors, this digital revolution enables the "fast Fourier transform telescope" or "omniscope" that we describe in this paper. We will show that it acts much like a single dish telescope with a dramatically larger field of view, yet is potentially much cheaper than a standard interferometer with comparable area. If a modern all-digital interferometer such as the Murchison widefield array (MWA) [1] is scaled up to a very large number of antennas N, its price becomes completely dominated by the computing hardware cost for performing of order N^2 correlations between all its antenna pairs. The key idea behind the FFT Telescope is that, if the antennas are arranged on a rectangular grid, this cost can be cut to scale merely as $N\log_2 N$ using fast Fourier transforms. As we will see, this design also eliminates the need for individual antennas that are pointable (mechanically or electronically), and has the potential to dramatically improve the sensitivity for some applications of future telescopes like the square kilometer array without increasing their cost.

This basic idea is rather obvious, so when we had it, we wondered why nothing like the massive all-sky low-frequency telescope that we are proposing had ever been built. We have since found other applications of the idea in the astronomy and engineering literature dating as far back as the early days of radio astronomy [2–11], and it is clear that the answer lies in lack of both computer power and good science applications. Moore's law has only recently enabled A/D conversion up to the GHz range, so in older work, Fourier transforms were done by analog means and usually in only one dimension (*e.g.*, using a so-called

Butler matrix [2]), severely limiting the number of antennas that could be used. For example, the 45 MHz interferometer in [3] used six elements. Moreover, to keep the number of elements modest while maintaining large collecting area, the elements themselves would be dishes or interconnected antennas that observed only a small fraction of the sky at any one time. A Japanese group worked on an analog 8×8 FFT Telescope about 15 years ago for studying transient radio sources [4,5], and then upgraded it to digital signal processing aiming for a 16×16 array with a field of view just under 1°. Electronics from this effort is also used in the 1-dimensional 8-element Nasu Interferometer [8].

Most traditional radio astronomy applications involve mapping objects subtending a small angle surrounded by darker background sky, requiring only enough sensitivity to detect the object itself. For most such cases, conventional radio dishes and interferometers work well, and an FFT Telescope (hereafter FFTT) is neither necessary nor advantageous. For the emerging field of 21 cm tomography, which holds the potential to 1 day overtake the microwave background as our most sensitive cosmological probe [12–18], the challenge is completely different: it involves mapping a faint and diffuse cosmic signal that covers all of the sky and needs to be separated from foreground contamination that is many orders of magnitude brighter, requiring extreme sensitivity and beam control. This 21 cm science application and the major efforts devoted to it by experiments such as MWA [1], LOFAR [19], PAPER [20], 21CMA [21], GMRT [22,23], and SKA [24] makes our paper timely.

An interesting recent development is a North American effort [9,10] to do 21 cm cosmology with a one-dimensional array of cylindrical telescopes that can be analyzed with FFT's, in the spirit of the Cambridge 1.7 m instrument from 1957, exploiting Earth rotation to fill in the missing two-dimensional information [9,10]. We will provide a detailed analysis of this design below, arguing that is complementary to the 2D FFTT at higher frequencies while a 2D FFTT provides sharper cosmological constraints at low frequencies.

The rest of this paper is organized as follows. In Sec. II, we describe our proposed design for FFT Telescopes. In Sec. III, we compare the figures of merit of different types of telescopes, and argue that the FFT Telescope is complementary to both single dish telescopes and standard interferometers. We identify the regimes where each of the three is preferable to the other two. In Sec. IV, we focus on the regime where the FFT Telescope is ideal, which is when you have strong needs for sensitivity and beam cleanliness but not resolution, and argue that 21 cm tomography may be a promising first application for it. We also comment briefly on cosmic microwave background applications. We summarize our conclusions in Sec. V and relegate various technical details to a series of appendices.

II. HOW THE FFT TELESCOPE WORKS

In this section, we describe the basic design and data processing algorithm for the FFT Telescope. We first summarize the relevant mathematical formalism, then discuss data processing, and conclude by discussing some practical issues. For a comprehensive discussion of radio interferometry techniques, see e.g. [25].

A. Interferometry without the flat sky approximation

Since the FFT Telescope images half the sky at once, the flat-sky approximation that is common in radio astronomy is not valid. We therefore start by briefly summarizing the general curved-sky results formalism. Suppose we have a set of antennas at positions $\mathbf{r}_n$ with sky responses $\mathbf{B}_n(\hat{\mathbf{k}})$ at a fixed frequency $\omega = ck$, $n = 1, \ldots$, and a sky signal $\mathbf{s}(\hat{\mathbf{k}})$ from the direction given by the unit vector $-\hat{\mathbf{k}}$ (this radiation thus travels in the direction $+\hat{\mathbf{k}}$). The data measured by each antenna in response to a sky signal $\mathbf{s}(\hat{\mathbf{k}})$ is then

$$\mathbf{d}_n = \int e^{-i[\mathbf{k}\cdot\mathbf{r}_n + \omega t]} \mathbf{B}_n(\hat{\mathbf{k}})\mathbf{s}(\hat{\mathbf{k}})d\Omega_k. \tag{1}$$

Details related to polarization are covered below in Appendix A, but are irrelevant for the present section. For now, all that matters is that $\mathbf{s}(\hat{\mathbf{k}})$ specifies the sky signal, $\mathbf{d}_n$ specifies the data that is recorded, and $\mathbf{B}_n(\hat{\mathbf{k}})$ specifies the relation between the two.[1] Specifically, $\mathbf{s}$ is the so-called Jones vector (a two-dimensional complex vector field giving the electric field components—with phase—in two orthogonal directions), $\mathbf{d}_n$ is a vector containing the two complex numbers measured by the antenna, and $\mathbf{B}_n(\hat{\mathbf{r}})$, the so-called *primary beam*, is a 2×2 complex matrix field that defines both the polarization response and the sky response (beam pattern) of the antenna. The only properties of Eq. (1) that matter for our derivation below are that it is a linear relation (which comes from the linearity of Maxwell's equations) and that it contains the phase factor $e^{-i\mathbf{k}\cdot\mathbf{r}_n}$ (which comes from the extra path length $\hat{\mathbf{k}} \cdot \mathbf{r}_n$ that a wave must travel to get to the antenna location $\mathbf{r}_n$).

The sky signal $\mathbf{s}(\hat{\mathbf{k}})$ has a slow time dependence because the sky rotates overhead, because of variable astronomical sources, and because of distorting atmospheric/ionospheric

---

[1]If one wishes to observe the sky at frequencies ν higher than current technology can sample directly ($\gtrsim 1$ GHz), then one can extract a bandwidth $\Delta\nu \lesssim 1$ GHz in this high frequency range using standard radio engineering techniques (first an analog frequency mixer multiplies the input signal with that from a local oscillator, then an analog low-pass-filter removes frequencies above $\Delta\nu$, and finally the signal is A/D converted). The net effect of this is simply to replace $e^{-i\omega t}$ in Eq. (1) by $e^{-i(\omega-\omega_0)t}$ for some conveniently chosen local oscillator frequency $\omega_0 = 2\pi\nu_0$. It is thus the bandwidth $\Delta\nu$ rather than the actual frequencies ν that are limited by Moore's Law.

fluctuations. However, since these changes are many orders of magnitude slower than the electric field fluctuation timescale ω^{-1}, we can to an excellent approximation treat Eq. (1) as exact for a snapshot of the sky. Below we derive how to recover the snapshot sky image from these raw measurements; only when coadding different snapshots does one need to take sky rotation and other variability into account.

The statements above hold for *any* telescope array. For the special case of the FFT Telescope, all antennas have approximately identical beam patterns ($\mathbf{B}_n = \mathbf{B}$) and lie in a plane, which we can without loss of generality take to be the $z = 0$ plane so that $\hat{\mathbf{z}} \cdot \mathbf{r}_n = 0$. Using the fact that

$$d\Omega_k = \sin\theta d\theta d\phi = \frac{dk_x dk_y}{k\sqrt{k^2 - k_\perp^2}}, \qquad (2)$$

where $k_\perp \equiv \sqrt{k_x^2 + k_y^2}$ is the length of the component of the $\mathbf{k}$-vector perpendicular to the z-axis, we can rewrite Eq. (1) as a two-dimensional Fourier transform

$$\mathbf{d}_n = \int e^{-i[\mathbf{q}\cdot\mathbf{x}_n + \omega t]} \frac{\mathbf{B}(\mathbf{q})\mathbf{s}(\mathbf{q})}{k\sqrt{k^2 - q^2}} d^2q = \hat{\mathbf{s}}_\mathbf{B}(\mathbf{x}_n)e^{-i\omega t}, \quad (3)$$

where we have defined the two-dimensional vectors

$$\mathbf{q} = \begin{pmatrix} k_x \\ k_y \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad (4)$$

and the function

$$\mathbf{s}_\mathbf{B}(\mathbf{q}) \equiv \frac{\mathbf{B}(\mathbf{q})\mathbf{s}(\mathbf{q})}{k\sqrt{k^2 - q^2}}. \qquad (5)$$

Here the two-dimensional function $\mathbf{s}(\mathbf{q})$ is defined to equal $\mathbf{s}(q_x, q_y, -[k^2 - q_x^2 - q_y^2]^{1/2})$ when $q \equiv |\mathbf{q}| < k$, zero otherwise, and $\mathbf{B}(\mathbf{q})$ is defined analogously. $\mathbf{s}_\mathbf{B}$ can therefore be thought of as the windowed, weighted and zero-padded sky signal. Equation (3) holds under the assumption that $\mathbf{B}(\hat{\mathbf{k}})$ vanishes for $k_z > 0$, *i.e.*, that a ground screen eliminates all response to radiation heading up from below the horizon, so that we can limit the integration over solid angle to radiation pointed towards the lower hemisphere. Note that for our application, the simple Fourier relation of Eq. (3) is exact, and that none of the approximations that are commonly used in radio astronomy for the so-called "w-term" (see Eq. 3.7 in [25]) are needed.

One usually models the fields arriving from different directions as uncorrelated, so that

$$\langle \mathbf{s}(\hat{\mathbf{k}})\mathbf{s}(\hat{\mathbf{k}}')^\dagger \rangle = \delta(\hat{\mathbf{k}}, \hat{\mathbf{k}}')\mathbf{S}(\hat{\mathbf{k}}), \qquad (6)$$

where $\mathbf{S}(\hat{\mathbf{k}})$ is the 2×2 sky intensity Stokes matrix and the spherical δ-function satisfies

$$\delta(\hat{\mathbf{k}}, \hat{\mathbf{k}}') = \delta(\mathbf{q} - \mathbf{q}')k\sqrt{k^2 - k_\perp^2} \qquad (7)$$

so that $\int \delta(\hat{\mathbf{k}}, \hat{\mathbf{k}}')g(\hat{\mathbf{k}}')d\Omega_k' = g(\hat{\mathbf{k}})$ for any function g.

Combining Eq. (3) with Eq. (6) implies that the correlation between two measurements, traditionally referred to as a *visibility*, has the expectation value

$$\langle \mathbf{d}_m\mathbf{d}_n^\dagger \rangle = \int e^{-i\mathbf{q}\cdot(\mathbf{x}_m - \mathbf{x}_n)} \frac{\mathbf{B}(\mathbf{q})^\dagger\mathbf{S}(\mathbf{q})\mathbf{B}(\mathbf{q})}{k\sqrt{k^2 - q^2}} d^2q$$

$$= \hat{\mathbf{S}}_B(\mathbf{x}_m - \mathbf{x}_n), \qquad (8)$$

where $\hat{\mathbf{S}}_B(\mathbf{x})$ is the Fourier transform of:

$$\mathbf{S}_\mathbf{B}(\mathbf{q}) \equiv \frac{\mathbf{B}(\mathbf{q})^\dagger\mathbf{S}(\mathbf{q})\mathbf{B}(\mathbf{q})}{k\sqrt{k^2 - q^2}}. \qquad (9)$$

is the beam-weighted, projection-weighted and zero-padded sky brightness map.

In summary, things are not significantly more complicated than in standard interferometry in small sky patches (where the flat sky approximation is customarily made). One can therefore follow the usual radio astronomy procedure with minimal modifications: first measure $\hat{\mathbf{S}}_\mathbf{B}(\Delta\mathbf{x})$ at a large number of baselines $\Delta\mathbf{x}$ corresponding to different antenna separations $\mathbf{x}_m - \mathbf{x}_n$, then use these measurements to estimate the Fourier transform of this function, $\mathbf{S}_\mathbf{B}(\mathbf{q})$, and finally recover the desired sky map $\mathbf{S}$ by inverting Eq. (9):

$$\mathbf{S}(\mathbf{q}) = k\sqrt{k^2 - q^2}\mathbf{B}(\mathbf{q})^{-\dagger}\mathbf{S}_\mathbf{B}(\mathbf{q})\mathbf{B}(\mathbf{q})^{-1}. \qquad (10)$$

B. FFTT analysis algorithm

Equation (8) shows that the Fourier-transformed beam-convolved sky $\hat{\mathbf{S}}_B$ is measured at each baseline, *i.e.*, at each separation vector $\mathbf{x}_m - \mathbf{x}_n$ for an antenna pair. A traditional correlating array with N_a antennas measures all $N_a(N_a - 1)/2$ such pairwise correlations, and optionally fills in more missing parts of the Fourier plane exploiting Earth rotation. Since the cost of antennas, amplifiers, A/D-converters, *etc.* scales roughly linearly with N_a, this means that the cost of a truly massive array (like what may be needed for precision cosmology with 21 cm tomography [18]) will be dominated by the cost of the computing power for calculating the correlations, which scales like N_a^2.

For the FFT Telescope, the N_a antenna positions $\mathbf{r}_n$ are chosen to form a rectangular grid. This means that the all $N_a(N_a - 1)/2 \sim N_a^2$ baselines also fall on a rectangular grid, typically with any given baseline being measured by many different antenna pairs.

The sums of $\mathbf{d}_m\mathbf{d}_n^\dagger$ for each baseline can be computed with only of order $N_a\log_2 N_a$ (as opposed to N_a^2) operations by using fast Fourier transforms. Essentially, what we wish to measure in the Fourier plane are the antenna measurements (laid out on a 2D grid) convolved with themselves, and this naively N_a^2 convolution can be reduced to an FFT, a squaring, and an inverse FFT.

In fact, Eq. (3) shows that after FFT-ing the 2D antenna grid of data $\mathbf{d}_n$, one already has the two electric field

components $\mathbf{s_B(q)}$ from each sky direction, and can multiply them to measure the sky intensity from each direction (Stokes I, Q, U and V) without any need to return to Fourier space, as illustrated in Fig. 1. This procedure is then repeated for each time sample and each frequency, and the many intensity maps at each frequency are averaged (after compensating for sky rotation, ionospheric motion, *etc.*) to improve signal-to-noise.

It should be noted that the computational cost for the entire FFT Telescope signal processing pipeline is (up to some relatively unimportant log factors) merely proportional to the total number of numbers measured by all antennas throughout the duration of the observations. In particular, the time required for the spatial FFT operations is of the same order as the time required for the time-domain FFT's that are used to separate out the different frequencies from the time signal using standard digital
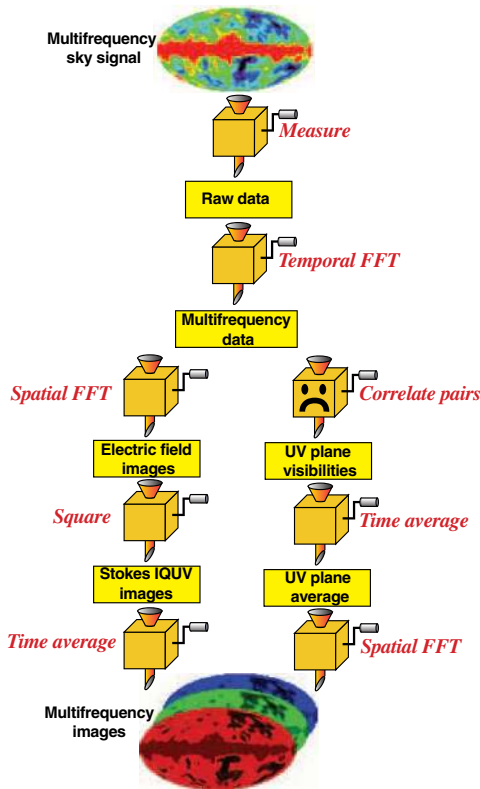

FIG. 1 (color online). When the antennas are arranged in a rectangular grid as in the FFT Telescope, the signal processing pipeline can be dramatically accelerated by eliminating the correlation step (indicated by a sad face): its computational cost scales as N_a^2, because it must be performed for all *pairs* of antennas, whereas all other steps shown scale linearly with N_a. The left and right branches recover the same images on average, but with slightly different noise. Alternatively, if desired, the FFT Telescope can produce images that are mathematically identical to those of the right branch (while retaining the speed advantage) by replacing the correlation step marked by the sad face by a spatial FFT, "squaring," and an inverse spatial FFT.

filtering. If the antennas form an $n_x \times n_y$ rectangular array, so that $N_a = n_x n_y$, and each antenna measures n_t different time samples (for a particular polarization), then it is helpful to imagine this data arranged in a three-dimensional $n_x \times n_y \times n_t$ block. The temporal and spatial FFT's (left branch in Fig. 1) together correspond to a 3D FFT of this block, performed by 3 one-dimensional FFT operations:

(1) For each antenna, FFT in the t-direction.
(2) For each time and antenna row, FFT in the x-direction.
(3) For each time and antenna column, FFT in the y-direction.

One processes one such block for each of the two polarizations. These three steps each involve of order $n_t n_x n_y$ multiplications (up to order-of-unity factors $\log n_t$, $\log n_x$ and $\log n_y$), and it is easy to show that the number of operations for the three steps combined scales as $(n_t n_x n_y) \log(n_t n_x n_y)$, *i.e.*, depends only on the total amount of data $n_t n_x n_y$. After step 3, one has the two electric field components from each direction at each frequency. Phase and amplitude calibration of each antenna/amplifier system is normally performed after step 1. If one is interested in sharp pulses that are not well-localized in frequency, one may opt to skip step 1 or perform a broad band-pass filtering rather than a full spectral separation.

The FFT Telescope cuts down not only on CPU time, but also on data storage costs, since the amount of data obtained at each snapshot scales as number of time samples taken times N_a rather than N_a^2.

In a conventional interferometer, antennas are correlated only with other antennas and not with themselves, to eliminate noise bias. This can be trivially incorporated in the FFTT analysis pipeline as well by setting the pixel at the origin of the UV plane (corresponding to zero baseline) to zero, and is mathematically equivalent to removing the mean form the recovered sky map.

C. Practical considerations

Although we have laid out the mathematical and computational framework for an FFT Telescope above, there are a number of practical issues that require better understanding before building a massive scale FFT Telescope.

As we will quantify in Sec. III below, the main advantages of an FFT Telescope relative to single dish telescopes and conventional interferometers emerge when the number of antennas N_a is very large. A successful FFTT design should therefore emphasize simplicity and mass-production, and minimize hardware costs. To exploit the FFT data processing speedup, care must be taken to make the antenna array as uniform as possible. The locations $\mathbf{r}_i$ of the antennas need to be kept in a planar rectangular grid to within a small fraction of a wavelength, so when selecting the construction site, it is important that the land is quite flat to start with, that bulldozing is feasible, and that there are no immovable obstacles. It is equally important

that the sky response **B**(**r**) be close to identical for all antennas. A ground screen, which can simply consist of cheap wire mesh laid out flat under the entire array, should therefore extend sufficiently far beyond the edges of the array that it can to reasonable accuracy be modeled as an infinite reflecting plane, affecting all antennas in the same way. The sky response **B**(**r**) of an antenna will also be affected by the presence of neighbors: whereas the response of antennas in the central parts of a large array will be essentially identical to one another (and essentially identical to that for an antenna in the middle of an infinite array), antennas near the edges of the array will have significantly different response. Instead of complicating the analysis to incorporate this, it is probably more cost effective to surround the desired array with enough rows of dummy antennas that the active ones can be accurately modeled as being in an infinite array. These dummy antennas could be relatively cheap, as they need not be equipped with amplifiers or other electronics (merely with an equivalent impedance), and no signals are extracted from them.

The FFT algorithm naturally lends itself to a rectangular array of antennas. However, this rectangle need not be square; we saw above that the processing time is independent of the shape of the rectangle, depending only on the total number of antennas, and below we will even discuss the extreme limit where the telescope is one-dimensional. Another interesting alternative to a square FFTT telescope is a circular one, consisting of only those $\pi/4 \approx 79\%$ of the antennas in the square grid that lie within a circle inscribed in the square. This in no way complicates the analysis algorithm, as the FFT's need to be zero-padded in any case, and increases the computational cost for a given collecting area by only about a quarter. The main advantage is a simple rotationally invariant synthesized beam as discussed below. Antennas can also be weighted in software before the spatial FFT do create beams with other desired properties, for example, edge tapering can be used to make the beam even more compact. A third variant is to place the antennas further apart to gain resolution at the price of undersampling the Fourier plane and picking up sidelobes.

III. COMPARISON OF DIFFERENT TYPES OF TELESCOPES

A. Telescopes generalized

In this section, we compare the figures of merit (resolution, sensitivity, cost, *etc.*) of different types of telescopes, summarized in Table I, and argue that the FFT Telescope is complementary to both single dish telescopes and standard interferometers. We identify the regimes where each of the three is preferable to the other two, as summarized in Fig. 2.

It is well-known that all telescopes can be analyzed within a single unified formalism that characterizes their linear response to sky signals and their noise properties. In particular, a single dish telescope can be thought of as an interferometer, where every little piece of the collecting area is an independent antenna, and the correlation is performed by approximate analog means using curved mirrors. This eliminates the costly computational step, but the approximations involved are only valid in a limited field of view (Table I). Traditional interferometers can attain larger field of view and better resolution for a given collecting area, but at a computational cost. The FFT Telescope is a hybrid of the two in the sense that it combines the resolution of a single dish telescope with the all-sky field-of-view of a dipole interferometer—at a potentially much lower cost than either a single dish or a traditional interferometer of the same collecting area. Let us now quantify these statements, starting with angular resolution and its generalization and then turning to sensitivity and cost. We first briefly review some well-known radio astronomy formalism that is required for our applications.

B. Angular resolution and the beam function B_ℓ

The angular resolution of the telescopes we will compare are all much better than a radian, so we can approximate the sky as flat for the purposes of this section. If we ignore polarization, then it is well-known that the response of an interferometer to radiation intensity coming from

TABLE I. How telescope properties scale with dish size D, collecting area A and wavelength λ. We assume that the standard interferometer has $N_a \sim A/D^2$ separate dishes with a maximum separation D_{max} that together cover a fraction $f^{cover} \sim A/D_{max}^2$ of the total array region rather uniformly.

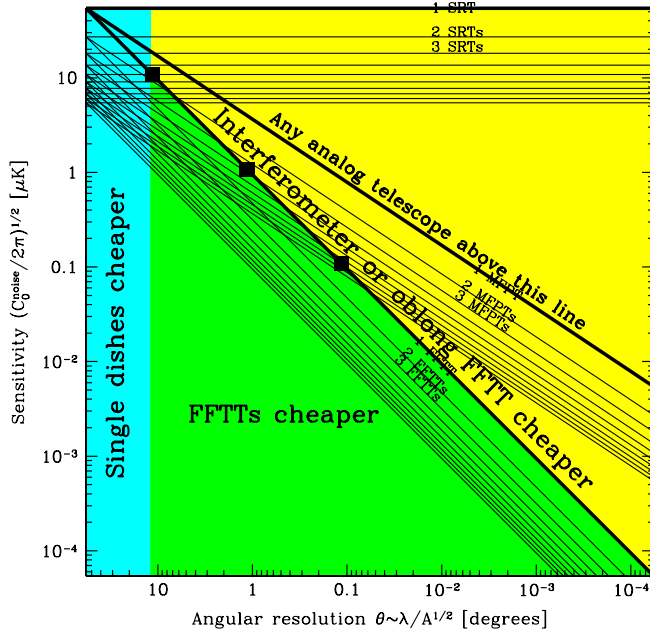|		Single Dish Telescopes		Interferometers	
		Single Receiver Telescope	Maximal Focal Plane Telescope	FFT Telescope	Standard Interferometric Telescope
Resolution	θ_{min}	$\frac{\lambda}{D}$	$\frac{\lambda}{D}$	$\frac{\lambda}{D}$	$\frac{\lambda}{D_{max}}$
Field of view	θ_{max}	$\frac{\lambda}{D}$	$\left(\frac{\lambda}{D}\right)^{1/3}$	1	$\frac{\lambda}{D}$
Resolution elements	n	1	$\left(\frac{D}{\lambda}\right)^{4/3}$	$\left(\frac{D}{\lambda}\right)^2$	$\left(\frac{D_{max}}{D}\right)^2$
Etendu	$A\Omega$	λ^2	$D^{4/3}\lambda^{2/3}$	D^2	$\frac{\lambda^2 A}{D^2}$
Sensitivity	C_0^{noise}	λT_{sys}^2	$\frac{\lambda^{7/3} T_{sys}^2}{A^{2/3}}$	$\frac{\lambda^3 T_{sys}^2}{A}$	$\frac{\lambda D^2 T_{sys}^2}{A f^{cover}}$
Cost	\$	$A^{1.35}$	$A^{1.35}$	A	A^2

FIG. 2 (color online). Angular resolution and sensitivity are compared for different telescope designs, assuming that half the sky is surveyed during 4000 hours at an observing frequency 150 MHz, with 0.1% bandwidth and 200 K system temperature. Since $\delta T_\ell = \sqrt{\ell^2 C_0^{\text{noise}}/2\pi}$, $\delta T_\ell = 10$ mK at $\ell = 1000$ corresponds to 10 μK on the vertical axis. The parameters of any analog telescope (using focusing optics rather than digital beamforming) lie in the upper right triangle between the limiting cases of the single receiver telescope (SRT; heavy horizontal line) and the single dish telescope with a maximal focal plane (MFPT; heavy line of slope 2/3). The parameters of a fast Fourier transform telescope (FFTT) lie on the heavy horizontal line of slope -1, with solid squares corresponding to squares FFTTs of side 10 m, 100 m and 1000 m, respectively. Moving their antennas further apart (reducing f^{cover} with A fixed) would move these squares along a 45° line up to the right. Improved sensitivity at fixed resolution can be attained by building multiple telescopes (thin parallel lines correspond to 2, 3, ..., 10 copies). As explained in the text, SDTs, SITs and FFTTs are complementary: the cheapest solution is offered by SDTs for low resolution, FFTTs for high sensitivity $(C_0^{\text{noise}})^{1/2} \lesssim \theta \times 2$ μK, and elongated FFTTs or standard interferometers for high resolution $\theta \lesssim (C_0^{\text{noise}})^{1/2}/2$ μK.

near the local zenith[2] and traveling in the direction **k** is $\check{W}(k_x, k_y)$, the inverse Fourier transform of the function $W(\Delta x)$ that gives the distribution of baselines Δx. For the classic example of a single dish telescope of radius R, this formula gives

---

[2]If we are imaging objects much smaller than a radian centered at a zenith angle θ, we recover the same formula as above, but with the synthesized beam compressed by a factor $\cos\theta$ in one direction, as the source effectively sees the array at a slanting angle, compressed by $\cos\theta$ in one direction.
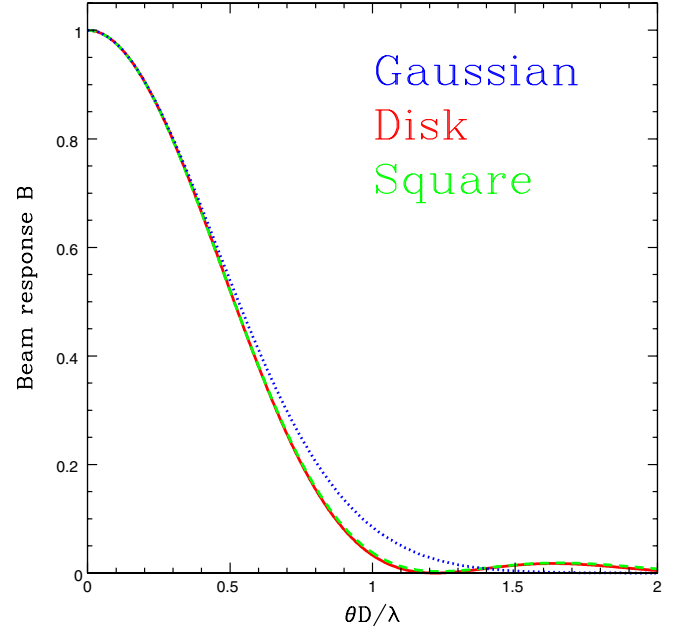
FIG. 3 (color online). The familiar Airy pattern that constitutes the sky response of a circular telescope dish of diameter D is compared with the azimuthally averaged response of a square telescope and one with a Gaussian tapered aperture. The square has side $0.87D$ to have the same FWHM, and the Gaussian has standard deviation $0.45D$ to give comparable response for $\theta D/\lambda \ll 1$.

$$\check{W}(k_x, k_y) = \left[\frac{2J_1(Rk_\perp)}{Rk_\perp}\right]^2, \qquad (11)$$

the famous Airy pattern plotted in Fig. 3. Here $k_\perp = (k_x^2 + k_y^2)^{1/2} = 2\pi\theta/\lambda$, where θ is the angle to the zenith. When the beam is asymmetric, we will mainly be interested in the azimuthally averaged beam which again depends only on θ; the result for a square telescope like the fully instrumented FFTT plotted for comparison.[3] The figure also shows a Gaussian beam, which may be a better approximation for an optical telescope when the seeing is poor.

For these three cases, the shapes are seen to be sufficiently similar that, for many purposes, all one needs to

---

[3]For a telescope with a square dish of side $D = 2R$, convolving the square with itself gives the baseline distribution

$$W(\Delta x) \propto (2R - |\Delta x|)(2R - |\Delta y|) \qquad (12)$$

when $|\Delta x| < 2R$ and $|\Delta y| < 2R$, zero otherwise. Writing $\Delta x = r(\cos\varphi, \sin\varphi)$ and averaging over the azimuthal angle φ gives

$$W(r) = \begin{cases} 1 - \frac{1}{\pi}(4 - \frac{r}{D})\frac{r}{D} & \text{if } r \leq D, \\ 1 - \frac{4}{\pi}[\frac{1}{2} - \sqrt{\frac{r^2}{D^2} - 1} + \frac{r^2}{4D^2} + \cos^{-1}\frac{D}{r}] & \text{if } r \geq D, \end{cases}$$

$$(13)$$

which is plotted in Fig. 4. The synthesized beam is simply

$$\check{W}(\mathbf{k}) = j_0(Rk_x)j_0(Rk_y), \qquad (14)$$

and the azimuthal average of this function is plotted in Fig. 3.

TABLE II.   Different measures of angular resolution, measured in units of D/λ.

	rms	FWHM	First null
Disk of diameter D	0.53	1.03	1.22
Square of side D	0.49	0.89	1.07
Gaussian with $\sigma = D$	1	2.35	∞

know about the beam can be encoded in a single number specifying its width. The most popular choices in astronomy are summarized in Table II: the rms (the root-mean-squared value of θ averaged across the beam), the FWHM (twice the θ-value where $B(\theta)$ has dropped to half its central value) and the first null (the smallest θ at which $B(\theta) = 0$). We will mainly focus on the FWHM in our cost comparison below.

The primary beam $\mathbf{B}(\mathbf{q})$ that was introduced in Sec. II A can itself be derived from this same formalism by considering each piece of an antenna as an independent element. For example, a single radio dish has $\mathbf{B} \propto \check{\mathbf{W}}$ with $\check{\mathbf{W}}$ given by Eq. (11) modulo polarization complications. To properly compute the polarization response that is encoded in the matrix $\mathbf{B}$, the full three-dimensional structure of the antenna and how it is connected to the two amplifiers must be taken into account, and the presence of nearby conducting objects affects $\mathbf{B}$ as well.

For applications like CMB and 21 cm mapping, where one wishes to measure a cosmological power spectrum, the key aspect of the synthesized beam that matters is how sensitive it is to different angular scales ℓ and their associated spherical harmonic coefficients. This response to different angular is encoded in the spherical harmonic expansion of the synthesized beam $\check{\mathbf{W}}(k_x, k_y)$. If the synthesized beam is rotationally symmetric (or made symmetric by averaging observations with different orientations as Earth rotates), then its spherical harmonic coefficients $\check{\mathbf{W}}_{\ell m}$ vanish except for $m = 0$, and we only need to keep track of the so-called *beam function*, the coefficients $B_\ell \equiv \check{\mathbf{W}}_{\ell 0}$ plotted in Fig. 4. In the flat-sky approximation, this beam function for a rotationally symmetric synthesized beam reduces to the two-dimensional Fourier transform of $\check{\mathbf{W}}(k_x, k_y)$, which is simply the baseline distribution $W(\Delta x)$:

$$B_\ell \approx W(\ell/k). \tag{15}$$

Figure 4 shows B_ℓ for the circular, square and Gaussian aperture cases mentioned above. WMAP and many other CMB experiments have published detailed measurements of their beam functions B_ℓ (*e.g.*, [26]), many of which are fairly well approximated by Gaussians. For interferometers, the beam functions can be significantly more interesting. Since B_ℓ^2 scales simply as the number of baselines at different separations, more complicated synthesized beams involving more than one scale can be designed if desirable.
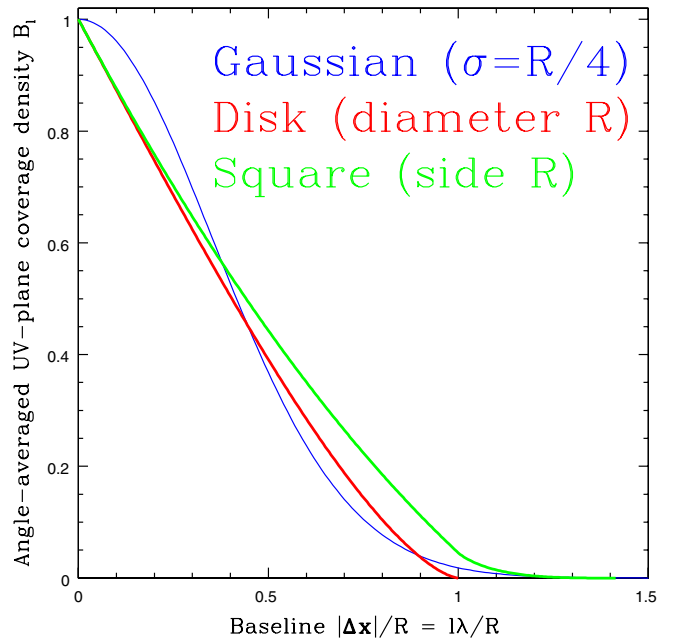

FIG. 4 (color online).   The angle-averaged UV plane sensitivity f_ℓ (the relative number of baselines width different lengths) is compared for telescopes of different shapes.

C. Sensitivity

1. How the noise power spectrum is defined and normalized

The sensitivity of an arbitrary telescope to signals on various angular scales is quantified by its noise power spectrum C_ℓ^{noise}. If the telescope were to make a uniformly observed all-sky map, then C_ℓ^{noise} would be the variance (due to detector noise) with which a spherical harmonic coefficient $a_{\ell m}$ could be measured. For a map that covers merely a small sky patch, the corresponding noise power spectrum is the C_ℓ^{noise} that would result if the whole sky were observed with this same sensitivity. Without loss of generality, we can factor the noise power spectrum as [27,28]

$$C_\ell^{\mathrm{noise}} = C_0^{\mathrm{noise}} B_\ell^{-2}, \tag{16}$$

where B_ℓ is the beam function from the previous section, and C_0^{noise} is an overall normalization constant. To avoid ambiguity in this factorization, we normalize the beam function B_ℓ so that its maximum value equals unity. For a single dish, the maximum is always at $\ell = 0$. This gives the normalization $B_0 = 1$, which given Eq. (15), which means that the synthesized beam $\check{\mathbf{W}}(\mathbf{q})$ integrates to unity and that we can interpret the signal as measuring a weighted average of the true sky map. In all cases, our normalization Most interferometers have $B_0 = 0$ and thus no sensitivity to the mean; in many such cases, B_ℓ is roughly constant on angular scales ℓ^{-1} much larger than

the synthesized beam but much smaller than the primary beam, taking its maximum on these intermediate scales.

This seemingly annoying lack of sensitivity to the mean is a conscious choice and indeed a key advantage of interferometers. The mean sensitivity can optionally be retained by simply including the antenna autocorrelations in the analysis (*i.e.*, not explicitly setting the pixel at the origin of the (u, v) plane equal to zero), but this pixel normally contains a large positive bias due to noise that is difficult to accurately subtract out. In contrast, the noise in all other (u, v) pixels normally has zero mean, because the noise in different antennas is uncorrelated. Since single-dish telescopes cannot exclude this zero mode, they often require other approaches to mitigate this noise bias, such as rapid scanning or d beamswitching.

2. How it depends on experimental details

Consider a telescope with total collecting area A observing for a time τ with a bandwidth $\Delta\nu$ around some frequency $\nu = c/\lambda$. If this telescope performs a single pointing in some direction, then the noise power spectrum for this observed region is [13]

$$C_0^{\text{noise}} = \frac{\lambda^2 \gamma^2 T_{\text{sys}}^2}{A f^{\text{cover}} \tau \Delta\nu}. \tag{17}$$

Here γ is a dimensionless factor of order unity that depends on the convention used to define the telescope system temperature T_{sys}; below we simply adopt the convention where $\gamma = 1$. For a single-dish telescope and for a maximally compact interferometer like the FFTT, $f^{\text{cover}} = 1$. For an interferometer where the antennas are rather uniformly spread out over a larger circular area, f^{cover} is the fraction of this area that they cover; if there are N_a antennas with diameter D in this larger area of diameter D_{max}, we thus have $f^{\text{cover}} = N_a(D/D_{\text{max}})^2$ and total collecting area $A = N_a\pi(D/2)^2$. For a general interferometer the noise power spectrum depends on the distribution of baselines and could be a complicated function of ℓ. We are absorbing all ℓ-dependence into the beam function B_ℓ as per Eq. (16).

If instead of just pointing at a fixed sky patch, the telescope scans the sky (using Earth rotation and/or pointing) to map a solid angle Ω_{map} that exceeds its field-of view Ω, and spends roughly the same amount of time covering all parts of the map, then a given point in the map is observed a fraction $\Omega/\Omega_{\text{map}}$ of the time. The resulting noise power spectrum for the map is then

$$C_0^{\text{noise}} = \frac{4\pi}{\eta} \frac{\lambda^3 f_{\text{sky}} T_{\text{sys}}^2}{f^{\text{cover}} A \Omega c \tau}. \tag{18}$$

Here $f_{\text{sky}} \equiv \Omega_{\text{map}}/4\pi$ is the fraction of the sky covered by the map, and we have introduced the dimensionless parameter $\eta \equiv \Delta\nu/\nu = \Delta\nu c/\lambda$ to denote the relative bandwidth.

D. The 3D noise power spectrum P^{noise}

For 21 cm applications, it is also important to know the three-dimensional noise power spectrum the "data cube" mapped by treating the frequency as the radial direction (the higher the frequency, the larger the redshift and hence the larger the distance to the hydrogen gas responsible for the 21 cm signal). In a comoving volume of space subtending a small angle $\theta \ll 1$ and a small redshift range $\Delta z/z \ll 1$ centered around z_*, we can linearize the relation between the comoving coordinate $\mathbf{r}$ and the observed quantities $(\theta_x, \theta_y, \nu)$ (*e.g.*, [18])

$$\Delta\mathbf{r}_\perp = d_A(z_*)\Delta\Theta \tag{19}$$

$$\Delta r = y(z_*)\Delta\nu. \tag{20}$$

Here $\Delta\Theta \equiv (\theta_x, \theta_y) = (\hat{\mathbf{k}}_x, \hat{\mathbf{k}}_y)$ gives the angular distance away from the center of the field being imaged, and $\Delta\mathbf{r}_\perp$ is the corresponding comoving distance transverse to the line of sight. $d_A(z)$ is the comoving angular diameter distance to redshift z, and

$$y(z) = \frac{\lambda_{21}(1 + z)^2}{H(z)}, \tag{21}$$

where $\lambda_{21} \approx 21$ cm is the rest-frame wavelength of the 21 cm line, and $H(z)$ is the cosmic expansion rate at redshift z. In Appendix B, we show that these two conversion functions can be accurately approximated by

$$d_A(z_*) \approx 14.8 \text{ Gpc} - \frac{16.7 \text{ Gpc}}{(1 + z)^{1/2}}, \tag{22}$$

$$y(z) \approx \frac{\lambda_{21}(1 + z)^{1/2}}{\Omega_m^{1/2}H_0} \approx \frac{18.5 \text{ Mpc}}{1 \text{ MHz}}\left(\frac{1 + z}{10}\right)^{1/2} \tag{23}$$

for the $z \gg 1$ regime most relevant to 21 cm tomography given the flat concordance cosmological parameter values $\Omega_m = 0.25$ and $H_0 = 72$ km s^{-1} Mpc^{-1} [29,30].

If a two-dimensional map is subdivided into pixels of area Ω_{pix} and the noise is uncorrelated with variance σ^2 in these pixels, then

$$C_\ell^{\text{noise}} = \sigma^2 \Omega_{\text{pix}} \tag{24}$$

for angular scales ℓ well above the pixel scale. Analogously, if a three-dimensional map is subdivided into pixels (voxels) of volume V_{pix} and the noise is uncorrelated with variance σ^2 in them, then

$$P^{\text{noise}} = \sigma^2 V_{\text{pix}} \tag{25}$$

on length scales well above the pixel scale. Since the volume of a 3D pixel is $V_{\text{pix}} = (d_A^2\Omega_{\text{pix}}) \times (y\Delta\nu)$, *i.e.*, its area times its depth, combining Eqs. (18), (24), and (25) gives the large-scale noise power spectrum

$$P^{\text{noise}} = \frac{4\pi f_{\text{sky}} \lambda^2 T_{\text{sys}}^2 y d_A^2}{A\Omega f^{\text{cover}} \tau}. \qquad (26)$$

When 2D and 3D power spectra are discussed in the cosmology literature, it is popular to introduce corresponding quantities

$$(\delta T_\ell)^2 \equiv \frac{\ell(\ell+1)}{2\pi} C_\ell, \qquad (27)$$

$$\Delta(k)^2 \equiv \frac{4\pi k^3}{(2\pi)^3} P(k), \qquad (28)$$

which give the variance contribution per logarithmic interval in scale. One typically has $\delta T_\ell \sim \Delta(k)$ when both the angular scale ℓ and the bandwidth $\Delta\nu$ are chosen to match the length scale $\Delta r = 2\pi/k$, i.e., when $\ell = k/d_A$ and $\Delta\nu = 2\pi/ky$. Beware that here (and only here) we use k to denote the wavenumber of cosmic fluctuations, while everywhere else in this paper, we use it to denote the wave vector of electromagnetic radiation.

1. Sensitivity to point sources

It is obviously good to have a small noise power spectrum C_ℓ^{noise} and a large field of view. However, the trade off between these two differs depending on the science goal at hand. Below we mention two cases of common interest.

If one wishes to measure the flux ϕ from an isolated point source, it is easy to show that the attainable accuracy $\Delta\phi$ is

$$\Delta\phi = \sqrt{4\pi C_0^{\text{noise}}} \left[\sum_{\ell=0}^{\infty} (2\ell+1) B_\ell^2 \right]^{-1/2}. \qquad (29)$$

In the approximation of a Gaussian beam $B_\ell = e^{-\theta^2\ell^2/2}$ with rms width $\theta \ll 1$, this simplifies to

$$\Delta\phi \approx \theta \sqrt{4\pi C_0^{\text{noise}}}$$

$$= 4\pi T_{\text{sys}} \theta \sqrt{\frac{\lambda^3 f_{\text{sky}}}{A\Omega f^{\text{cover}} \eta c \tau}} \sim T_{\text{sys}} \sqrt{\frac{\lambda^4 f_{\text{sky}}}{A^2 \Omega \tau \Delta\nu}}. \qquad (30)$$

In the last step, we used the fact that the angular resolution $\theta \sim \lambda/(A/f^{\text{cover}})^{1/2}$. The total information (inverse variance) in the map about the point source flux thus scales as $A^2\Omega\tau\Delta\nu$. That this information is proportional to the field of view Ω, the observing time τ and te bandwidth $\Delta\nu$ is rather obvious. That it scales like the collecting area as A^2 rather than A is because every baseline carries an equal amount of information about the flux ϕ, and the number of baselines scales quadratically with the area. It is independent of f^{cover} because it does not matter how long the baselines are; therefore the result is the same regardless of where the antennas are placed. This last result also provides intuition for the f^{cover}-factor in Eq. (17): since $\theta^2 C_\ell^{\text{noise}}$ is independent of f^{cover} and $\theta \propto \lambda/(A/f^{\text{cover}})^{1/2} \propto \sqrt{f^{\text{cover}}}$, we must have $C_\ell^{\text{noise}} \propto 1/f^{\text{cover}}$. As f^{cover} drops and

the same total amount of information is spread out over an area in the UV plane that is a factor $1/f^{\text{cover}}$ larger, the information in any given ℓ-bin that was previously observed must drop by the same factor, increasing its variance by a factor $1/f^{\text{cover}}$.

2. Power spectrum sensitivity

For CMB and 21 cm applications, one is interested in measuring the power spectrum C_ℓ of the sky signal. The accuracy with which this can be done depends not only on C_ℓ^{noise}, but also on the signal C_ℓ itself (which contributes sample variance) and on the mapped sky fraction f_{sky}. The average power spectrum across a band consisting of $\Delta\ell$ multipoles centered around ℓ can be measured to precision [31,32]

$$\Delta C_\ell \approx \sqrt{\frac{2}{(2\ell+1)\Delta\ell f_{\text{sky}}}} (C_\ell + C_\ell^{\text{noise}}). \qquad (31)$$

Since $C_\ell^{\text{noise}} \propto f_{\text{sky}}$, there is an optimal choice of f_{sky} that minimizes ΔC_ℓ. In cases where $f_{\text{sky}} < 1$ is optimal, this best choice corresponds to $C_\ell^{\text{noise}} \sim C_\ell$, so that sample variance and noise make comparable contributions [31,32]. This means that optimized measurements tend to fall into one of three regimes:

(1) No detection: $\Delta C_\ell \gtrsim C_\ell$ even when f_{sky} is made as small as the telescope permits. Upper limit $\propto C_\ell^{\text{noise}}$.

(2) Improvable detection: $C_\ell^{\text{noise}} \sim C_\ell$, and $\Delta C_\ell \propto f_{\text{sky}}^{-1/2} \propto (C_\ell^{\text{noise}})^{1/2}$.

(3) Cosmic variance limited detection: $C_\ell^{\text{noise}} \ll C_\ell$, and further noise reductions do not help.

The regime depends normally depends on ℓ, since C_ℓ and C_ℓ^{noise} tend to have different shapes. For example, the WMAP measurement of the unpolarized CMB is in regimes 1, 2 and 3 at $\ell \sim 1000$, $\ell \sim 300$ and $\ell \sim 100$, respectively.

E. Field of view Ω

The field of view of a telescope is the solid angle Ω that it can map in a single pointing. For a telescope with a single dish of diameter D and a single receiver/detector pixel in its focal plane (a dish for satellite TV reception, say), the receiver will simply map a sky patch corresponding to the angular resolution $\sim\lambda/D$, giving $\Omega \sim (\lambda/D)^2$. The opposite extreme is to fill the entire focal plane with receivers, as is often done for, e.g., microwave and optical telescopes. In Appendix C, we show that the largest focal plane possible covers an angle of order $(\lambda/D)^{1/3}$, corresponding to $\Omega \sim (\lambda/D)^{2/3}$. This upper bound comes from the fact that the analog Fourier transform performed by telescope optics is only approximate. Many actual multireceiver telescopes fall somewhere between these two extremes. In summary, single-dish telescopes have a field of view somewhere in the range

$$\left(\frac{\lambda}{D}\right)^2 \lesssim \Omega \lesssim \left(\frac{\lambda}{D}\right)^{2/3}. \tag{32}$$

We refer to the two extreme cases in this inequality as the single receiver telescope (SRT) and the maximal focal plane telescope (MFPT), respectively.

Since the performs its Fourier transform with no approximations, it can in principle observe the entire sky above the horizon, corresponding to $\Omega = 2\pi$. However, the useful field of view is only of order half of this, because the image quality degrades near the horizon: viewed from a zenith angle θ, one dimension of the telescope appears foreshortened by a factor $\cos\theta$, causing loss of both angular resolution and collecting area (and thus sensitivity) near the horizon.

F. Cost

Detailed cost estimates for telescopes are notoriously difficult to make, and will not be attempted here. We will instead limit our analysis to the approximate scaling of cost with collecting area, as summarized in Table I, which qualitatively determines which telescopes are cheapest in the different parts of the parameter space of Fig. 2.

For a single-dish telescope, the cost usually grows slightly faster than linearly with area. Specifically, it has been estimated that the cost $\underset{\sim}{\propto} A^{1.35}$ for radio telescopes [33].

For a standard interferometric telescope consisting of N separate dishes, the total cost for the dishes themselves is of course proportional to N. However, the cost for the correlator hardware that computes the correlations between all the $N(N-1)/2$ pairs of dishes scales as N^2, and thus completely dominates the total cost in the large N limit that is the focus of the present paper (already at the modest scale of the MWA experiment, where $N = 512$, the N and N^2 parts of the hardware cost are comparable). For fixed dish size, the total collecting area $A \propto N$ so that the cost $\propto A^2$. For an FFT Telescope, the cost of antennas, ground screen and amplifiers are all proportional to the number of antennas and hence to the area. As described in Sec. II, the computational hardware is also proportional to the area, up to some small logarithmic factors that we to first approximation can ignore.

The above-mentioned approximate scalings are of course only valid over a certain range. All telescopes must have $A \gtrsim \lambda^2$. The cost of single dishes grows more rapidly once their structural integrity becomes an issue—for example, engineering challenges appear to make a single-dish radio telescope with $A = 1 \text{ km}^2$ daunting with current technology,[4] and for an FFTT with diameter $A^{1/2} \gg 10$ km, compensating for Earth's curvature could

---

[4]However, an interesting design for which this might be feasible has been proposed in [34], where an almost flat telescope rests close to the ground and the focal plane is carried by a steerable Helium balloon.

become a major cost.[5] Finally, an all-digital telescope like the FFTT is currently limited to by Moore's law for computer processing speed to frequencies below a few GHz, and analog interferometry has not yet been successfully carried out above optical frequencies.

G. Which telescope is best for what?

Let us now put together the results from the previous subsections to investigate which telescope design is most cost effective for various science goals.

We will use the noise power spectrum C_ℓ^{noise} to quantify sensitivity. We will begin our discussion focusing on only two parameters, the large-scale sensitivity C_0^{noise} and the angular resolution θ, since the parametrization $C_\ell^{\text{noise}} = C_0^{\text{noise}} e^{\theta^2 \ell^2}$ is a reasonable approximation for many of the telescope designs that we have discussed. We then turn to more general noise power spectra when discussing elongated FFTs, general interferometers and the issue of point source subtraction.

1. Complementarity

If we need a telescope with angular resolution θ and large-scale sensitivity C_0^{noise}, then which design will meet out requirements at the lowest cost? The answer is summarized in Fig. 2 for a $\nu = 150$ MHz example. First of all, we see that SDTs, FFTTs and SITs and are highly complementary: the cheapest solution is offered by SDTs for low resolution, FFTTs for high sensitivity $(C_0^{\text{noise}})^{1/2} \lesssim \theta \times 2 \ \mu\text{K}$, and standard interferometers or elongated FFTTs for high resolution $\theta \lesssim (C_0^{\text{noise}})^{1/2}/2 \ \mu\text{K}$.

2. Calculational details

A few comments are in order about how these results were obtained.

For a single SRT, MFPT or FFTT, both the resolution and the sensitivity are determined by their area alone, so as the area is scaled up, they each trace a line through the $(\theta, (C_0^{\text{noise}})^{1/2})$ parameter space of Fig. 2. The cheapest way to attain a better sensitivity at the same resolution is simply to build multiple telescopes of the same area (except for the

---

[5]If Earth were a perfect sphere of radius $R \approx 6400$ km, then a planar telescope of radius r would be a height

$$h \approx \frac{r^2}{2R} \approx 8 \text{ m}\left(\frac{r}{10 \text{ km}}\right)^2 \tag{33}$$

above the ground at its edges. If the telescope is not planar, one cannot use the straightforward FFT analysis method. In practice, this might only be a problem if *both* of the dimensions of the FFTT are $\gg 10$ km: as long as the telescope can be kept flat in the *narrowest* dimension, it will have no intrinsic (Gaussian) curvature even if the telescope has Earth's circular shape in its wide direction. An interesting question for future work is whether some algorithm incorporating an FFT along the long axis can be found that provides and exact and efficient recovery of the sky map for this generalized case.

FFTT, where cost$\underset{\sim}{\propto}A$, so that one might as well build a single larger telescope instead and get extra resolution for free). Since $C_0^{\mathrm{noise}} \propto 1/N\Omega$, where N is the number of telescopes whose images are averaged together, the sensitivity of an FFTT with a given resolution can be matched by building $N = \Omega_{\mathrm{FFTT}}/\Omega$ telescopes, where $N \sim A^{1/3}/\lambda^{2/3}$ for the MFPT and $N \sim A/\lambda^2$ for the SRT. The cost relative to an FFTT of the same resolution and sensitivity thus grows as $A^{0.65}$ for MFPTs and as $A^{1.33}$ for SRT's. The area below which single dish telescopes are cheaper depends strongly on wavelength; for the illustrative purposes of Fig. 2, we have taken this to be $(10m)^2$ at 150 GHz based on crude hardware cost estimates for the GMRT [22] and MWA [1] telescopes.

For regions to the right of the FFTT line in Fig. 2, one has the option of either building a square (or circular) FFTT with unnecessarily high sensitivity to attain the required resolution, or to build an elongated FFTT or a conventional interferometer—we return to this below, and argue that the latter is generally cheaper.

3. How the results depend on frequency and survey details

Although the Fig. 2 is for a specific example, these qualitative results hold more generally. Survey duration, bandwidth, system temperature and sky coverage all merely rescale the numbers on the vertical axis, leaving the figure otherwise unchanged. As one alters the observing wavelength, the resolution and sensitivity remains the same if one alters the other scales accordingly: $A \propto \lambda^2$, $c\tau \propto \lambda$, except that T_{sys} grows rapidly towards very low frequencies as the brightness temperature of synchrotron radiation exceeds the instrument temperature. The cost depends strongly and nonlinearly on frequency. As discussed in Sec. III F, both the FFTT and digital SITs are currently feasible only below about 1 GHz, and analog interferometry has not yet been successfully carried out above optical frequencies.

4. The advantage of an FFTT over a single dish telescope

The results above show that the FFTT can be thought of as simply a cheap single-dish telescope with a 180° field of view. Compared to single-dish telescope, the FFTT has two important advantages:
  (1) It is cheaper in the limit of a large collecting area, with the cost scaling roughly like A rather than $A^{1.35}$ or more.
  (2) It has better power spectrum sensitivity even for fixed area A, because of a field of view that is larger by a factor between $(D/\lambda)^{2/3}$ and $(D/\lambda)^2$.
An important disadvantage of the FFTT is that it currently only works below a about 1 GHz. Even if it were not for this limitation, since the computational cost of interferometry depends on the number of resolution elements $N \sim \Omega/\theta^2$, which grows fast toward higher frequencies (as $\nu^{4/3}$

for the MFPT and as ν^2 for the FFTT), single-dish telescopes become comparatively more advantageous at higher frequencies. However, as Moore's law marches on, the critical frequency where an FFTT loses out to an SDT should grow exponentially over time.

5. The advantage of an FFT Telescope over a traditional correlating interferometer

The results above also show that the FFTT can be thought of as a cheap maximally compact interferometer with a full-sky primary beam. To convert a state-of-the-art interferometers such as MWA [1], LOFAR [19], PAPER [20], or 21CMA [21] into an FFTT, one would need to do three things:
  (1) Move all antenna tiles together so that they nearly touch.
  (2) Get rid of any beamformer that "points" tiles towards a specific sky direction by adding relative phases to its component antennas, and treat each antenna as independent instead, thus allowing the array to image all sky directions simultaneously.
  (3) Move the antennas onto a rectangular grid to cut the correlator cost from N^2 to $N\log_2 N$.
This highlights both advantages and disadvantages of the FFTT compared to traditional interferometers. There are three important advantages:
  (1) It is cheaper in the limit of a large collecting area, with the cost scaling roughly like A rather than A^2.
  (2) It has better power spectrum sensitivity even for fixed area A, because of a field of view that is larger than for an interferometer whose primary beam is not full sky (because its array elements are either single-dish radio telescopes or antenna tiles that are pointed with beamformers).
  (3) The synthesized beam is as clean and compact as for a SDT, corresponding to something like a simple Airy pattern. This has advantages for multifrequency point source subtraction as discussed below, and also for high fidelity mapmaking.
The most obvious drawback of a square or circular FFTT is that the angular resolution is much poorer than what a traditional interferometer can deliver. This makes it unsuitable for many traditional radio astronomy applications. We discuss below how this drawback can be partly mitigated by a rectangular rather than square design.

A second drawback is the lack of flexibility in antenna positioning. Whereas traditional interferometry allows one to place the antennae wherever it is convenient given the existing terrain, the construction of a large FFTT requires bulldozing.

6. The advantage of a 2D FFTT over a 1D FFTT exploiting Earth rotation

There are two fundamentally different approaches to fully sampling a disk around the origin of the Fourier plane

(usually referred to as the UV plane in the radio astronomy terminology): build a two-dimensional array (like a square FFTT) whose baselines cover this disk, or build a more sparse array that fills the disk gradually, after adding together observations made at multiple times, when Earth rotation has rotated the available baselines. Equation (18) shows that, given a fixed number of antennas and hence a fixed collecting area, the former option gives lower C_0^{noise} and hence more accurate power spectrum measurements as long as the angular resolution is sufficient. The reason is that the factor f^{cover} in the denominator equals unity for the former case, and is otherwise smaller. For a rectangular FFTT of dimensions $D_{min} \times D_{max}$, $B_\ell^2 f^{cover}$ depends on the angular scale ℓ and it is easy to show that

$$B_\ell^2 f^{cover} \sim \begin{cases} 1 & \text{for } \ell \lesssim \frac{D_{min}}{\lambda} \\ \frac{D_{min}}{\lambda \ell} & \text{for } \frac{D_{min}}{\lambda} \lesssim \ell \lesssim \frac{D_{max}}{\lambda} . \\ 0 & \text{for } \ell \gtrsim \frac{D_{min}}{\lambda} \end{cases} \quad (34)$$

In essence, making the telescope more oblong simply dilutes the same total amount of information out over a broader range of ℓ-space, thus giving poorer sensitivity on the angular scales originally probed.

What telescope configuration is desirable depends on the science goal at hand. It has been argued [18,35] that for doing cosmology with 21 cm tomography in the near term, it is best to make the telescope as compact as possible, *i.e.*, to build a square or circular telescope. The basic origin of this conclusion is the result "a rolling stone gathers no moss" mentioned in Sec. III D 2: for power spectrum measurement, it is optimal to focus the efforts to make the signal-to-noise of order unity. The first generation of experiments have much lower signal-to-noise than this, and thus benefit from focusing on large angular scales and measuring them as accurately as possible rather than measuring a larger range of angular scales with even poorer sensitivity. Of course, none of these 1st generation telescopes were funded for 21cm cosmology alone, and their ability to perform other science hinges on having better angular resolution, explaining why they were designed with less compact configurations. Better angular resolution can also aid point source removal.

For other applications where high angular resolution required, an oblong telescope is preferable. An interesting proposal of this type is the higher-frequency mapping proposed Pittsburgh Cylinder telescope [9,10], which is one-dimensional. Instead rather omnidirectional antennas, it takes advantage of its one-dimensional nature by having a long cylindrical mirror, which increases the collecting area at higher frequencies. This is advantageous because its goal is to map 21 cm emission at the lower redshifts (higher frequencies $\gtrsim 200$ MHz) corresponding to the period after cosmic reionization, to detect neutral hydrogen in galaxies and use this to measure the baryon acoustic oscillation scale as a function of redshift. If one wishes to perform rotation synthesis with an oblong or 1D FFTT, it will

probably be advantageous to build multiple telescopes rotated relative to one another (say in an L-shaped layout, or like spokes on a wheel), to reduce the amount of integration time needed to fill the UV plane. Cross-correlating the N antennas between the telescopes would incur a prohibitive N^2 computational cost, so such a design with T separate telescopes would probably need to discard all but of a fraction $1/T$ if the total information, corresponding to the intratelescope baselines.

Another array layout giving higher resolution is to build an array whose elements consist of FFTTs placed far apart. After performing a spatial FFT of their individual outputs, these can then be multiplied and inverse-transformed pairwise, and the resulting block coverage of the UV plane can be filled in by Earth rotation. As long as the number of separate FFTTs is modest, the extra numerical cost for this may be acceptable.

Above we discussed the trade off between different shapes for fixed collecting area. If one instead replaces a $D \times D$ two-dimensional FFTT by a one-dimensional FFTT of length D using rotation synthesis, then Eq. (18) shows that one loses sensitivity in two separate ways: at the angular scale $\ell \sim D/\lambda$ where the power spectrum error bar ΔC_ℓ from Eq. (31) is the smallest, one loses one factor of D/λ from the drop in f^{cover}, and a second factor of D/λ from the drop in collecting area A. Another way of seeing this is to note that the available information scales as the number of baselines, which scales as the square of the number of antennas and hence as A^2. This quadratic scaling can also be seen in Eq. (30): the total amount of information $(\Delta\phi)^{-2}$ scales as $A^2\Omega\tau\Delta\nu$, so whereas field of view, observing time and bandwidth help only linearly, area helps quadratically. This is because we can correlate electromagnetic radiation at different points in the telescope, but not at different times, at different frequencies or from different points in the sky. The common statement that the information gathered scales as the etendu $A\Omega$ is thus true only at fixed ℓ; when all angular scales are counted, the scaling becomes $A^2\Omega$.

If in the quest of more sensitivity, one keeps lengthening an oblong or one-dimensional FFT to increase the collecting area, one eventually hits a limit: the curvature of Earth's surface makes a flat $D \gg 10$ km exceedingly costly, requiring instead telescope curving along Earth's surface and the alternative analysis framework mentioned above in Sec. III F. If one desires maximally straightforward data analysis, one thus wants to grow the telescope in the other dimension to make it less oblong, as discussed in Sec. III F. This means that if one needs $\gg 10^4$ antennas for adequate 21 cm cosmology sensitivity, one is forced to build a 2D rather than 1D telescope. For comparison, even the currently funded MWA experiment with its $512 \times 4^2 = 8192$ antennas is close to this number.

One final science application where 2D is required is the study of transient phenomena that vary on a time scale

much shorter than a day, invalidating the static sky approximation that underlies rotation synthesis. This was the key motivation behind the aforementioned Waseda telescope [4–6].

IV. APPLICATION TO 21 CM TOMOGRAPHY

In the previous section we discussed the pros and cons of the FFTT telescope, and found that its main strength is for mapping below about 1 GHz when extreme sensitivity is required. This suggests that the emerging field of 21 cm tomography is an ideal first science application of the FFTT: it requires sky mapping in the sub-GHz frequency range, and the sensitivity requirements, especially to improve cosmic microwave background constraints on cosmological parameters, are far beyond what has been achieved in the past [18,36–38].

A. 21 cm tomography science

It is becoming increasingly clear that 21 cm tomography has great scientific potential for both astrophysics [12–15,35] and fundamental physics [18,36–39]. The basic idea is to produce a three-dimensional map of the matter distribution throughout our Universe through precision measurements of the redshifted 21 cm hydrogen line. For astrophysics, much of the excitement centers around probing the cosmic dark ages and the subsequent epoch of reionization caused by the first stars. Here we will focus mainly on fundamental physics, as this arguably involves both the most extreme sensitivity requirements and the greatest potential for funding extremely sensitive measurements.

1. Three physics frontiers

Future measurements of the redshifted 21 cm hydrogen line have the potential to probe hitherto unexplored regions of parameter space, pushing three separate frontiers: time, scale, and sensitivity. Figure 5 shows a scaled sketch of our observable Universe, our Hubble patch. It serves to show the regions that can be mapped with various cosmological probes, and illustrates that the vast majority of our observable universe is still not mapped. We are located at the center of the diagram. Galaxies [from the Sloan Digital Sky Survey (SDSS) in the plot] map the distribution of matter in a three-dimensional region at low redshifts. Other popular probes like gravitational lensing, supernovae Ia, galaxy clusters and the Lyman α forest are currently also limited to the small volume fraction corresponding to redshifts $\lesssim 3$ or less, and in many cases much less. The CMB can be used to infer the distribution of matter in a thin shell at the so-called "surface of last scattering", whose thickness corresponds to the width of the black circle at $z \sim 1100$ and thus covers only a tiny fraction of the total volume. The region available for observation with the 21 cm line of hydrogen is shown in light blue/grey.
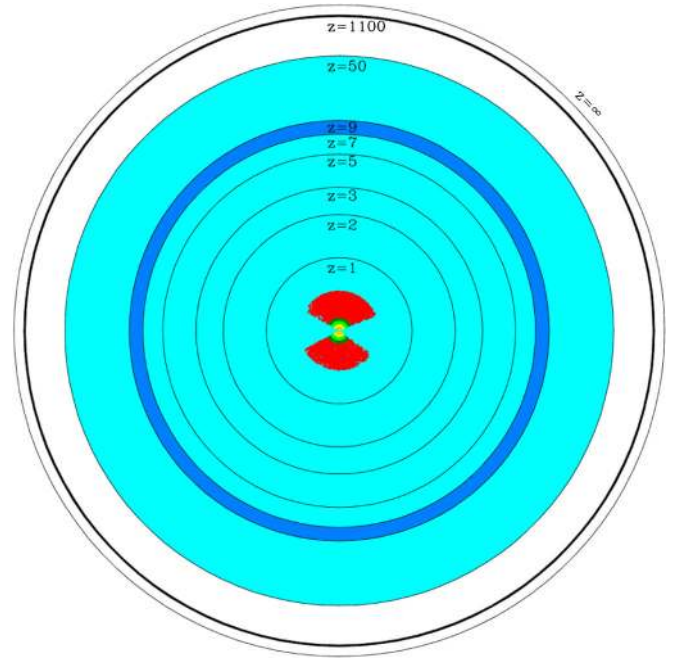

FIG. 5 (color online). 21 cm tomography can potentially map most of our observable universe (light blue/gray), whereas the CMB probes mainly a thin shell at $z \approx 1100$ and current large-scale structure maps (here exemplified by the Sloan Digital Sky Survey and its luminous red galaxies) map only small volumes near the center. Half of the comoving volume lies at $z > 29$ (Appendix B). This paper focuses on the convenient $7 \lesssim z \lesssim 9$ region (dark blue/grey).

Clearly the 21 cm line of hydrogen has the potential of allowing us to map the largest fraction of our observable universe and thus obtain the largest amount of cosmological information.

At the high redshift end ($z \gtrsim 30$) the 21 cm signal is relatively simple to model as perturbations are still linear and "gastrophysics" related to stars and quasars is expected to be unimportant. At intermediate times, during the epoch of reionization (EOR) around redshift $z \sim 8$, the signal is strongly affected by the first generation of sources of radiation that heat the gas and ionize hydrogen. Modeling this era requires understanding a wide range of astrophysical processes. At low redshifts, after the epoch of reionization, the 21 cm line can be used to trace neutral gas in galaxies and map the large-scale distribution of those galaxies.

2. The time frontier

Figure 5 illustrates that observations of the 21 cm line from the EOR and higher redshifts would map the distribution of hydrogen at times where we currently have no other observational probe, pushing the redshift frontier. Measurements of the 21 cm signal as a function of redshift will constrain the expansion history of the universe, the growth rate of perturbations and the thermal history of the gas during an epoch that has yet to be probed.

 (i) Tests of the standard model predictions for our cosmic thermal history $T(z)$, expansion history $H(z)$ (which can be measured independently using both expansion and the angular diameter distances), and linear clustering growth.
 (ii) Constraints on modified gravity from the above-mentioned measurements of $H(z)$ and clustering growth.
(iii) Constraints on decay or annihilation of dark matter particles, or any other long-lived relic, from the above-mentioned measurement of our thermal history [40–42]. Here 21 cm is so sensitive that even the expected annihilation of ''vanilla'' neutralino WIMP cold dark matter may be detectable [42].
(iv) Constraints on evaporating primordial black holes from the thermal history measurement [43].
 (v) Constraints on time-variation of fundamental physical constants such as the fine structure constant [44].

3. The scale frontier

These observations can potentially push the ''scale frontier'', significantly extending the range of scales that are accessible to do cosmology. This is illustrated in Fig. 6, where the scales probed by different techniques are compared to what is available in 21 cm. Neutral hydrogen is a

good probe of the small scales for two separate but related reasons. First, one can potentially make observations at higher redshifts, where more of the scales of interest are in the linear regime and thus can be better modeled. Second, at early times in the history of our Universe, hydrogen is still very cold and thus its distribution is expected to trace that of the dark matter up to very small scales, the so-called Jeans scale, where pressure forces in the gas can compete with gravity [45].

 (i) Precision tests of inflation, since smaller scales provide a longer lever arm for constraining the spectral index and its running (illustrated in Fig. 6) for the power spectrum of inflationary seed fluctuations [18]
 (ii) Precision tests of inflation by constraining small-scale non-Gaussianity [46].
(iii) Precision constraints on noncold dark matter from probing galactic scales while they were still linear.

4. The sensitivity frontier

This combination of a large available volume with the presence of fluctuations on small scales that can be used to constrain cosmology implies that the amount of information that at least in principle can be obtained with the 21 cm is extremely large. This can be illustrated by calculating the number of Fourier modes available to do cosmology that can be measured with this technique. This number can be compared with the number of modes measured to date with various other techniques such as galaxy surveys, the CMB, etc. In Fig. 7, we show the number of modes measured by past surveys and some planned probes including 21 cm experiments.[6] The figure illustrates a trend akin to Moore's law: exponential progress as a function of year. It is striking that the improvement of the 1 km^2 FFTT over WMAP is comparable to that of WMAP over COBE. Moreover, the ultimate number of modes available to be observed with 21 cm tomography is dramatically larger still, upward of 10^{16}, so although many practical issues will most certainly limit what can be achieved in the near future, the ultimate potential is vast.

The FFTT sensitivity improvement translates into better measurement accuracy for many of the usual cosmological parameters. It has been shown that even the limited redshift range $7 \lesssim z \lesssim 9$ (dark shading in Fig. 5) has the potential to greatly improve on cosmic microwave background con-
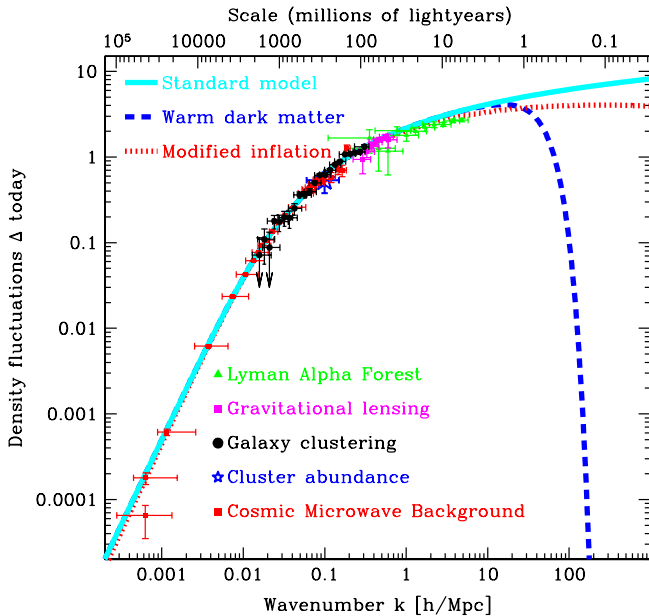

FIG. 6 (color online).   21 cm tomography can push the scale frontier far beyond that of current measurements of cosmic clustering, potentially all the way down to the Jeans scale at the right edge of the figure. This allows distinguishing between a host of alternative inflation and dark matter models that are consistent with all current data, for example, a warm dark matter with mass 14 keV (dashed curve) or greater and inflation with a running spectral index more extreme than $dn_s/d\ln k = -0.03$ (dotted).

---

[6]Although the number of modes gives an estimate of the statistical power of a survey, constraints on specific parameters will depend on how strongly each of the power spectra varies as a function of the parameter of interest. Furthermore, when considering probes such as the Lyman-α forest that probes modes in the nonlinear regime, our numbers based on the Gaussian formula overestimates the constraining power. In constructing this figure, only modes in the linear regime $k < 0.1$ h Mpc^{-1} were included for galaxy surveys. These are the range of modes that are typically used for doing cosmology. If the galaxy formation process becomes sufficiently well understood it may become feasible to increase the number of useful modes.
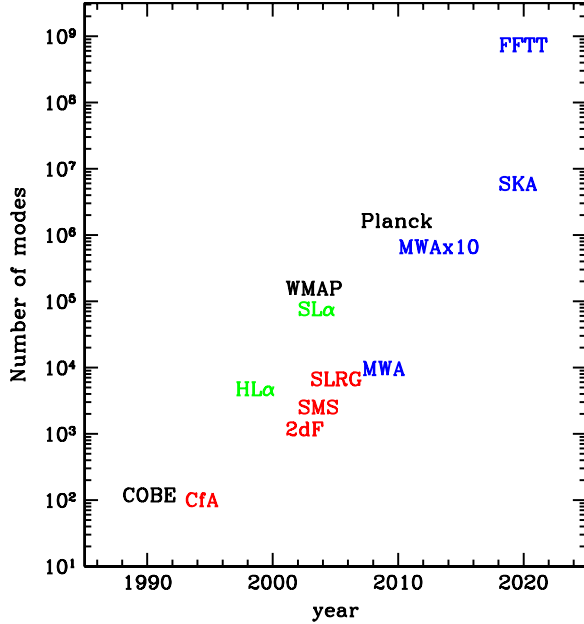
FIG. 7 (color online). Number of modes measured with different cosmological probes. We show illustrative examples of galaxy redshift surveys [CfA, PsCz, 2dF, SDSS main sample (SMS), SDSS Luminous red galaxies (SLRG)], CMB experiments (COBE, WMAP and Planck), Lyman-α forest measurements (using high resolution spectra (HLα) and SDSS spectra (SLα)) and 21 cm experiments (MWA, an extension of MWA with 10 times the collecting area, the square kilometer array (SKA) and a 1 km^2 FFTT). The number of modes is calculated from the constraints these experiments can place on the overall amplitude of the power spectrum ($\delta P/P$) and then using the formula for Gaussian random fields $\delta P/P = \sqrt{2/N_{modes}}$.

straints from WMAP and Planck: it could improve the sensitivity to spatial curvature and neutrino masses by up to 2 orders of magnitude, to $\Delta\Omega_k \approx 0.0002$ and $\Delta m_\nu \approx 0.007$ eV, and give a 4σ detection of the spectral index running predicted by the simplest inflation models [18]. Indeed, it may even be possible to measure three individual neutrino masses from the scale and time dependence of clustering [18,47].

Measuring the 21 cm power spectrum and using it to constrain physics and astrophysics does not require pushing the noise level down to the signal level, since the noise can be averaged down by combining many Fourier modes probing the same range of scales. This is analogous to how the COBE satellite produced the first measurement of the CMB power spectrum even though individual pixels in its sky maps were dominated by noise rather than signal [48]. Further boosting the sensitivity to allow imaging (with signal-to-noise per pixel exceeding unity) allows a number of improvements:

(i) Improving quantification, modeling and understanding of foregrounds and systematic errors
(ii) Pushing down residual foregrounds with better cleaning (like in the CMB field, the residual fore-

ground level after cleaning is likely to be comparable to the noise level)
(iii) Enabling power spectrum and non-Gaussianity estimation after masking out ionized bubbles, thus greatly reducing the hard-to-model gastrophysics contribution
(iv) Constraining small-scale properties of dark matter by using 21 cm maps as backgrounds for gravitational lensing experiments that could detect the presence of dark substructure in lower redshifts halos [49–51]
(v) Pushing to higher redshift where the physics is simpler

B. The cost of sensitivity

There is thus little doubt that sensitivity improvements can be put to good use. Equation (26) implies that the high-redshift frontier, in particular, has an almost insatiable appetite for sensitivity: since $\lambda \propto (1 + z)$, $y \propto (1 + z)^{1/2}$, d_A depends only weakly on z, and the diffuse synchrotron foreground that dominates T_{sys} at low frequencies scales roughly as $\nu^{-2.6} \propto (1 + z)^{2.6}$ in the cleanest parts of the sky for $50 \lesssim \nu \lesssim 200$ MHz [52], Eq. (26) gives a sensitivity

$$\delta T \underset{\sim}{\propto} [k^3 P^{\text{noise}}]^{1/2} \propto \frac{k^{3/2}(1 + z)^{3.85} f_{\text{sky}}^{1/2}}{(A\Omega f^{\text{cover}}\tau)^{1/2}}. \quad (35)$$

if the observing time and field of view is held fixed (like for the FFTT). Pushing from $z = 9$ to $z = 20$ with the same sensitivity thus requires increasing the collecting area by a factor around 300. This would keep the signal-to-noise level roughly the same if the 21 cm fluctuation amplitude is comparable and peaks at similar angular scales at the two redshifts, as suggested by the calculations of [17]. Equation (35) shows that imaging smaller scales is expensive too, with an order of magnitude smaller scales (multiplying k by 10) requiring a thousandfold increase in collecting area.

Figures 8 and 9 illustrate the rough cost of attaining the sensitivity levels required for various physics milestones mentioned above. Our cost estimates are very crude, and making more accurate ones would go beyond the scope of the present paper, but the qualitative scalings seen in the figures should nonetheless give a good indication of how the different telescope designs complement each other.[7] For our estimates, we have assumed $\tau = 4000$ hours of

---

[7]For interferometer arrays, we use the following hardware cost estimate loosely based on the MWA hardware budget [1]: $1M \times (A/8000 \text{ m}^2)\beta + 1M \times (A/8000 \text{ m}^2)^\gamma$, where $(\beta, \gamma) = (1.2, 1)$ for the FFTT and $(\beta, \gamma) = (1, 2)$ for a conventional interferometer. The first term corresponds to per-antenna costs (with β reflecting the extra construction cost related to land leveling *etc.*), and the second term corresponds to the computational cost. For a single dish, we assume a hardware cost $0.4M \times [A/(1600 \text{ m}^2)]^{1.35}$ based on Wilkinson's scaling [33] from the GMRT budget [22].
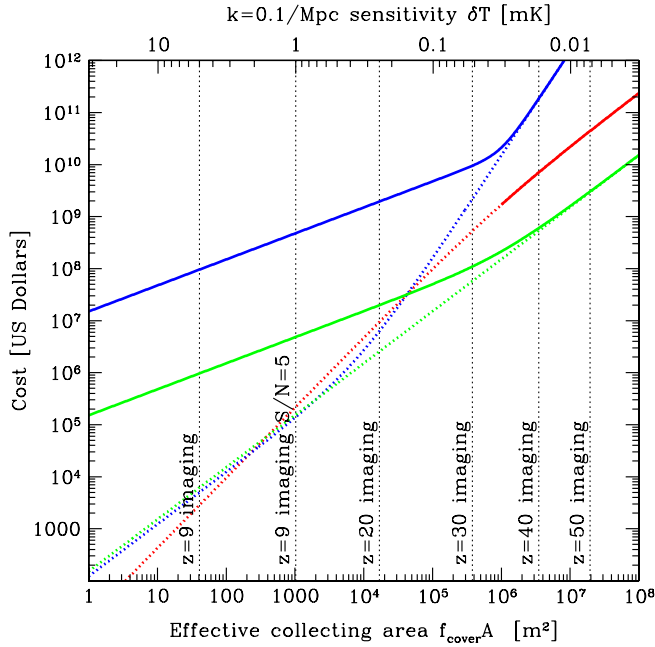
FIG. 8 (color online). The rough hardware cost in 2008 US Dollars of attaining various sensitivities at the $k = 0.1/\text{Mpc}$ scale with the FFTT telescope (green curves), a maximally compact regular interferometer (blue curves) and a single-dish telescope (red curve) always pointing towards the same patch of sky ($4\pi f_{\text{sky}}\Omega$). The dashed curves have angular resolution poorer than $\ell = 500$ at redshift $z = 9$; for the SIT and FFTT, this resolution can be achieved by making the telescope array oblong a higher cost (solid curves), since the area must be increased to compensate for the drop in f^{cover}. Note that cost is a function of A only so to plot is as a function of sensitivity $A f^{\text{cover}}$ a design dependent relation between A and f^{cover} is required.

observation with a system temperature $T_{\text{sys}} = 200 \text{ K} \times [(1 + z)/10]^{2.6}$. We assume the cosmic signal to be of order $\delta T = 5$ mK at the redshifts of interest [17]. Our baseline estimates are an observing frequency of 142 MHz, corresponding to 21 cm emission at redshift $z = 9$.

Figure 8 is for the case when all we care about is sensitivity, not how large a sky area is mapped with this sensitivity. We thus keep the telescope pointing at the same sky patch and get $4\pi f_{\text{sky}}\Omega$, so Eq. (35) gives a sensitivity $\delta T \propto (k^{3/2}(1 + z)^{3.85})/(A f^{\text{cover}})^{1/2}$. For a fixed spatial scale k and redshift z, the sensitivity thus depends only on the collecting effective area $f^{\text{cover}}A$ plotted on the horizontal axis. The solid curves in the figure all have maximally compact configurations with $f^{\text{cover}} = 1$, corresponding to angular resolution $\ell \sim A^{1/2}/\lambda$. The lines are dotted where this resolution $\ell < 500$ for the baseline wavelength $\lambda = 2.1$ m. If we insist on the higher resolution $\ell = 500$, we can achieve this goal by making the FFT or SIT oblong or otherwise sparse, with $f^{\text{cover}} \sim A/(\lambda\ell)^2 \propto A$, so in this regime, $(A f^{\text{cover}}) \propto A^2$ and hence $A \propto (A f^{\text{cover}})^{1/2}$, and this area in turn determines the cost—this is why the
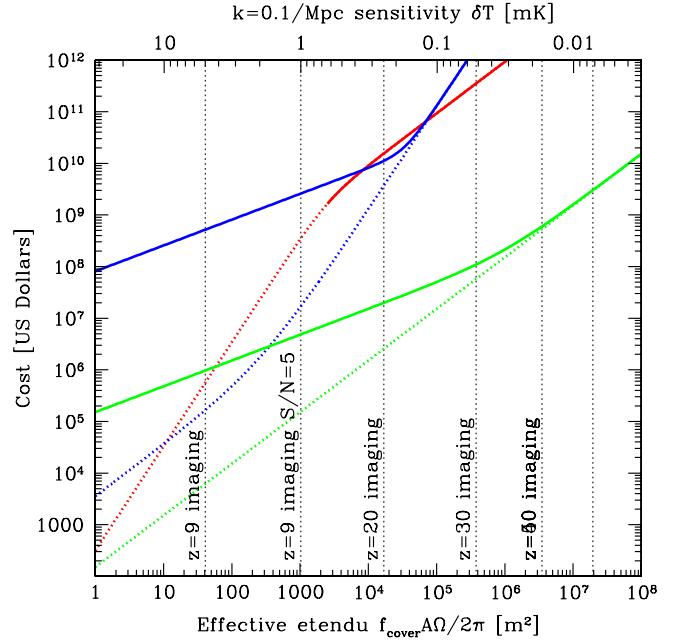

FIG. 9 (color online). Same as previous figure, but when half the sky is mapped ($f_{\text{sky}} = 2\pi$). The standard interferometric telescope (SIT) and single dish maximal focal plane telescope (MFPT) take an additional cost his here, needing to further increase the area to compensate for the drop in field of view Ω with A.

solid curves in Fig. 8 lie above the corresponding dotted ones.

Figure 9 is for the case when we want a map of a fixed area (WMAP-style), in this case covering half the sky ($f_{\text{sky}} = 0.5$), so Eq. (35) gives a sensitivity $\delta T \propto (k^{3/2}(1 + z)^{3.85})/(A\Omega f^{\text{cover}})^{1/2}$. For a fixed spatial scale and redshift, the sensitivity thus depends only on the collecting effective etendu $f^{\text{cover}}A\Omega$ plotted on the horizontal axis. Since Ω drops with area for both the SRT and MFPT, in order to boost sensitivity, these telescopes now need an extra area boost to make up for the drop in Ω. Although an MFPT has cost $\propto A^{1.35}$, it also has $\Omega \propto A^{-1/3}$, so that $A\Omega \propto A^{2/3}$ and the cost $\propto (A\Omega)^{1.35 \times 3/2} \approx (A\Omega)^2$. Once A is large enough to give sufficient resolution ($A \gtrsim \lambda^2\ell^2$) it becomes smarter to simply build multiple telescopes, giving cost $\propto A$.

For comparison, we have indicated some sensitivity benchmarks as vertical lines. Equation (35) shows that $\delta T \propto k^{3/2}(1 + z)^{3.85}$; this redshift scaling is illustrated by these vertical lines. Additional sensitivity can also be put to good use for probing smaller scales, since an order of magnitude change in k corresponds to 3 orders of magnitude on the horizontal axis.

C. 21 cm foregrounds

Aside from its extreme sensitivity requirements, another unique feature of 21 cm cosmology is the magnitude of its

foreground problem: it involves mapping a faint and diffuse cosmic signal that needs to be separated from foreground contamination that is many orders of magnitude brighter [14,16,52], requiring extreme sensitivity and beam control. Fortunately, the foreground emission (mainly synchrotron radiation) has a rather smooth frequency spectrum, while the cosmological signal varies rapidly with frequency (corresponding to variations in physical conditions along the line of sight). Early work on 21 cm foregrounds [53–55] has indicated that this can be exploited to clean out the foregrounds down to an acceptable level, effectively by high-pass filtering the data cube in the frequency direction.

However, these papers have generally not treated the additional complication that the synthesized beam $\check{W}(\theta_x, \theta_y)$ is frequency dependent, dilating like λ, which means that when raw sky maps at two different frequencies cannot be readily compared. For a single-dish telescope or an FFTT, the synthesized beam is compact and simple enough that this complication can be modeled and remedied exactly (say by convolving maps at all frequencies to have the same resolution before foreground cleaning), but for a standard interferometer, complicated low-level ''frizz'' extending far from the central parts of the synthesized beam appears to make this unfeasible at the present time. Recent work [56–58] has indicated that this is a serious problem: whereas the foreground emission from our own galaxy is smooth enough that these off-beam contributions average down to low levels, emission from other galaxies appears as point sources to which the telescope response varies rapidly with frequency because of the beam dilation effect. The ability to mitigate this problem is still subject to significant uncertainty [10], and may therefore limit the ultimate potential of 21 cm cosmology with a conventional interferometer. The ability to deal with foreground contamination is thus another valuable advantage of the FFT Telescope.

V. CONCLUSIONS

We have presented a detailed analysis of an all-digital telescope design where mirrors are replaced by fast Fourier transforms, showing how it complements conventional telescope designs. The main advantages over a single dish telescope are cost and orders of magnitude larger field-of-view, translating into dramatically better sensitivity for large-area surveys. The key advantages over traditional interferometers are cost (the correlator computational cost for an N-element array scales as $N\log_2 N$ rather than N^2) and a compact synthesized beam. These traits make the FFT Telescope ideal for applications where the angular resolution requirements are modest while those on sensitivity are extreme. We have argued that the emerging field of 21 cm tomography could provide an ideal first application of a very large FFT Telescope, since it could provide massive sensitivity im-

provements per dollar as well as mitigate the off-beam point source foreground problem with its clean beam.

A. Outstanding challenges

There are a number of interesting challenges and design questions that would need to be addressed before building a massive FFT Telescope for 21 cm cosmology. For example:

(1) To what extent can the massive redundancy of an FFT Telescope (where the same baseline is typically measured by $\sim N_a^{1/2}$ independent antenna pairs) be exploited to calibrate the antennas against one another in a computationally feasible way?

(2) To what extent, if any, are more distant antennas outside the FFTT needed to resolve bright point sources and calibrate the FFTT antennas?

(3) After calibration, how do gain fluctuations in the individual array elements affect the noise properties of the recovered sky map?

(4) How do variations in primary beam $\mathbf{B}(\hat{\mathbf{k}})$ from Eq. (1) from between individual antennas affect the properties of the recovered sky map?

(5) How many layers of dummy antennas are needed around the active instrumented part of the array to ensure that the beam patterns of all utilized antennas are sufficiently identical?

(6) What antenna design is optimal for a particular FFT Telescope science application, maximizing gain in the relevant frequency range? The limit of an infinite square grid of antennas on an infinite ground screen is quite different from the limit of a single isolated antenna, and modeling mutual coupling effects becomes crucial when computing the primary beam $\mathbf{B}(\hat{\mathbf{k}})$ from Eq. (1)

(7) What unforeseen challenges does the FFT Telescope entail, and how can they be overcome?

(8) Can performing the first stages of the spatial FFT by analog means (say connecting adjacent 2×2 or 4×4 antenna blocks with Butler matrices [2]) lower the effective system temperature in parts of the sky with overall lower levels of synchrotron emission?

Answering these questions will require a combination of theoretical and experimental work. The authors are currently designing a small FFTT prototype with a group of radio astronomy colleagues to address these questions and to identify unforeseen obstacles.

B. Outlook

Looking further ahead, we would like to encourage theorists to think big and look into what additional physics may be learned from the sort of massive sensitivity gains that an FFTT could offer, as this can in turn increase the motivation for hard work on experimental challenges like those listed above.

Perhaps in a distant future, almost all telescopes will be FFT Telescopes, simultaneously observing light of all wavelengths from all directions. In the more immediate future, as Moore's law enables FFTT's with higher bandwidth, cosmic microwave background polarization may be an interesting application besides 21 cm cosmology. By using an analog frequency mixer to extract of order a GHz of bandwidth in the CMB frequency range (around say 30 GHz or 100 GHz), it would be possible to obtain a much greater instantaneous sky coverage than current CMB experiments provide, and this gain in Ω could outweigh the disadvantage of lower bandwidth $\Delta\nu$ in Eq. (18) to provide overall better sensitivity. The fact that extremely high spectral resolution would be available essentially for free may also help ground-based measurements, allowing exploitation of the fact that some atmospheric lines are rather narrow.

ACKNOWLEDGMENTS

APPENDIX A: POLARIZATION ISSUES

The Stokes matrix $\mathbf{S}$ defined by Eq. (6) is related to the usual Stokes parameters I, Q, U, V by

$$\mathbf{S} = \frac{1}{2}\begin{pmatrix} I+Q & U-iV \\ U+iV & I-Q \end{pmatrix} = \frac{1}{2}\mathbf{v}\cdot\sigma. \quad (A1)$$

In the dot product,

$$\mathbf{v} \equiv \begin{pmatrix} I & Q & U & V \end{pmatrix} \quad (A2)$$

and

$$\sigma \equiv \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \right\}, \quad (A3)$$

contains the four Pauli matrices. As usual, I denotes the total intensity, Q and U quantify the linear polarization and V the circular polarization (which normally vanishes for astrophysical sources). It is easy to invert Eq. (A1) to solve for the Stokes parameters:

$$\mathbf{v} = \mathrm{tr}\{\sigma \cdot \mathbf{S}\}. \quad (A4)$$

An annoying but harmless nuisance when dealing with large-area polarization maps is the well-known fact that "you cannot comb a sphere", *i.e.*, that there is no global choice of reference vector to define the Jones vector and the Stokes parameters (Q, U) all across the sky. In practice, it never matters until at the very last analysis step, since one can collect the data $\mathbf{d}_i$ and reconstruct both $\hat{\mathbf{S}}_B$ and $\mathbf{S}_B$ without worrying about this issue. To compute $\mathbf{B}$ and solve for the Stokes parameters, any convention for defining the Stokes parameters will suffice, even one involving separate schemes for a number of partially overlapping sky patches; it is easy to see that the choice of convention has no effect on the accuracy or numerical stability of the inversion method.

APPENDIX B: COSMIC GEOMETRY

In this Appendix, we derive Eqs. (22) and (23). For a flat universe (which is an excellent approximation for ours [29,30]), the comoving angular diameter distance is given by [59]

$$d_A(z) = \int_0^z \frac{c\,dz'}{H(z')}, \quad (B1)$$

where

$$H(z) = H_0\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}, \quad (B2)$$

where $\Omega_\Lambda = 1 - \Omega_m$. The second term in the square root becomes negligible for $z \gg (\Omega_\Lambda/1 - \Omega_m)^{1/3} - 1 \approx 0.4$ for $\Omega_m = 0.25$ [30], which gives Eq. (23). The dark energy density is completely negligible at the high redshift regime relevant to 21 cm cosmology also in most models where this density evolves with time. For such high redshifts, we can therefore approximate Eq. (B1) as follows:

$$d_A(z) = \int_0^\infty \frac{c\,dz'}{H(z')} - \int_z^\infty \frac{cH_0^{-1}dz'}{\Omega_m^{1/2}(1+z')^{3/2}}$$
$$\approx H_0^{-1}c\left[3.56 - \frac{4}{(1+z)^{1/2}}\right] \quad (B3)$$

for $\Omega_m = 0.25$, which gives Eq. (22). The accuracy of Eq. (B3) better than 1% for $z > 2.2$, *i.e.*, better that with which the relevant cosmological parameters have currently been measured.

Equation (B3) shows that, surveying our observable universe as illustrated in Fig. 5, we reach half the comoving distance at $z \approx (4/(3.56 \times 0.5))^2 - 1 \approx 4$ and half the comoving volume at $z \approx \{4/[3.56 \times (1 - 0.5^{1/3})]\}^2 - 1 \approx 29$.

APPENDIX C: FIELD-OF-VIEW ESTIMATES

In this appendix, we derive the restriction on the field of view for a single dish telescope. Consider a parabolic mirror of height z given by:

$$z = \frac{x^2 + y^2}{R} \tag{C1}$$

where x and y are the coordinates in the plane of the ground and R determines the radius of curvature. The mirror has a diameter D such that

$$x^2 + y^2 < \frac{D^2}{4}. \tag{C2}$$

We consider radiation initially traveling with wave vector $\hat{\mathbf{k}} = k(\sin\theta, 0, \cos\theta)$ with $k = 2\pi/\lambda$. We will calculate the phase of the radiation that scatters at the location $(x, y, z) = (\rho\cos\phi, \rho\sin\phi, \rho^2/R)$ on the surface of the mirror and then arrives at a detector located at (x_f, y_f, z_f). For simplicity we will consider a point in the mirror with $y = 0$ so that $\phi = 0$ and then also set $y_f = 0$. After some simple algebra one obtains the following expression for the phase ψ:

$$\psi = k\left[\sqrt{(x_f - \rho)^2 + \left(z_f - \frac{\rho^2}{R}\right)^2} - \frac{\rho(\rho\cos\theta + R\sin\theta)}{R} \right]. \tag{C3}$$

Because of the parabolic shape chosen for the mirror, the phase of radiation coming with normal incidence ($\theta = 0$) comes to a perfect focus at $x_f = 0$, $z_f = R/4$. By perfect focus we mean that the phase $\psi(x_f, z_f, \rho)$ is independent of ρ for $x_f = 0$, $z_f = z/4$, $\theta = 0$ For radiation incident at an angle, there will be no point in space where one can locate the detector so that the radiation reflected everywhere in the mirror will be in phase. We will find the field of view of the telescope by demanding that the phase difference between radiation incident in different parts of the telescope be less than a radian at the location of the detector.

To obtain a formula, we expand ψ in a Taylor series as a function of ρ. By choosing x_f and z_f, we can make the terms linear and quadratic in ρ vanish, but the cubic term will in general be nonzero, except for normal incidence. For a given telescope diameter we will then find the field of view by demanding that the cubic contribution to the phase be smaller than a radian. The Taylor series of ψ is:

$$k^{-1}\psi \approx \sqrt{x_f^2 + z_f^2} - \left(\frac{x_f}{\sqrt{x_f^2 + z_f^2}} + \sin\theta \right)\rho$$
$$+ \frac{\left(\frac{z_f((R-2z_f)z_f - 2x_f^2)}{(x_f^2 + z_f^2)^{3/2}} - 2\cos\theta \right)\rho^2}{2R}$$
$$+ \frac{x_f z_f((R - 2z_f)z_f - 2x_f^2)\rho^3}{2R(x_f^2 + z_f^2)^{5/2}} + \cdots \tag{C4}$$

By choosing

$$x_f = -z_f\tan\theta \approx -z_f\theta, \tag{C5}$$

we can eliminate the term linear in ρ, and by choosing

$$z_f = \frac{R}{8}(\cos2\theta + 1) \approx \frac{R}{4} \tag{C6}$$

the quadratic one. Thus we get

$$\psi \approx k\left(\frac{1}{4}R\cos\theta - \frac{4\rho^3\sin\theta.}{R^2} + \cdots \right). \tag{C7}$$

For small values of θ, demanding that ψ changes by less than a radian as we move from the center to the edge of the telescope, and using $k = 2\pi/\lambda$, we obtain,

$$\theta < \frac{R^2\lambda}{D^3\pi} = \frac{R^2}{D^2\pi} \times \frac{\lambda}{D}. \tag{C8}$$

Thus by increasing the radius of curvature R, one can increase the field of view. In fact, $(R/D)^2$ basically gives the number of resolution elements in each linear dimension in the focal plane.

The upper bound on the size on the curvature radius comes from demanding that the focal plane not cover the entire telescope. Using Eqs. (C5) and (C6) and demanding that the size of the focal plane be smaller than $D/2$ (a very conservative assumption), we get another constraint on the field of view:

$$\theta < \frac{2D}{R}. \tag{C9}$$

While the size of the field of view increases with R in (C8), it decreases with R in (C9) and thus the largest field of view is obtained when both constraints are equal, and corresponds to

$$R = \left(\frac{2\pi D}{\lambda} \right)^{1/3} D; \qquad \theta < \left(\frac{4\lambda}{\pi D} \right)^{1/3}. \tag{C10}$$

The inequality that we have been derived can be pushed somewhat with clever multimirror designs (for example, the optical large synoptic telescope uses three mirrors [60]). In contrast, radio telescopes typically use only one mirror. In this case, the value of R required to attain the maximal field of view that we have derived is a factor $(2\pi D/\lambda)^{1/3}$ larger than D and can thus get very large for sufficiently small wavelengths. Mechanical constraints can make building such a radio telescope impractical as the focal plane would be very far away from the telescope, making the upper bound $\theta \lesssim (\lambda/D)^{1/3}$ that we have derived for the field of view a rather conservative one.

[1] http://www.haystack.mit.edu/ast/arrays/mwa/.

[2] J. Butler and R. Lowe, Electronic Design **9**, 170 (1961).

[3] J. May, F. Reyes, J. Aparici, M. Bitran, H. Alvarez, and F. Olmos, Astron. Astrophys. **140**, 377 (1984)

[4] J. Nakajima *et al.*, Publ. Astron. Soc. Jpn. **44L**, 35 (1993).

[5] J. Nakajima *et al.*, Publ. Astron. Soc. Jpn. **45**, 477 (1993).

[6] N. Tanaka *et al.*, Publ. Astron. Soc. Jpn. **52**, 447 (2000)

[7] U. Pen, New Astron. Rev. **9**, 417 (2004).

[8] H. Takeuchi *et al.*, Publ. Astron. Soc. Jpn. **57**, 815 (2005).

[9] J. B. Peterson, K. Bandura, and U. L. Pen, arXiv:astro-ph/0606104.

[10] T. Chang, U. Pen, J. B. Peterson, and P. McDonald, Phys. Rev. Lett. **100**, 091303 (2008).

[11] After presenting this work at a May 2008 conference, we were told that the Square Kilometer Array collaboration is considering similar ideas for the central part of their proposed telescope (Jeffrey B. Peterson, private communication).

[12] R. Barkana and A. Loeb, Phys. Rep. **349**, 125 (2001).

[13] M. Zaldarriaga, S. Furlanetto, and L. Hernquist, Astrophys. J. **608**, 622 (2004).

[14] S. Furlanetto, S. P. Oh, and F. Briggs, Phys. Rep. **433**, 181 (2006).

[15] A. Loeb, Sci. Am. **295**, 46 (2006).

[16] M. G. Santos, A. Amblard, J. Pritchard, H. Trac, R. Cen, and A. Cooray, Astrophys. J. **689**, 1 (2008).

[17] J. R. Pritchard and A. Loeb, Phys. Rev. D **78**, 103511 (2008).

[18] Y. Mao, M. Tegmark, M. McQuinn, O. Zahn, and M. Zaldarriaga, Phys. Rev. D **78**, 023529 (2008).

[19] http://www.lofar.org/.

[20] D. Backer (private communication).

[21] http://web.phys.cmu.edu/~past/, formerly known as *PaST*.

[22] http://gmrt.ncra.tifr.res.in/.

[23] U. Pen, T. Chang, J. B. Peterson, J. Roy, Y. Gupta, and K. Bandura, in *The GMRT Search for Reionization*, AIP Conf. Proc. No. 1035 (AIP, New York, 2008).

[24] http://www.skatelescope.org/.

[25] A. R. Thompson and J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy* (Wiley, New York, 2001), 2nd ed..

[26] R. S. Hill, Astrophys. J. Suppl. Ser. **180**, 246 (2009).

[27] L. Knox, Phys. Rev. D **52**, 4307 (1995).

[28] M. Tegmark and G. Efstathiou, Mon. Not. R. Astron. Soc. **281**, 1297 (1996).

[29] M. Tegmark *et al.*, Phys. Rev. D **74**, 123507 (2006).

[30] J. Dunkley, Astrophys. J. Suppl. Ser. **180**, 306 (2009).

[31] D. Scott, M. Srednicki, and M. White, Astrophys. J. **421**, L5 (1994).

[32] M. Tegmark, Phys. Rev. D **56**, 4514 (1997).

[33] P. N. Wilkinson, in *Radio Interferometry: Theory, Techniques, and Applications: Proceedings of the 131st IAU Colloquium, Astronomical Society of the Pacific,*

1990.

[34] T. H. Legg, Astron. Astrophys. Suppl. Ser. **130**, 369 (1998).

[35] A. Lidz, O. Zahn, M. McQuinn, M. Zaldarriaga, and L. Hernquist, Astrophys. J. **680**, 962 (2008).

[36] J. D. Bowman, M. F. Morales, and J. N. Hewitt, Astrophys. J. **661**, 1 (2007).

[37] M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist, and S. R. Furlanetto, Astrophys. J. **653**, 815 (2006).

[38] A. Loeb and S. Wyithe, Phys. Rev. Lett. **100**, 161301 (2008).

[39] R. Barkana and A. Loeb, Astrophys. J. Lett. **624**, L65 (2005).

[40] S. Furlanetto, P. O. Oh, and E. Pierpaoli, Phys. Rev. D **74**, 103502 (2006).

[41] M. Valdes, A. Ferrara, M. Mapelli, and E. Ripamonti, Mon. Not. R. Astron. Soc. **377**, 245 (2007).

[42] Z. Myers and A. Musser, Mon. Not. R. Astron. Soc. **384**, 727 (2008).

[43] K. J. Mack, J. P. Ostriker, and M. Ricotti, Astrophys. J. **680**, 829 (2008).

[44] R. Khatri and B. D. Wandelt, Phys. Rev. Lett. **98**, 111301 (2007).

[45] S. Naoz and R. Barkana, Mon. Not. R. Astron. Soc. **362**, 1047 (2005)

[46] A. Cooray, C. Li, and A. Melchiorri, Phys. Rev. D **77**, 103506 (2008).

[47] J. R. Pritchard and E. Pierpaoli, Phys. Rev. D **78**, 065009 (2008).

[48] G. F. Smoot *et al.*, Astrophys. J. **396**, L1 (1992).

[49] O. Zahn and M. Zaldarriaga, Astrophys. J. **653**, 922 (2006).

[50] S. Hilbert, R. B. Metcalf, and S. D. M. White, Mon. Not. R. Astron. Soc. **382**, 1494 (2007).

[51] R. B. Metcalf and S. D. M. White, Mon. Not. R. Astron. Soc. **381**, 447 (2007).

[52] de Oliveira-Costa, M. Tegmark, B. M. Gaensler, J. Jonas, T. L. Landecker, and P. Reich, Mon. Not. R. Astron. Soc. **388**, 247 (2008).

[53] X. Wang, M. Tegmark, M. Santos, and L. Knox, Astrophys. J. **650**, 529 (2006).

[54] M. F. Morales, J. D. Bowman, and J. N. Hewitt, Astrophys. J. **648**, 767 (2006).

[55] V. Jelic *et al.*, Mon. Not. R. Astron. Soc. **389**, 1319 (2008).

[56] J. D. Bowman, Ph.D. thesis, MIT, 2007.

[57] J. D. Bowman, M. F. Morales, and J. N. Hewitt, arXiv:0807.3956.

[58] A. Liu, M. Tegmark, and M. Zaldarriaga, arXiv:0807.3952.

[59] E. W. Kolb and M. S. Turner, *The Early Universe* (Addison Wesley, New York, 1990).

[60] Z. Ivezic *et al.*, arXiv:0805.2366.