

Precision calibration of radio interferometers using redundant baselines

Adrian Liu,^{1*} Max Tegmark,¹ Scott Morrison,¹ Andrew Lutomirski¹
and Matias Zaldarriaga²

¹*Department of Physics and MIT Kavli Institute, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA*

²*Institute for Advanced Study, School of Natural Sciences, Einstein Drive, Princeton, NJ 08540, USA*

Accepted 2010 June 9. Received 2010 June 8; in original form 2010 February 5

ABSTRACT

Growing interest in 21-cm tomography has led to the design and construction of broad-band radio interferometers with low noise, moderate angular resolution, high spectral resolution and wide fields of view. With characteristics somewhat different from traditional radio instruments, these interferometers may require new calibration techniques in order to reach their design sensitivities. Self-calibration or redundant calibration techniques that allow an instrument to be calibrated off complicated sky emission structures are ideal. In particular, the large number of redundant baselines possessed by these new instruments makes redundant calibration an especially attractive option. In this paper, we explore the errors and biases in existing redundant calibration schemes through simulations, and show how statistical biases can be eliminated. We also develop a general calibration formalism that includes both redundant baseline methods and basic point source calibration methods as special cases, and show how slight deviations from perfect redundancy and coplanarity can be taken into account.

Key words: instrumentation: interferometers – methods: data analysis – cosmology: observations – radio continuum: general.

1 INTRODUCTION

Technological advances of recent years have enabled the design and construction of radio telescopes with broad-band frequency coverage, low instrumental noise, moderate angular resolution, high spectral resolution and wide fields of view. These instruments have generated considerable scientific interest due to their unique potential to do 21-cm cosmology, mapping the high-redshift Universe using redshifted radio emission from neutral hydrogen. Not only would this provide the only direct measurements of the cosmological dark ages and the epoch of re-ionization (Madau, Meiksin & Rees 1997; Tozzi et al. 2000a,b; Iliev et al. 2002; Furlanetto, Sokasian & Hernquist 2004a; Furlanetto, Zaldarriaga & Hernquist 2004b; Loeb & Zaldarriaga 2004; Barkana & Loeb 2005; Furlanetto et al. 2009b), but it also has the potential to overtake the cosmic microwave background as the most powerful cosmological probe, measuring cosmological parameters related to inflation, dark matter and neutrinos to unprecedented accuracy (McQuinn et al. 2006; Santos & Cooray 2006; Bowman, Morales & Hewitt 2007; Mao et al. 2008; Furlanetto et al. 2009a) and improving provide dark energy constraints after re-ionization through the detection of baryon acoustic oscillations (Chang et al. 2008; Wyithe, Loeb & Geil 2008; Morales & Wyithe 2009).

From a practical standpoint, however, none of the aforementioned applications will be feasible without reliable algorithms for instrumental calibration. Compared to traditional radio astronomy experiments, instruments designed for 21-cm tomography will require particularly precise calibration because of the immense foreground subtraction challenges involved. Epoch of Re-ionization experiments, for example, will be attempting to extract cosmological signals from beneath foregrounds that are of order 10^4 times brighter (de Oliveira-Costa et al. 2008). Such procedures, where one must subtract strong foregrounds from a bright sky to uncover weak cosmological signals, are particularly susceptible to calibration errors. A prime example of this is the effect that a miscalibration has on the subtraction of bright, resolved point sources. Datta, Bhatnagar & Carilli (2009), for instance, argued that bright point sources must be localized to extremely high positional accuracy in order for their subtraction to be successful, and Liu, Tegmark & Zaldarriaga (2009) showed that failure to do so could comprise one's subsequent ability to subtract any *unresolved* point sources, making the detection of the 21-cm cosmological signal difficult. This places stringent requirements on calibration accuracy.

While standard algorithms are available for calibrating radio interferometers, the unprecedented challenges faced by 21-cm tomography and the advent of a new class of large radio arrays [such as Precision Array to Probe the Epoch of Reionization (PAPER; Parsons et al. 2010),

*E-mail: acliu@mit.edu

Low Frequency Array (LOFAR; Garrett 2009), Murchison Widefield Array (MWA; Lonsdale et al. 2009), Cylinder Radio Telescope (CRT; Peterson, Bandura & Pen 2006), the omniscopes (Tegmark & Zaldarriaga 2009b) and the Square Kilometer Array (SKA; Carilli & Rawlings 2004)] mean that it is timely to consider the relative merits of various algorithms. For example, several authors (Bhatnagar et al. 2008; Morales & Matejek 2009) have pointed out that these new high precision radio arrays require close attention to be paid to heterogeneities in the primary beams of different antenna elements. In this paper, we focus on the technique of *redundant calibration*, which takes advantage of the fact that these new telescopes generically possess a large number of redundant or near-redundant baselines. This property of exact or near-redundancy is not an accident, and arises from the science requirements of 21-cm tomography,¹ where the demanding need for high sensitivity on large spatial scales calls for arrays with a large number of short baselines (Morales 2005; Lidz et al. 2008). Since these arrays typically also have very wide fields-of-view and in many cases operate at frequencies where all-sky survey data is scarce, the fewer assumptions one needs to make about the structure of the sky signal, the better. This is an advantage that redundant calibration has over other calibration methods – no a priori assumptions need to be made about calibrator sources.

The use of redundant baselines for calibration goes far back. For example, Noordam & de Bruyn (1982) used this approach to produce high dynamic range images of 3C 84, and similar techniques have been used in solar imaging (Ishiguro 1974; Ramesh, Subramanian & Sastry 1999). The methods used are described in detail in Wieringa (1992), Yang (1988) and Pearson & Readhead (1984), and closely resemble the logarithmic implementation we review in Section 2.3. In this paper, we build on this previous work and extend it in several ways:

- (i) we provide a detailed examination of the errors and biases involved in existing redundant baseline methods;
- (ii) we introduce methods to mitigate (or, in the case of biases, eliminate) these problems;
- (iii) we show how slight deviations from perfect redundancy and coplanarity can be incorporated into the analysis,

with the view that it may be necessary to correct for all these effects in order to achieve the precision calibration needed for precision cosmology.

The rest of the paper is organized as follows. After defining our notation, we describe the simplest possible calibration algorithm (i.e. point source calibration) in Section 2.1. In Section 2.3 we review the redundant baseline calibration proposed by Wieringa (1992). Readers familiar with radio interferometer calibration techniques are advised to skip to Section 2.4, where we examine Wieringa’s scheme more closely by deriving the optimal weightings for its fitting procedure. In Section 2.5 we focus on simulations of redundant calibration for monolithic square arrays, which show that existing redundant calibration algorithms have certain undesirable error properties. These properties suggest a linearized variation that we propose in Section 2.7. In Section 3 we consider redundant calibration algorithms and point source calibration schemes as extreme cases of a generalized calibration formalism, which allows us to treat the case of near-redundant baseline calibration. We summarize our conclusions in Section 4.

2 CALIBRATION USING REDUNDANT BASELINES

We begin by defining some basic notation. Suppose an antenna i measures a signal s_i at a given instant.² This signal can be written in terms of a complex gain factor g_i , the antenna’s instrumental noise contribution n_i and the true sky signal x_i that would be measured in the limit of perfect gain and no noise:

$$s_i = g_i x_i + n_i. \quad (1)$$

Each baseline measures the correlation between the two signals from the two participating antennas:

$$c_{ij} \equiv \langle s_i^* s_j \rangle \quad (2a)$$

$$= g_i^* g_j \langle x_i^* x_j \rangle + \langle n_i^* n_j \rangle + g_i^* \langle x_i^* n_j \rangle + g_j \langle n_i^* x_j \rangle \quad (2b)$$

$$= g_i^* g_j \langle x_i^* x_j \rangle + n_{ij}^{\text{res}} \equiv g_i^* g_j y_{i-j} + n_{ij}^{\text{res}}, \quad (2c)$$

where we have denoted the true correlation $\langle x_i^* x_j \rangle$ by y_{i-j} , and the angled brackets $\langle \dots \rangle$ denote time averages. The index $i - j$ is not to be taken literally but is intended to signify the fact that any given correlation should depend only on the *relative* positions of the two antennas. As is usual in radio interferometry, we have assumed that noise contributions from different antennas are uncorrelated, and that instrumental noise is uncorrelated with the sky signal. This means that $\langle n_i^* x_j \rangle$ reduces to $\langle n_i^* \rangle \langle x_j \rangle$, and since $\langle n_i \rangle \rightarrow 0$ in the limit of long integration times, we can assume that the last three terms in equation (2b) average to some residual noise level n^{res} . The size of n^{res} will of course depend on the details of the instrument, but in general will be on the order of $T_{\text{sys}}/\sqrt{\tau \Delta\nu}$, where T_{sys} is some fiducial system temperature, τ is the total integration time and $\Delta\nu$ is the bandwidth. For the purposes of this paper, the details of the noise term do not matter, and for notational

¹ Note that we are *not* advocating the redesign of arrays to achieve redundancy for the sake of redundant baseline calibration. The design of an array should be driven by its science requirements, and the ability to use redundant calibration should simply be viewed as an option to be explored if there happens to be redundancy.

² More precisely, s_i refers to frequency domain data at a given instant. In other words, it is the result of dividing the time-ordered data into several blocks, each of which is then Fourier transformed in the time direction to give a signal s_i that is both a function of frequency and time (albeit at a coarser time resolution than the original time-ordered data).

convenience we will drop the superscript ‘res’ from now on. The key qualitative results of this paper (such as the fact that many existing redundant baseline calibration schemes are biased) do not depend on the noise level.

It should be noted that the equation for c_{ij} presented above does not represent the most general form possible for the measured correlations. So-called non-closing errors, for example, can result in gain factors that do not factor neatly on an antenna-by-antenna basis. In this paper we will forgo a discussion of these errors, with the view that a good number of these contributions can be mitigated (although not eliminated completely) by good hardware design. Alternatively, our results may be interpreted as best-case scenario *upper limits* on calibration quality.

Since the gain factors in equation (2c) are in general complex, we can parametrize them by an amplitude gain η and a phase φ by letting $g_i \equiv e^{\eta_i + i\varphi_i}$. Our equation then becomes

$$c_{ij} = \exp[(\eta_i + \eta_j) + i(\varphi_j - \varphi_i)]y_{i-j} + n_{ij}. \quad (3)$$

The goal of any radio interferometer is to extract the true correlations y_{i-j} (from which one can construct a sky map using Fourier transforms) from the measured correlations c_{ij} . Formally, however, this is an unsolvable problem for the generic case, as one can see from equation (3) – with N antennas, one has $N(N-1)/2$ measured correlations (c_{ij} s), which are insufficient to solve for the $N(N-1)/2$ true correlations (y_{i-j} s) in addition to the $2N$ unknown η s and φ s. (The n_{ij} noise terms are also unknown quantities, but are generally not treated as ones that need to be solved for, but as noise to be minimized in a least-squares fit.) In abstract terms, all calibration schemes can be thought of as methods for reducing the number of unknowns on the right-hand side of equation (3) so that the system of equations becomes overdetermined and the η s and φ s can be solved (or fit) for.

2.1 Basic point source calibration

The most straightforward way to calibrate a radio telescope is to image a single bright point source. A point source has constant amplitude in uv space, and thus y_{i-j} is a complex number whose complex amplitude is independent of baseline. Moreover, the phase of y_{i-j} must also be zero, since by definition we are only ‘looking’ at the bright point source if our radio array is phased so that the phase centre lies on the source. Thus, $y_{i-j} = S$, and equation (3) reduces to

$$c_{ij} = \exp[(\eta_i + \eta_j) + i(\varphi_j - \varphi_i)]S + n_{ij}, \quad (4)$$

which is an overdetermined set of equations for all but the smallest arrays, since we have $N(N-1)/2$ complex inputs on the left-hand side but only $2N$ real numbers (η s and φ s) to solve for on the right-hand side. The system is therefore solvable, up to two intrinsic degeneracies: the overall gain $\sum_i \eta_i$ is indeterminate, as is the overall phase $\sum_i \varphi_i$. These degeneracies, however, can be easily dealt with by having some knowledge of our instrument and our calibration source. For instance, the overall gain can be computed if the calibrator source is a catalogued source with known flux density.

It is important to note that the point source calibration method we have presented in this section represents only the simplest, most straightforward way in which one could calibrate a radio telescope. There exist far more sophisticated methods,³ such as the many self-calibration schemes that are capable of calibrating a radio telescope off most reasonable sky signals. In practice, one almost never needs to use basic point source calibration, for it is just as easy to use a self-calibration algorithm. The scheme described here should therefore be considered as no more than a ‘toy model’ for calibration.

2.2 Redundant baseline calibration

The main drawback of the approach described in the previous section is that it requires the existence of an isolated bright point source. This requirement becomes increasingly difficult to fulfil as cosmological science drivers push radio telescopes towards regimes of higher sensitivity and wider fields of view. While self-calibration methods have been shown to produce high dynamic range maps over wide fields of view and do *not* require the existence of isolated bright point sources, they are reliant on having a reasonable a priori model of the sky. Thus, if possible it may be preferable to opt for redundant calibration, which is completely model independent. With the emergence of telescopes with high levels of redundancy [such as the CRT – see Peterson et al. (2006) – or any omniscopes⁴ like those described in Tegmark & Zaldarriaga (2009a,2009b)] redundant baseline calibration becomes a competitive possibility.

Mathematically, if an array has a large number of redundant baselines, then one can reduce the number of unknowns on the right-hand side of equation (3) by demanding that the true visibilities y_{i-j} for a set of identical baselines be the same. Since the number of *measured* visibilities c_{ij} s stays the same, our system of equations becomes overdetermined provided there are a sufficient number of redundant baselines, and it becomes possible to fit for the η and φ parameters.

As an example, consider a one-dimensional array of five radio antennas where the antennas are equally spaced. In this case, one has four unique baselines (of lengths 1, 2, 3 and 4), and 10 measured correlations. Thus, one must fit for 18 real numbers (four complex numbers

³ See Cornwell & Fomalont (1999) or Rau et al. (2009), for example, for nice reviews.

⁴ We define an omniscopes as any instrument where antenna elements are arranged on a regular grid, and full digitized data are collected on a per element basis without any beam forming.

from the true visibilities of the four unique baselines, plus five η s and five φ s from 20 numbers (10 complex measured correlations):

$$\begin{aligned}
c_{1,2} &= \exp[(\eta_1 + \eta_2) + i(\varphi_2 - \varphi_1)] y_1 + n_{1,2}, \\
c_{2,3} &= \exp[(\eta_2 + \eta_3) + i(\varphi_3 - \varphi_2)] y_1 + n_{2,3}, \\
&\vdots \text{ (all four baselines of length 1)} \\
c_{1,3} &= \exp[(\eta_1 + \eta_3) + i(\varphi_3 - \varphi_1)] y_2 + n_{1,3}, \\
c_{2,4} &= \exp[(\eta_2 + \eta_4) + i(\varphi_4 - \varphi_2)] y_2 + n_{2,4}, \\
&\vdots \text{ (all three baselines of length 2)} \\
c_{1,4} &= \exp[(\eta_1 + \eta_4) + i(\varphi_4 - \varphi_1)] y_3 + n_{1,4}, \\
&\vdots \text{ (both baselines of length 3)} \\
c_{1,5} &= \exp[(\eta_1 + \eta_5) + i(\varphi_5 - \varphi_1)] y_4 + n_{1,5}.
\end{aligned} \tag{5}$$

As is usual in radio interferometry, we have omitted equations corresponding to baselines of length zero, i.e. autocorrelations of the data from a single antenna. This is because autocorrelation measurements carry with them correlated noise terms, and thus are likely to be much noisier than other correlated measurements.

In the following sections we will present two methods for explicitly solving equation (5). We begin in Section 2.3 with a logarithmic method, which is very similar to one described by Wieringa (1992). This method is the simplest way to implement a redundant baseline calibration scheme, but suffers from several problems which we detail in Sections 2.5 and 2.6. In Section 2.7, we show how these problems can be solved with an alternative method based on linearization.

2.3 The logarithmic implementation

We proceed by rewriting our equations in the following way:

$$c_{ij} = g_i^* g_j y_{i-j} \left(1 + \frac{n_{ij}}{g_i^* g_j y_{i-j}} \right). \tag{6}$$

Taking the logarithm gives

$$\ln c_{ij} = \eta_i + \eta_j + i(\varphi_j - \varphi_i) + \ln y_{i-j} + \underbrace{\ln \left(1 + \frac{n_{ij}}{g_i^* g_j y_{i-j}} \right)}_{\equiv w_{ij}}, \tag{7}$$

and requiring that the real and imaginary parts be separately equal gives linear equations of the form

$$\ln |c_{i,j}| = \eta_i + \eta_j + \ln |y_{i-j}| + \text{Re } w_{ij}, \tag{8a}$$

$$\arg |c_{i,j}| = \varphi_j - \varphi_i + \arg |y_{i-j}| + \text{Im } w_{ij}. \tag{8b}$$

Note that the amplitude and phase calibrations have now decoupled,⁵ and can thus be written as two separate matrix systems. The amplitude calibration, for example, takes the form

$$\begin{pmatrix} \ln |c_{1,2}| \\ \ln |c_{2,3}| \\ \ln |c_{3,4}| \\ \ln |c_{4,5}| \\ \ln |c_{1,3}| \\ \ln |c_{2,4}| \\ \ln |c_{3,5}| \\ \vdots \\ \ln |c_{2,5}| \\ \ln |c_{1,5}| \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ \vdots & & & & & & & & \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\equiv \mathbf{A}} \underbrace{\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \ln |y_1| \\ \ln |y_2| \\ \ln |y_3| \\ \ln |y_4| \end{pmatrix}}_{\equiv \mathbf{x}} + \underbrace{\begin{pmatrix} s_{1,2} \\ s_{2,3} \\ s_{3,4} \\ s_{4,5} \\ s_{1,3} \\ s_{2,4} \\ s_{3,5} \\ \vdots \\ s_{2,5} \\ s_{1,5} \end{pmatrix}}_{\equiv \text{Re } \mathbf{w}} \tag{9}$$

⁵ From a rigorous standpoint, it is not strictly true that the amplitude and phase calibrations of an interferometer decouple. Indeed, this is evident even in our simple model, where $\text{Re } w_{ij}$ and $\text{Im } w_{ij}$ both contain factors of the complex gain $g \equiv e^{\eta+i\varphi}$, which means that equations (8a) and (8b) are in principle coupled, even if only weakly so. However, one must remember that the noise parameters are treated simply as given numbers which are *not* solved for when solving these calibration equations. This means that in finding the best-fitting solution to equations (8a) and (8b), the η and φ calibrations *can be performed separately*. It is only in this sense that the systems are ‘decoupled’.

for the one-dimensional, five-element array considered in Section 2.2. The vector \mathbf{d} stores the measured correlations, the matrix \mathbf{A} is completely determined by the array configuration and the \mathbf{x} is what we hope to solve for.

As it currently stands, however, this set of linear equations has no unique solution even in the absence of noise (i.e. even when $\mathbf{w} = 0$). This can be seen from the fact that the vector $\mathbf{x}_{\text{null}} = (1, 1, 1, 1, -2, -2, -2, -2)$ (written here as a row vector for convenience) lies in the null space of \mathbf{A} (i.e. $\mathbf{A}\mathbf{x}_{\text{null}} = 0$), so for any solution \mathbf{x}_0 , one can form a new solution $\mathbf{x}_0 + \mathbf{x}_{\text{null}}$. The new solution corresponds to adding a constant to all η s, which – since $g \equiv \exp(\eta + i\varphi)$ – is equivalent to multiplying all gains by a constant factor and simultaneously dividing all the sky signals by the same factor. Physically, this is an expression of the fact that internal calibration schemes (i.e. schemes like this one where the calibration is done by exploiting internal mathematical consistencies) are insensitive to the absolute gain of the radio interferometer as a whole, a problem which is present even performing a simple point source calibration, as we noted in Section 2.1.

This mathematical degeneracy in absolute amplitude calibration can be broken by arbitrarily specifying an absolute gain level. For example, one can add the equation

$$0 = \sum_i \eta_i \quad (10)$$

to the system of equations, thus ensuring that one cannot uniformly elevate the gain levels and still satisfy all constraints. The set of equations governing the phase calibration require the addition of a similar equation, since the absolute phase of a system is not a physical meaningful quantity and therefore must be arbitrarily specified:

$$0 = \sum_i \varphi_i. \quad (11)$$

With the phase calibration, however, there exist two additional degeneracies: the calibration is insensitive to tilts of the entire telescope in either the x or the y direction, since such tilts are equivalent to rotations of the sky and redundant algorithms are (by design) independent of the nature of the sky signal. To break these degeneracies, one adds the following equations:

$$0 = \sum_i r_{x,i} \varphi_i, \quad (12a)$$

$$0 = \sum_i r_{y,i} \varphi_i, \quad (12b)$$

where $\mathbf{r}_i \equiv (r_{x,i}, r_{y,i})$ denotes the physical location of the i th antenna in the array. While these extra equations are somewhat arbitrary (the left-hand side of equation 10, for instance, can be any real number), the true gain, phase and tilt parameters can always be fixed after-the-fact by referring to published flux densities and locations of known, catalogued bright point sources in the final sky maps.

With these four degeneracies broken, our equations can be solved using the familiar least-squares estimator $\hat{\mathbf{x}}$ for the parameter vector:

$$\hat{\mathbf{x}} = [\mathbf{A}^t \mathbf{N}^{-1} \mathbf{A}]^{-1} \mathbf{A}^t \mathbf{N}^{-1} \mathbf{d}. \quad (13)$$

Here, \mathbf{N} is the noise covariance matrix, which takes the form $\langle \text{Re } \mathbf{w} \text{ Re } \mathbf{w}^t \rangle$ for the amplitude calibration and $\langle \text{Im } \mathbf{w} \text{ Im } \mathbf{w}^t \rangle$ for the phase calibration. In the following section we examine the form of this matrix in detail. Note that since the parameter vector \mathbf{x} contains not only the calibration factors η and φ but also the true sky correlations y_{i-j} , one can solve for the true sky signal directly from the uncalibrated correlations \mathbf{d} , as one expects from any self or redundant calibration scheme.

This concludes our review of the Wieringa (1992) redundant calibration method, cast in notation conducive to the development of the new material that follows. It is worth noting that in Wieringa (1992), the noise covariance matrix \mathbf{N} is set to the identity. In the next section, we show that this is *not* the optimal weighting to use for the calibration fit, and derive the noise covariance matrix that corresponds to inverse variance weighting.

2.4 The noise covariance matrix

The form of \mathbf{N} will depend on one's model for the instrumental noise. In what follows, we will for convenience assume that the noise is Gaussian, small compared to the signal, and uncorrelated between baselines; generalization to other noise models is straightforward.⁶ Although these assumptions are not necessary for calibration (since equation 13 can be implemented for any invertible choice of \mathbf{N}), making them will allow us to gain a more intuitive understanding of the errors involved. Recalling the form of the noise term after taking the logarithm of our equations, we can say

$$w_{ij} = \ln \left(1 + \frac{n_{ij}}{g_i^* g_j y_{i-j}} \right) \approx \frac{n_{ij}}{g_i^* g_j y_{i-j}}, \quad (14)$$

where we have invoked the assumption of high signal ($g_i^* g_j y_{i-j}$) to noise (n_{ij}) to expand the logarithm. Since $c_{ij} = g_i^* g_j y_{i-j} + n_{ij}$, we can rewrite this as

$$w_{ij} \approx \frac{n_{ij}}{c_{ij} - n_{ij}} \approx \frac{n_{ij}}{c_{ij}}, \quad (15)$$

⁶ Typically, the electronic noise will be uncorrelated between different antennas/amplifier chains. This means that the noise correlation matrix \mathbf{N} for the baseline will be highly sparse but not diagonal: away from its diagonal, it will have non-zero entries for precisely those pairs of baselines which have an antenna in common. This sparseness is very helpful, making the computations numerically feasible even for massive arrays as described below.

where we have neglected higher order terms. Equation (15) is more convenient than equation (14) because it is written in terms of the measured correlation c_{ij} instead of the noiseless $g_i^* g_j y_{i-j}$. Since the logarithmic scheme separates the real and imaginary parts of the calibration, the crucial quantity is not w_{ij} but rather its real (or imaginary) part:

$$w_{ij} \approx \frac{n_{ij}}{c_{ij}} = \frac{e^{-i\phi}}{|c_{ij}|} (n_x + in_y) \quad (16a)$$

$$= \frac{1}{|c_{ij}|} (\cos \phi - i \sin \phi) (n_x + in_y) \quad (16b)$$

$$\Rightarrow \text{Re } w_{ij} = \frac{1}{|c_{ij}|} (n_x \cos \phi + n_y \sin \phi), \quad (16c)$$

where $\phi \equiv \arg c_{ij}$. With this result, we can form the noise covariance matrix $\mathbf{N} \equiv \langle \text{Re } \mathbf{w} \text{Re } \mathbf{w}^\dagger \rangle$. Since we are assuming that the noise is uncorrelated between different baselines, our matrix must be diagonal, and its components can be written in the form $\mathbf{N}_{\alpha\beta} = \langle (\text{Re } \mathbf{w}_\alpha)^2 \rangle \delta_{\alpha\beta}$, where $\delta_{\alpha\beta}$ is the Kronecker delta, and α and β are Greek *baseline* indices formed from *pairs* of Latin *antenna* indices. For example, the baseline formed by antennas $i = 1$ and $j = 2$ might be labelled baseline $\alpha = 1$, whereas that formed by antennas $i = 1$ and $j = 3$ might be given baseline index $\alpha = 2$. The diagonal components are given by

$$\mathbf{N}_{\alpha\alpha} = \langle n_\alpha n_\alpha \rangle \quad (17a)$$

$$= \frac{1}{|c_\alpha|^2} \left(\underbrace{\langle n_x n_x \rangle}_{\equiv \sigma^2} \cos^2 \phi + 2 \underbrace{\langle n_x n_y \rangle}_{=0} \cos \phi \sin \phi + \underbrace{\langle n_y n_y \rangle}_{\equiv \sigma^2} \sin^2 \phi \right) \quad (17b)$$

$$= \frac{\sigma^2}{|c_\alpha|^2}, \quad (17c)$$

where we have assumed that the real and imaginary parts of the residual noise n are independently Gaussian distributed with spread σ^2 . This expression provides a simple prescription for the inverse variance weighting of equation (13): one simply weights each measured correlation by the square modulus of itself (the σ^2 in the numerator is irrelevant in the computation of the estimator $\hat{\mathbf{x}}$ because the two factors of \mathbf{N} in equation 13 ensure that any constants of proportionality cancel out).

For the phase calibration, the fact that inverse variance weighting corresponds to weighting by $1/|c|^2$ has a simple interpretation. In Fig. 1, we plot various correlations on the complex plane. The green dots represent a set of simulated noiseless visibilities, while the red triangles show what the baselines of a noisy 4×4 regular square antenna array would measure. From the plot, it is clear that for a given amount of noise, visibilities that are closer to the origin of the complex plane (i.e. those with smaller $|c|$) are more susceptible to loss of phase ($\arg c$) information from noise. The gain ($\log |c|$) errors are similarly large for such visibilities. These visibilities should be given less weight in the overall fit, which is what the inverse variance weighting achieves.

2.5 Simulation results and error properties

In Figs 2 and 3, we show simulated calibration results of a 16-antenna element interferometer. The sky signal used in the simulations was that of a Gaussian random field. That such a sky is somewhat unrealistic is irrelevant, since redundant calibration schemes are sky independent⁷ by construction.

Like in Fig. 1, the antennas in these simulations were arranged in a 4×4 square grid, with the antenna spacings assumed to be completely regular. The precise physical separation between adjacent antennas need not be specified, because a dilation of the entire array layout changes only the sky signal measured by the interferometer, and not the *pattern* of baseline redundancies. Repeating the simulations with rescaled versions of the same array is therefore unnecessary, again because of the fact that we can calibrate from any sky signal. The same reasoning makes it unnecessary to specify an observing frequency for the simulations.

In the plots, it is clear that there is some scatter in the recovered antenna parameters. This is due to non-zero instrumental noise, which was simulated to be Gaussian, i.e. n_{ij} in equation (2c) was taken to be a Gaussian random variable. Rather than using a specific noise model to fix the amplitude of the noise, we simply parametrize the problem in terms of the signal-to-noise ratio (S/N). Figs 2 and 3 show the results for an S/N of 10. (Such an S/N would be typical for an interferometer with bandwidth $\Delta\nu \approx 10$ kHz and integration time $\tau \approx 10$ s observing Galactic synchrotron emission at 150 MHz.) The scatter seen in these figures can be viewed as error bars in the final calibration parameters outputted from the calibration. To quantify these errors, we can compute the deviation $\boldsymbol{\varepsilon} \equiv \mathbf{x} - \hat{\mathbf{x}}$ and form the quantity

$$\boldsymbol{\Sigma} \equiv \langle \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\dagger \rangle = [\mathbf{A}^\dagger \mathbf{N}^{-1} \mathbf{A}]^{-1}, \quad (18)$$

where the last equality can be proved with a little algebra and equation (13) (Tegmark 1997). The square root of the diagonal elements of $\boldsymbol{\Sigma}$ then give us the expected error bars on the corresponding elements in $\hat{\mathbf{x}}$. For example, σ_1 , the expected error in the calibration of the first antenna's η , is given by $(\boldsymbol{\Sigma}_{11})^{1/2}$ if one arranges the elements of the parameter vector \mathbf{x} in the manner shown in equation (9).

⁷ That is, provided we exclude unreasonable mathematical exceptions such as a sky devoid of any sources.

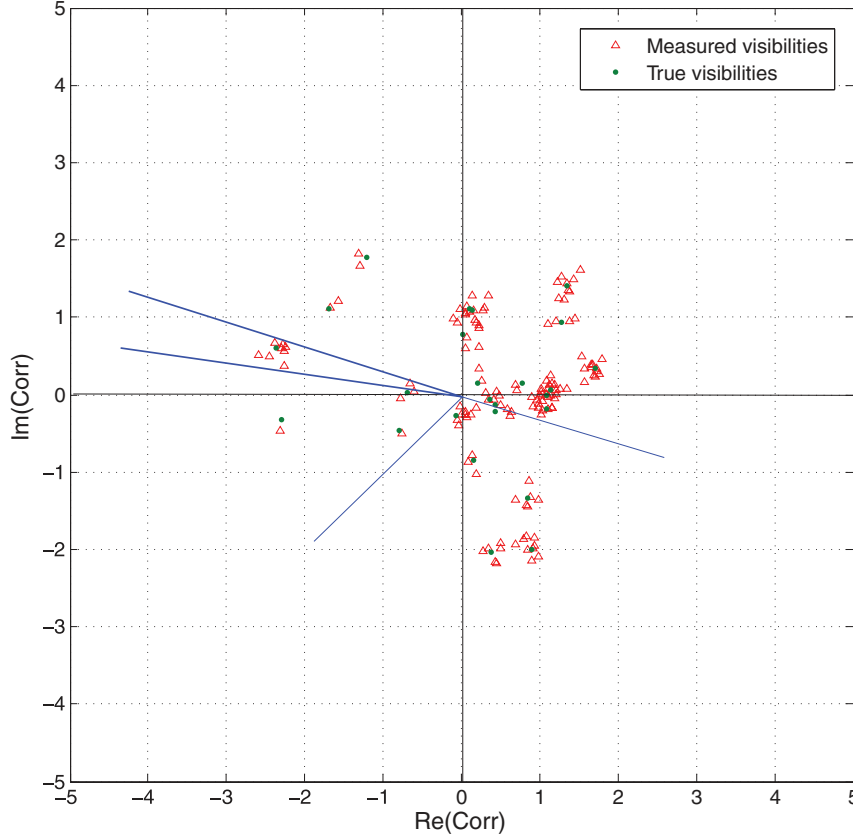


Figure 1. A plot of true visibilities (green dots) and measured noisy visibilities (red triangles) on the complex plane. The blue lines delineate the phase extent of the noisy visibilities from two sets of redundant baselines. One sees that in the presence of noise, visibilities closer to the origin are more susceptible to loss of phase information.

The fact that our system is linear means that to understand the errors in our calibration, we need only compute error bars from simulations of systems with an S/N of 1; the errors from systems of arbitrary S/N can be scaled accordingly. To see this, consider equation (17c), which we can rewrite by scaling out some fiducial mean value $|c_0|$ for the magnitude of the visibilities:

$$\mathbf{N}_{\alpha\beta} = \frac{\sigma^2}{|c_\alpha|^2} \delta_{\alpha\beta} = \frac{\sigma^2}{|c_0|^2 |c'_\alpha|^2} \delta_{\alpha\beta} = \frac{1}{(\text{S/N})^2} \frac{\delta_{\alpha\beta}}{|c'_\alpha|^2} \quad (19a)$$

$$\equiv \frac{\mathbf{N}'_{\alpha\beta}}{(\text{S/N})^2}, \quad (19b)$$

where we have defined normalized visibilities c'_α and a normalized noise covariance matrix $\mathbf{N}' \equiv \delta_{\alpha\beta}/|c'_\alpha|^2$. Plugging this into equation (18) gives

$$\boldsymbol{\Sigma} = [\mathbf{A}'\mathbf{N}'^{-1}\mathbf{A}]^{-1} = \frac{1}{(\text{S/N})^2} [\mathbf{A}'\mathbf{N}'^{-1}\mathbf{A}]^{-1}. \quad (20)$$

We therefore see that in what follows, general conclusions for any S/N can be scaled from our simulations of a system with an S/N of 1.

Although redundant calibration schemes make it *possible* to calibrate an interferometer using any reasonable sky signal, it is important to point out that the *quality* of the calibration, or in other words, the error bars predicted by equation (20), *are* sky dependent. To see this, consider a situation where the sky signal is dominated by a small number of Fourier modes whose wavevectors are of precisely the right magnitude and orientation for its power to be picked up by only the longest baselines of an array. Since our noise covariance matrix is proportional to $1/|c|$, using such a sky to calibrate our array will give small error bars for the antennas that are part of the longest baselines and relatively large error bars for all other antennas. On the other hand, a sky that has comparable power across all scales is likely to yield calibration error bars that are more consistent across different antennas. Given this dependence of calibration quality on sky signal, in what follows we compute ensemble-averaged error bars based on random realizations of our Gaussian skies to give a sense of the typically attainable accuracy.

From our simulations, we find that the dependence of calibration quality on antenna location is quite weak. This can be seen from the scale on Fig. 4, where we show the expected error bars (estimated by taking an average of 30 random realizations) in the recovered η for the antennas that comprise a regularly spaced 18×18 square interferometer. Far more important than antenna location in determining the

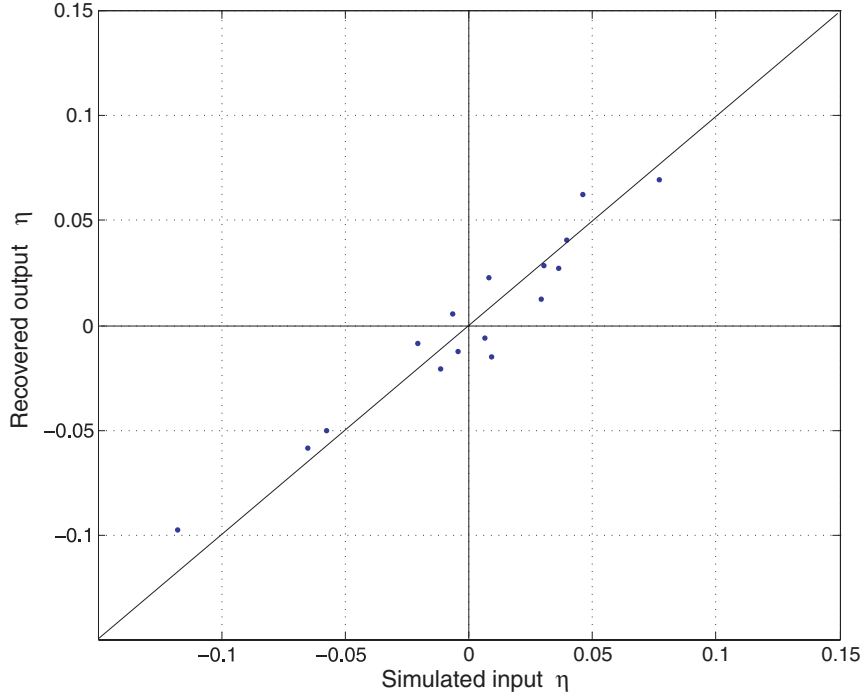


Figure 2. Scatter plot of simulated versus recovered antenna gain parameter η for a 4×4 square array with a S/N of 10.

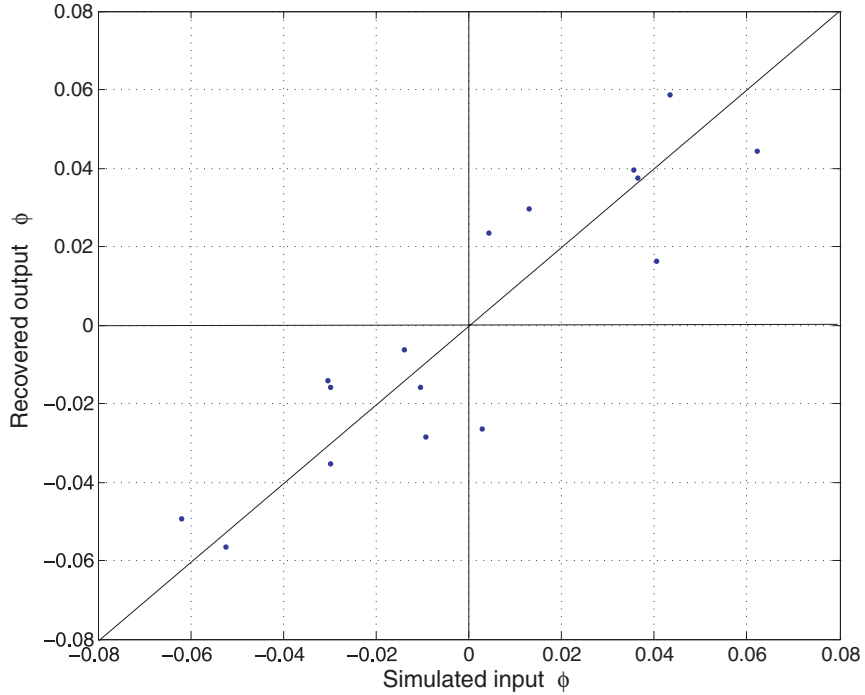


Figure 3. Scatter plot of simulated versus recovered antenna phase parameter ϕ for a 4×4 square array with a S/N of 10.

calibration errors is the *total* number of antennas in an array. In Fig. 5, we show the average⁸ error in η as a function of the number of antennas in a square array. The error $\Sigma^{1/2}$ roughly asymptotes to a $1/\sqrt{N}$ dependence, where N is the total number of antennas in the array, as would be expected if the calibration algorithm is thought of as combining independent information from different redundant baselines. Precisely the same trend is seen in the ϕ errors.

⁸ Averaging both over different antennas in an array and over different simulations.

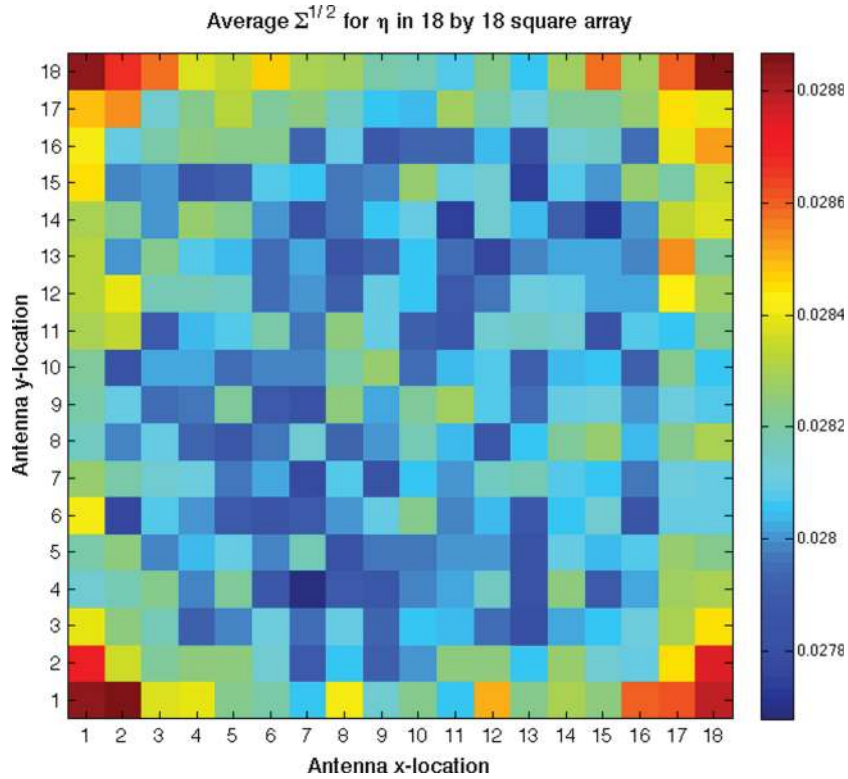


Figure 4. Expected error bars in recovered η for different antennas in an 18×18 square array with a S/N of 1, although in a realistic situation one would expect a more favourable S/N.

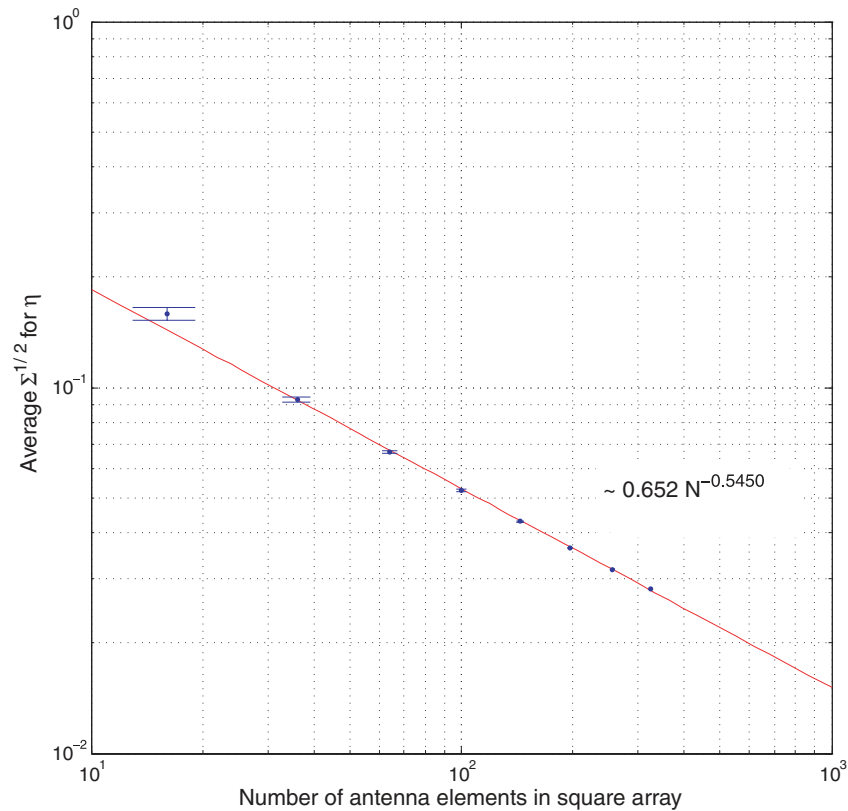


Figure 5. Expected average error bars in recovered η as a function of the size of the simulated square array. All simulations have a S/N of 1, although in a realistic situation one would expect a more favourable S/N.

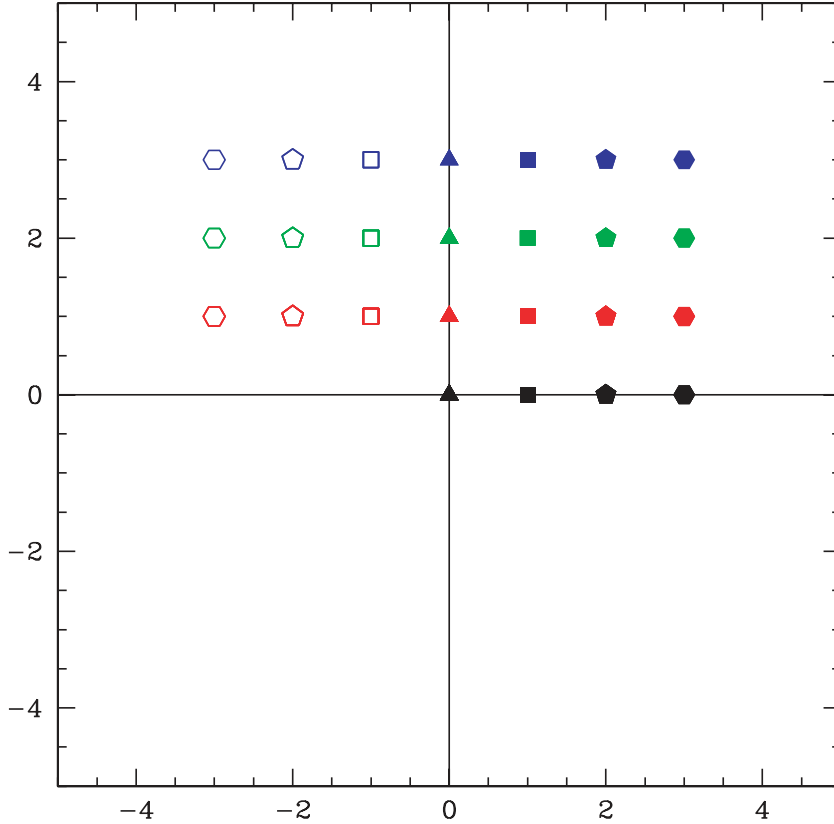


Figure 6. Legend of baselines for Figs 7 and 8.

In running the simulations, we find that minor variations in the baseline distribution have little effect on the final calibration errors.⁹ For instance, varying the aspect ratio of a 100 antenna system from a 10×10 array, to a 5×20 , 4×25 and a 1×100 array has mostly minimal effects on the η and φ errors. Thus, for an array with baseline redundancies that are at roughly the same level as for a square array, we can put everything together and arrive at the following rule of thumb for the calibration errors:

$$\Delta\eta \equiv \Sigma_{\eta}^{1/2} \approx \Delta\varphi \equiv \Sigma_{\varphi}^{1/2} \approx \frac{0.5}{S/N} \frac{1}{\sqrt{N}}, \quad (21)$$

where N is again the total number of antennas in the array.

Not captured by equation (21) is the fact that antennas closer to the centre of an array tend to have lower calibration errors than those that are peripherally located, as can be seen in Fig. 4. Intuitively, this is due to the fact that while every antenna forms the same number of baselines (because each antenna forms a baseline with all other antennas in the array), antennas that are located in different parts of the array form different *sets* of baselines. An antenna that is in the corner of a square array, for example, forms $N - 1$ baselines that are all different from each other, whereas an antenna that is close to the centre will form baselines in redundant pairs because of the symmetry of the square array. In the system of equations, the centrally located antenna will therefore be more tightly constrained by the data, and the resulting error bars in its calibration parameters will be smaller. These differences are, however, rather minimal and are unlikely to significantly affect the actual calibration of a real (non-simulated) array.

As mentioned above, in addition to the antenna calibration parameters η and φ , redundant calibration algorithms yield estimates y_{ij} for the true sky visibilities. In Figs 6–8, we show the visibility results that fall out of the calibration. Fig. 6 serves as a legend for the uv plane, with different colours and symbols signifying the different *unique* baselines found in a 4×4 square array. These symbols are used to plot the uncalibrated input correlations measured by the interferometer (i.e. the c_{ij} s) on a complex plane in Fig. 7. Each colour and symbol combination may appear multiple times in Fig. 7, because two baselines that are identical in principle may give different readings in practice because of instrumental noise and the fact that different antennas will have different calibration parameters. Indeed, one can see from Fig. 7 that the uncalibrated correlations are rather spread out over the complex plane, and do not seem close to the simulated true sky visibilities, which are denoted in the figure by solid magenta dots.

⁹ *Major* variations, of course, can drastically affect the errors. An extreme example of this would be a change in baseline distribution that completely destroys all redundancy, which would render the algorithm unusable and formally give rise to infinite error bars.

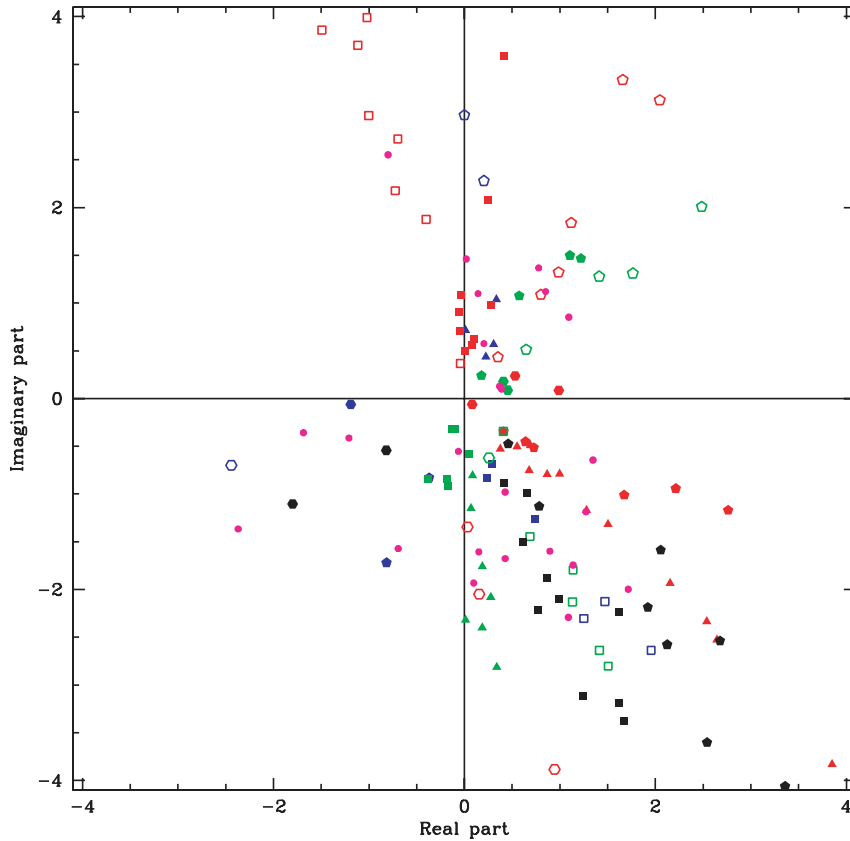


Figure 7. Uncalibrated visibility measurements plotted on the complex plane according to the uv plane baseline legend provided by Fig. 6. The simulated (i.e. true) sky visibilities are given by the magenta dots.

After calibration, however, one has estimates for the antenna calibration parameters, and thus these uncalibrated correlations can be corrected to yield measurements that are close to the correct sky visibilities. One simply divides out the complex gain factors, since the measured correlations c_{ij} are related to the true sky visibilities y_{i-j} by $c_{ij} = g_i^* g_j y_{i-j}$. The results are shown in Fig. 8, where it is clear from the colour and symbol combinations that identical baselines now yield measurements that cluster around the simulated true values, which are still given by the solid magenta dots. The scatter that remains within a given set of identical baselines is due to instrumental noise (set at a level corresponding to an S/N of 10 for this simulation). It should be noted, however, that while this is a perfectly workable method for computing estimates of the sky visibilities, it is unnecessary. Redundant calibration outputs a parameter vector estimate $\hat{\mathbf{x}}$ (see equation (13)) that contains the true visibilities. These are shown in Fig. 8 using magenta xs, and emerge from the calibration at no additional computational cost.¹⁰

Since the true visibilities (like the antenna gain parameters) are simply components of the parameter vector \mathbf{x} , we can again use the machinery of equation (20) to compute the expected error bars in the estimated visibilities. The results are similar to what was found for the antenna calibration parameters, and are summarized in Fig. 9. The errors are again found to asymptote to a $1/\sqrt{N}$ dependence, but here N refers to the *number of redundant baselines that go into determining a given visibility*, and not the total number of antennas in the array. That the error bars do not depend on the total number of antennas is clear from Fig. 9, where data points from simulations of arrays of various sizes lie on top of each other. It is also a property that is intuitively expected, since adding extra antennas increases the mathematical constraints on a given visibility only if the extra antennas form new baselines that correspond to that visibility’s location on the uv plane.

2.6 Problems with the logarithmic implementation

While our simulations and real-world applications like those detailed in Noordam & de Bruyn (1982), Ishiguro (1974), Ramesh et al. (1999), Wieringa (1992), Yang (1988) and Pearson & Readhead (1984) have shown above that taking logarithms of equation (3) yields a linear system that can be successfully used to fit for calibration parameters and visibilities, this logarithmic implementation is not without its drawbacks. In particular, the implementation has two undesirable properties.

¹⁰ In fact, it is not even strictly valid to consider any ‘extra’ computational cost associated with solving for the y_{ij} s, for as one can see from equation (9), the inclusion of the y_{ij} s is *necessary* for redundant calibration to work.

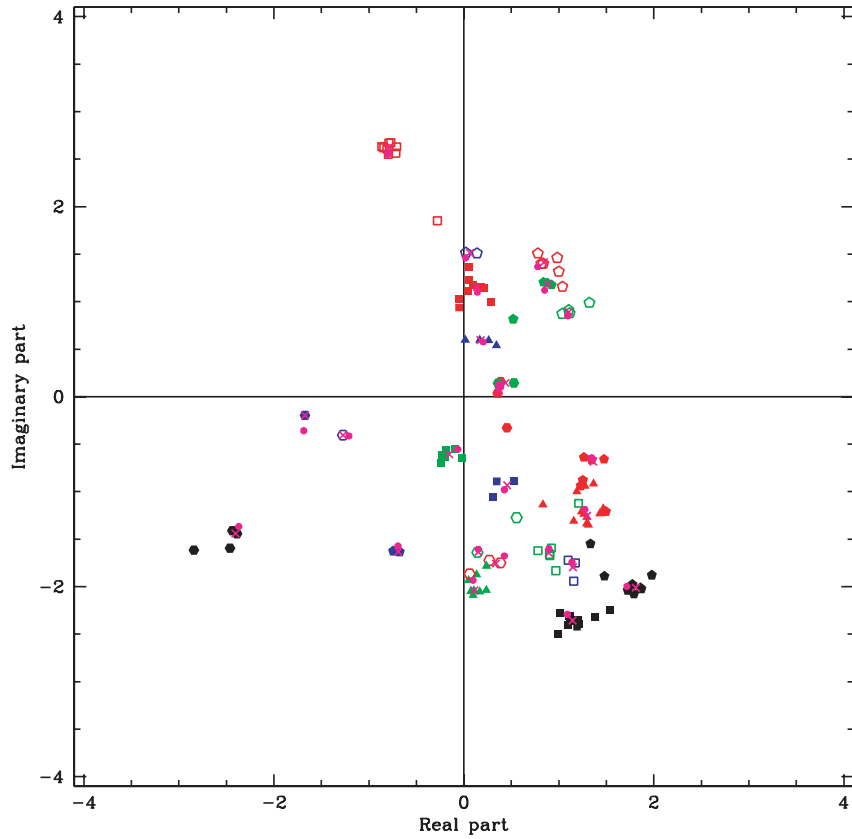


Figure 8. Same as Fig. 7, except the visibilities have been calibrated. The simulated sky visibilities are again given by the magenta dots, while the best estimates from the calibration algorithm for these visibilities are given by the magenta xs.

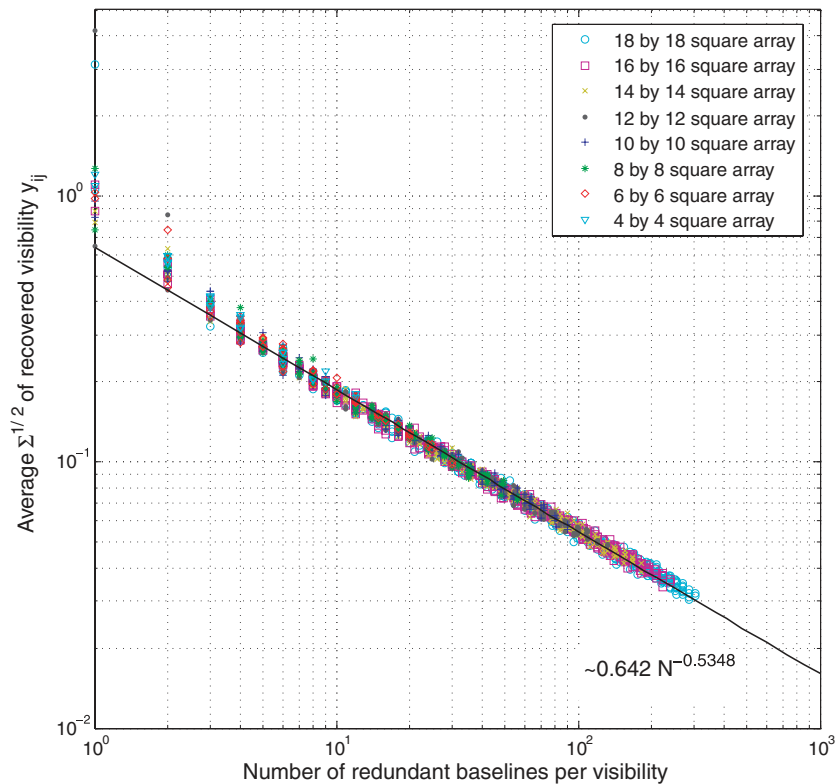


Figure 9. Expected error bars on the estimated visibilities, as a function of the number of redundant baselines that go into determining each visibility. A S/N of 1 is assumed, although in a realistic situation one would expect a more favourable S/N .

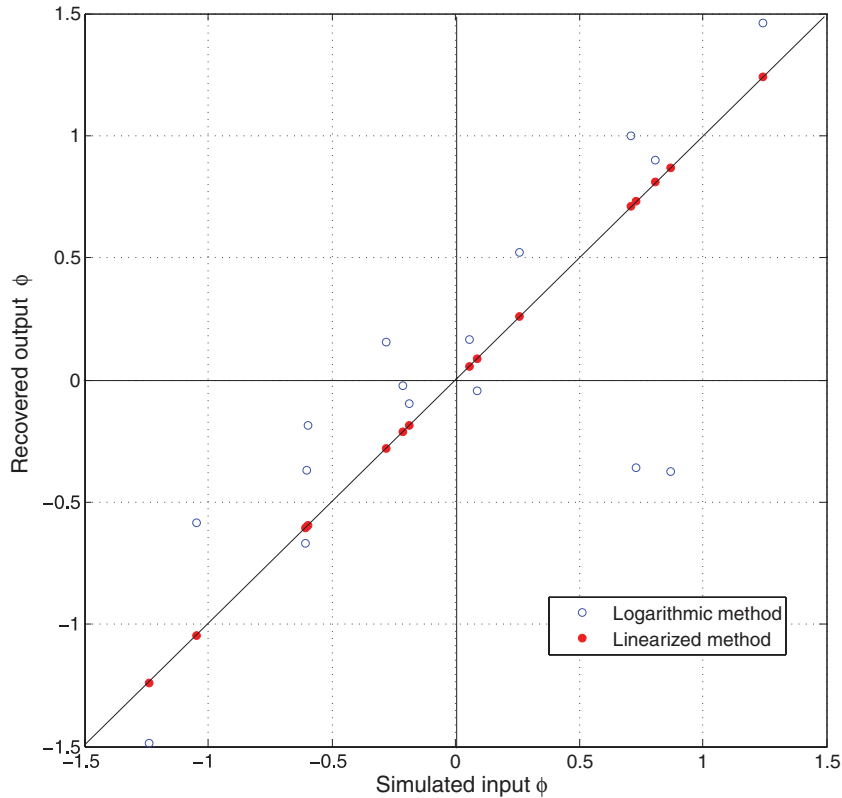


Figure 10. Scatter plots of simulated versus recovered antenna phase parameter ϕ for a *noiseless* 4×4 square array with relatively large phases. The results from the logarithmic implementation described in Section 2.3 are shown using open blue circles, whereas the results from the linear implementation described in Section 2.7 are shown using solid red circles. Large phases are seen to affect the accuracy of the logarithmic method, but not the linear method.

2.6.1 Phase wrapping

For the logarithmic implementation to work, the phase calibration parameter ϕ needs to be close to zero for all antennas. This is because one must take the complex logarithm of $c_{ij} = g_i^* g_j y_{i-j}$, which is an operation that is determined only up to additions or subtractions of multiples of 2π in the imaginary part (i.e. in the phase of the original number). Thus, while the calibration solution may correctly recover the *measured* correlations (since one is insensitive to additions or subtractions of 2π in the phase when re-exponentiating to find $c_{ij} = g_i^* g_j y_{i-j}$), it does not give correct phases for g_i^* , g_j and y_{i-j} , which are ultimately the quantities of interest. This is a problem even when only a small number of antennas have large phase calibration parameters, since to find the visibilities and antenna gain parameters redundant algorithms essentially performs a *global* fit, where all antennas are coupled to each other via the visibilities.

In Fig. 10, we show the phase calibration results for a *noiseless* simulation of a 4×4 array with phases that are significantly greater than zero. The logarithmic method is shown using the open blue circles, and since a noiseless simulation should yield perfect parameter recovery, the scatter in the trend suggests a problem with the method. A detailed examination of the simulation's numerical results reveals that it is indeed the multivalued property of the complex logarithm that causes the trouble, and from the plot one also sees that a small number of very badly fit antennas can give rise to inaccuracies for all antennas, as is expected from the global nature of the fit.

It is important to note that even when the phase calibration fails because of large phases, the *amplitude* calibration can still be implemented successfully. This is because the taking of the complex logarithm means that in our system of equations, the amplitudes and phases separate into the real and imaginary parts, respectively, as we can see from equations (8a) and (8b). The two calibrations are therefore completely decoupled, and instabilities in one do not affect the other. The fact that an accurate gain calibration can be extracted regardless of the quality of one's phase calibration will be important in Section 2.8.

2.6.2 Bias

The logarithmic method is not unbiased, in the sense that ensemble averages of noisy simulations do not converge to the true simulated parameter values. While equation (13) can be shown to be unbiased (Tegmark 1997), the proof assumes that the noise covariance matrix \mathbf{N} is Gaussian and that the weights used in the fits (encoded by \mathbf{N}^{-1}) are independent of the data. In our case both assumptions are violated. While we may assume that n_x and n_y in equation (17b) are Gaussian distributed, this is not the case in amplitude/phase angle space. In addition, since the diagonal entries of \mathbf{N} are taken to be $1/|c|$, the measured data enter the least-squares fit (equation 13) not just through the data vector \mathbf{d} but also via \mathbf{N}^{-1} , making the fitting process a non-linear one.

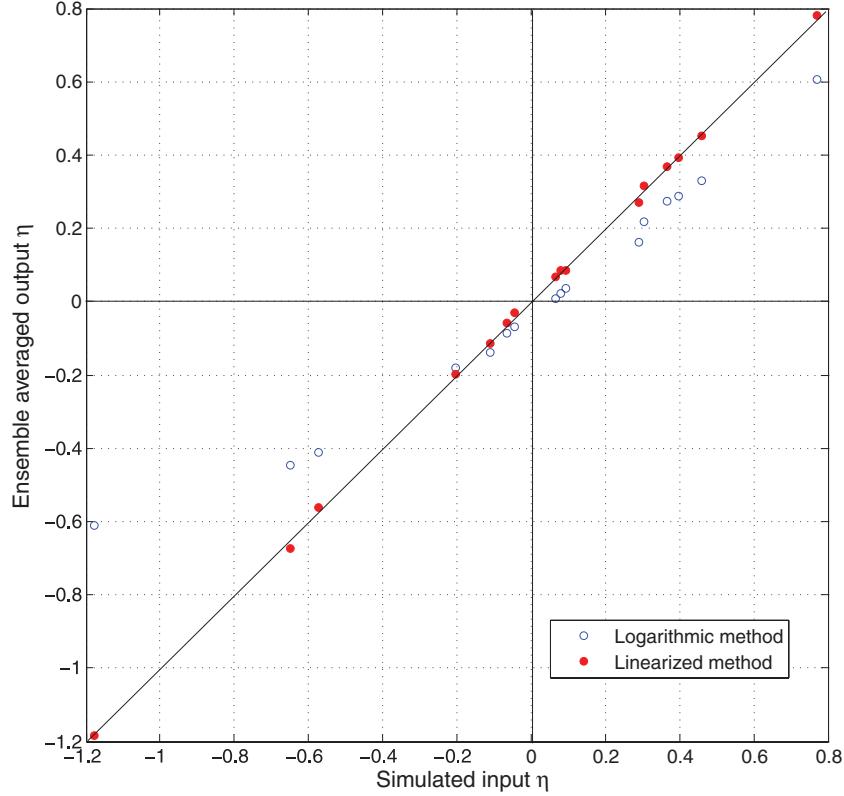


Figure 11. Scatter plots of simulated versus ensemble averaged recovered antenna phase parameter φ for a 4×4 square array with a S/N of 2. The results from the logarithmic implementation described in Section 2.3 are shown using open blue circles, whereas the results from the linear implementation described in Section 2.7 are shown using solid red circles. The logarithmic algorithm gives a biased result, whereas the linear implementation does not.

The bias in the method can be easily seen in Fig. 11, where the blue dots show the simulated results for η , averaged over 90 random ensemble realizations of a 4×4 array with an S/N of 2. There is clearly a systematic bias in the recovered output η s.

2.7 The linearized approach

Since both of the problems discussed in the previous section arise at least partly from taking the logarithm of equation (6), we now propose a method that does not require that step. As an alternative to taking logarithms, one can instead linearize the equations in Section 2.2. We do so by Taylor expanding our expressions about some fiducial guesses η_0 , φ_0 and y_0 for the parameters η , φ and y_0 . The gain of a particular antenna, for example, becomes

$$g \equiv e^{\eta+i\varphi} \quad (22a)$$

$$= g_0 + \left. \frac{\partial g}{\partial \eta} \right|_{\eta_0, \varphi_0} (\eta - \eta_0) + \left. \frac{\partial g}{\partial \varphi} \right|_{\eta_0, \varphi_0} (\varphi - \varphi_0) + \dots \quad (22b)$$

$$= \exp(\eta_0 + i\varphi_0) [1 + (\eta - \eta_0) + i(\varphi - \varphi_0)] + \dots \quad (22c)$$

Along with our fiducial guess for the true correlations (the y s), we define guesses for our measured correlations:

$$c_{ij}^0 \equiv y_{i-j}^0 \exp(\eta_i^0 - i\varphi_i^0) \exp(\eta_j^0 + i\varphi_j^0). \quad (23)$$

In addition, we define the *deviation* δ_{ij} of the guessed correlations from the measured correlations:

$$\delta_{ij} \equiv c_{ij} - c_{ij}^0, \quad (24)$$

as well as an analogous quantity for the true sky correlations:

$$y_{i-j}^1 \equiv y_{i-j} - y_{i-j}^0. \quad (25)$$

To calibrate our telescope, we take the measured deviations δ_{ij} as our inputs and compute the corrections to the guessed antenna parameters and guessed true sky correlations that are required to match these deviations. Plugging our definitions and expansions into $c_{ij} = g_i^* g_j y_{i-j}$ gives

$$\delta_{ij} \approx \exp(\eta_i^0 - i\varphi_i^0) \exp(\eta_j^0 + i\varphi_j^0) [y_{i-j}^1 + y_{i-j}^0 (\Delta\eta_i + \Delta\eta_j - i\Delta\varphi_i + i\Delta\varphi_j)], \quad (26)$$

where we have defined $\Delta\eta \equiv \eta - \eta_0$, $\Delta\varphi \equiv \varphi - \varphi_0$, and have discarded second order terms in the small quantities $\Delta\eta$, $\Delta\varphi$ and y^1 . The result is a linear system in $\Delta\eta$, $\Delta\varphi$ and y^1 , which means we can use the matrix formalism presented in the previous section, the only differences

being that the η and φ equations are now coupled, and that the matrix analogous to the \mathbf{A} used in the logarithmic method – which we will call \mathbf{B} for the linearized method – now depends on the fiducial guesses as well as the array layout. A least-squares fit can once again be employed to solve for the $\Delta\eta$ s, the $\Delta\varphi$ s and y_{i-j}^1 . Once the fit has been obtained, it can be used to update our fiducial guesses, and if desired, one can continue to improve the quality of the calibration by repeating the algorithm iteratively – the updated calibration parameters are simply used as the fiducial guesses for the next cycle of fitting. Note that unlike before, it is unnecessary to use a weighted fit, i.e. one can set $\mathbf{N} = \mathbf{I}$ in equation (13). This is because the linearized algorithm works with visibilities directly (as opposed to, say, their logarithm or some other function of the visibilities), and since our assumption of uncorrelated baseline errors means that individual baselines have the same noise level, all correlations should be weighted the same in the fitting.

In Fig. 12, we show plots of simulated and recovered visibilities on the complex plane as one steps through each iteration of the algorithm. The top left-hand plane shows the simulated correlations as well as the measured correlations. After one iteration, we plot the corrected visibilities as well as the estimated true visibilities in the top right. The recovered visibilities are seen to be still quite different from the simulated ones. The bottom left- and bottom right-hand planes show the second and fourth iterations, respectively, and one can see that the recovered visibilities converge toward their simulated values.

Scatter plots showing the correlation between the simulated and estimated antenna calibration parameters are qualitatively identical to Figs 2 and 3 for the logarithmic implementation, so we do not show them here. However, the linearized approach solves the two problems with the logarithmic algorithm discussed in Section 2.6. In Fig. 10, the filled red circles representing the linearized method show a perfect recovery of simulated phases in the noiseless limit. This demonstrates that the linearized algorithm can tolerate arbitrarily large phases, which the logarithmic method could not. Fig. 11 shows that the ensemble-averaged parameter estimates coming out of the linearized method converge to their true values, demonstrating that the linearized algorithm is unbiased. That this is the case is particularly important if one applies the algorithm to highly noisy systems, for then one can compensate for the low S/N by using massive arrays with large numbers of baselines and by averaging the results over long time periods.¹¹

2.8 Numerical issues for the linearized approach

Although the linearized algorithm evades the problems with the logarithmic approach, it is slightly more complicated to put into practice. Below we discuss two problems that may arise in the numerical implementation of the method and how to solve them.

2.8.1 Convergence

Bad fiducial guesses can result in misfits, although such fits can be easily identified and improved upon. Given that our basic measurement equation (equation 3) is non-linear in our parameters, the linearization that we have presented in this section can only be expected to stably yield good fits when our fiducial guesses are reasonably close to their true values. In a naive implementation of the algorithm, bad fiducial guesses can result in an inability to iterate out of a local maximum in goodness-of-fit. However, bad fits can be readily identified by simply computing the goodness-of-fit parameter, and if necessary, one can repeat the fit with a better search algorithm such as simulated annealing to find the correct maximum.

Another way to avoid misfits is to use the results of the logarithmic algorithm as our initial guess for the linearized method. Even though the phase calibration has been shown to be unreliable when the phases are large, the amplitude calibration can still be trusted since it comes from solving a completely separate system of equations. The fact that the amplitudes estimated using the logarithmic method are biased is not an issue, since our fiducial guess serves merely as a starting point for our iterative linearized scheme.

An attractive approach to implementing our method in practice could be to update the calibration on the time-scale over which the calibration parameters are expected to drift by non-negligible amounts (say once per second or once per minute), using the previously determined calibration parameters as the initial guess and performing merely a single iteration each time. Avoiding multiple iterations in this fashion has the advantage that the noise from any given observation affects the output only linearly, thus avoiding non-linear mappings that can potentially bias the results. The optimal duration between calibrations should be determined to minimize the expected errors: if we calibrate too frequently, the calibration parameters get unnecessarily noisy because so little data are used to determine them, while if we calibrate too rarely, the true calibration parameters can drift appreciably before we recalibrate. An alternative way to determine this trade-off is to re-estimate the calibration parameters more frequently than needed and then smooth the time series, each time replacing the calibration parameter vector actually used by p times the new estimate and $(1 - p)$ times the last estimate, and determining which choice for $p \in (0, 1)$ minimizes the rms calibration errors.

2.8.2 Computational cost

The linear method is more computationally intensive than the logarithmic one, especially if more than one iteration is required. However, the computational cost is still acceptable if one takes advantage of the sparseness of \mathbf{B} . At first sight, the setting of $\mathbf{N} = \mathbf{I}$ in equation (13) for the

¹¹ That is, provided any natural drift in the calibration parameters occur over longer time-scales than the required averaging time.

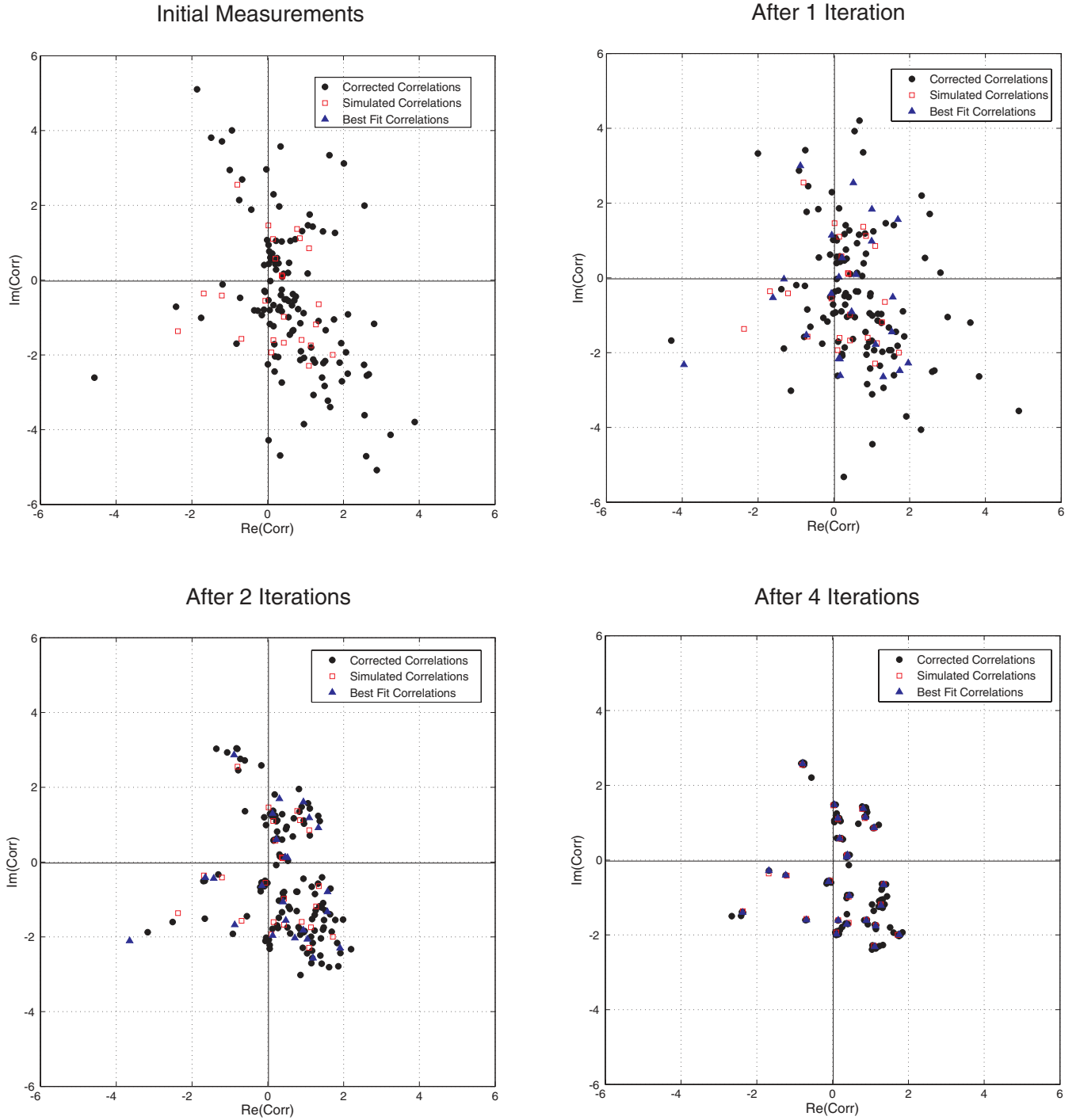


Figure 12. Visibility simulations, measurements and estimated visibilities plotted on complex planes. The top left-hand plane shows the simulated and measured visibilities before any calibration scheme has been applied. The top right-hand plane shows the simulated visibilities, corrected visibilities and best-fitting visibilities after one iteration. The bottom planes show the same quantities after two and four iterations, respectively.

linearized method but not for the logarithmic method would seem to make the linear method computationally quicker. However, the presence of a non-trivial \mathbf{N} in the logarithmic method does not increase the computational cost, for in the absence of cross-talk \mathbf{N} is diagonal. The computational cost of its inversion is therefore negligible, and its presence in equation (13) simply changes the weighting of matrix elements when finding matrix products.

That the linear method is slightly more computationally intensive than the logarithmic method is due to the fact that the matrices for the linear method are larger than those in the logarithmic method. This is because the amplitude and phase calibrations do not decouple in the linearized scheme, and all the parameters must be solved for in one big system. In particular, all vectors are now twice as long as before, since a system where gains and phases are coupled is mathematically equivalent to one where the real and imaginary parts of all complex

numbers are dealt with together. This has its greatest computational impact on the matrix inversion of $\mathbf{B}^t\mathbf{B}$, which can scale as strongly as the vector length cubed if one simply implements the most straightforward matrix inversion routines. Since this inversion must be performed every time the fiducial values (which are part of \mathbf{B}) are updated, the computational time also scales linearly with the number of iterations, and so naively the linearized method seems rather expensive. However, the matrix \mathbf{B} is by construction always sparse, which means the matrix inversion can be performed more efficiently. As an example, one may use the conjugate gradient method, where the sparseness of \mathbf{B} means that the inversion scales linearly with vector length, improving the numerical cost scaling from $\mathcal{O}(N^3)$ to $\mathcal{O}(N)$. This is very similar to the map-making approach successfully implemented by the *Wilkinson Microwave Anisotropy Probe* (WMAP) team in Hinshaw et al. (2003).

It should also be noted that if only one iteration is required and the gain parameters drift only by small amounts over time, then the linearized method can be extremely efficient over long integrations. This is because a *series* of calibrations can then be performed using the same fiducial values for the antenna parameters, and thus the matrix $[\mathbf{B}^t\mathbf{B}]^{-1}\mathbf{B}^t$ never changes and can be computed once and for all.¹² One then simply stores this matrix, which is applied to the data whenever calibration is required. Note that this cannot be done with the logarithmic approach, since the relevant matrix there is $[\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}$ and from equation (17c) we see that \mathbf{N} changes as the measured correlations change, so the combination $[\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}$ must be recomputed every time one would like to calibrate.

If an instrument with N antennas possesses a very large number of redundant baselines (like any reasonably large omniscopes, say, where almost all of the $\sim N^2/2$ baselines are redundant), Fig. 9 suggests that the errors will be extremely small, especially if the S/N is high. With such an instrument, one will be able to calibrate accurately by utilizing merely a small subset of the redundant baselines; for example, if each antenna is correlated with merely 20 others, the resulting system of equations will still be safely overdetermined while keeping the size of the \mathbf{B} matrix and the computational cost reasonable. This is important for any array utilizing fast Fourier transforms for the correlation process, since its attractive $N \log_2 N$ cost scaling would be lost if an $\mathcal{O}(N^2)$ correlator were needed for calibration.

In summary, our proposed calibration method can be performed rather cheaply, especially when one compares the computational cost to that of initially computing the interferometric correlations. To see this, note that one must compute correlations roughly once every 10^{-8} s for an interferometer operating at a frequency of ~ 100 MHz. This is much more expensive than the calibration step (especially if one incorporates the computational tricks suggested in this section), which only needs to be performed, say, once every second or minute. In other words, the correlation needs to be repeated on the time-scale on which the electromagnetic waves oscillate whereas the calibration only needs to be repeated the time-scale on which the calibration parameters drift.

3 A GENERALIZED CALIBRATION FORMALISM: CALIBRATION AS A PERTURBATIVE EXPANSION

In previous sections, the calibration algorithms that were discussed represented two idealized limits: with the bright point source calibration (Section 2.1), the existence of an *isolated* bright point source was key, while with redundant baseline algorithms (Section 2.2) a large number of baselines of equal length and orientation were required. Both algorithms were ‘zeroth-order algorithms’ in that they required exact adherence to these requirements. In this section, we perturb around these idealized assumptions, and show that point source calibration and redundant calibration are simply two extremes in a generalized continuum of calibration schemes.

Consider the sky response (i.e. the visibility) from a baseline \mathbf{b} formed by two antennas at locations \mathbf{r}_i and \mathbf{r}_j (i.e. $\mathbf{b} \equiv \mathbf{r}_i - \mathbf{r}_j$). Assuming that our array is coplanar, we have

$$c(\mathbf{b}) = \int_{\text{sky}} g_i^* g_j \frac{J(\theta)B(\theta)}{\sqrt{1 - \theta_x^2 - \theta_y^2}} \exp\left(-\frac{i2\pi\mathbf{b} \cdot \boldsymbol{\theta}}{\lambda}\right) d^2\boldsymbol{\theta}, \quad (27)$$

where $J(\theta)$ is the intensity of the sky signal from the θ direction, and $B(\theta)$ is the primary beam. Suppose we now define an effective sky signal $I(\boldsymbol{\theta}) \equiv J(\boldsymbol{\theta})B(\boldsymbol{\theta})/\sqrt{1 - \theta_x^2 - \theta_y^2}$. Writing this in terms of its Fourier space description,

$$I(\boldsymbol{\theta}) = \int \tilde{I}(\mathbf{k}) \exp(i2\pi\mathbf{k} \cdot \boldsymbol{\theta}) d^2\mathbf{k}, \quad (28)$$

and inserting this into equation (27), we get

$$c(\mathbf{b}) = g_i^* g_j \tilde{I}\left(\frac{\mathbf{b}}{\lambda}\right), \quad (29)$$

which is a well-known result: baselines of an interferometer probe different Fourier modes of the effective sky signal.

Now suppose that our interferometer possesses a number of nearly redundant baselines. In other words, suppose that there exist a set of vectors $\mathbf{b}_0^\alpha = \{\mathbf{b}_0^1, \mathbf{b}_0^2, \dots\}$ on the uv plane that baselines tend to cluster around. Nearly redundant baselines are considered part of the same cluster, and are associated with a single \mathbf{b}_0^α . On the other hand, any baselines that are isolated on the uv plane are associated with ‘their own’ \mathbf{b}_0^α . Thus, for an array with N antennas the index α runs from 1 to $N(N-1)/2$ if there are zero near or exact redundancies in the baselines,

¹² In principle, one also requires that transients are not present in the data, since transients may cause sudden changes in the visibilities y_{ij} and thus one cannot assume that the same fiducial guesses are good ones over all time. In practice, however, transients can be easily identified in the data, and one can simply ignore the data taken over time intervals where transients dominate.

and to some number substantially smaller than $N(N - 1)/2$ if there are many near or exact redundancies. In other words, we have grouped the baselines into nearly redundant clusters.

We now reparametrize our baseline distribution in terms of $\delta\mathbf{b}_{ij} \equiv \mathbf{b}_{ij} - \mathbf{b}_0^\alpha$, which measures each baseline's deviation from perfect redundancy. Since $\delta\mathbf{b}_{ij}$ is a small quantity, we can Taylor expand equation (29):

$$c_{ij} = g_i^* g_j \tilde{I} \left(\frac{\mathbf{b}_0^\alpha + \delta\mathbf{b}_{ij}}{\lambda} \right) \quad (30a)$$

$$\approx g_i^* g_j \left[\tilde{I} \left(\frac{\mathbf{b}_0^\alpha}{\lambda} \right) + \nabla_u \tilde{I} \Big|_{\mathbf{b}=\mathbf{b}_0^\alpha} \cdot \frac{\delta\mathbf{b}_{ij}}{\lambda} + \dots \right], \quad (30b)$$

where ∇_u denotes a two-dimensional gradient in the uv plane. If we define $c_0^\alpha \equiv \tilde{I}(\mathbf{b}_0^\alpha/\lambda)$, we can rewrite our equation as

$$c_{ij} \approx g_i^* g_j c_0^\alpha \left(1 + \mathbf{h}_0^\alpha \cdot \frac{\delta\mathbf{b}_{ij}}{\lambda} + \dots \right), \quad (31)$$

where $\mathbf{h}_0^\alpha \equiv \nabla \ln \tilde{I} |_{\mathbf{b}=\mathbf{b}_0^\alpha}$.

For algebraic brevity, we proceed using the logarithmic method of Section 2.3, the generalization to the linearized method being straightforward. Taking logarithms of both sides yields

$$\begin{aligned} \ln c_{ij} &\approx (\eta_i + \eta_j) + i(\varphi_j - \varphi_i) + \ln c_0^\alpha + \ln \left(1 + \mathbf{h}_0^\alpha \cdot \frac{\delta\mathbf{b}_{ij}}{\lambda} \right) \\ &\approx (\eta_i + \eta_j) + i(\varphi_j - \varphi_i) + \ln c_0^\alpha + \mathbf{h}_0^\alpha \cdot \frac{\delta\mathbf{b}_{ij}}{\lambda}, \end{aligned} \quad (32)$$

where we have assumed that $\mathbf{h}_0^\alpha \cdot (\delta\mathbf{b}_{ij})/\lambda$ is small in order to expand the last logarithm.

Before we proceed, it is worth examining the precise conditions under which our Taylor series expansions are valid. The crucial quantity is $\mathbf{h}_0^\alpha \cdot (\delta\mathbf{b}_{ij})/\lambda$, which will be small if either \mathbf{h}_0^α or $\delta\mathbf{b}_{ij}/\lambda$ are small. Plugging equation (28) into our definition of \mathbf{h} , we have

$$\mathbf{h} = \frac{\nabla \tilde{I}}{\tilde{I}} = -i2\pi \frac{\int \boldsymbol{\theta} I(\boldsymbol{\theta}) \exp(-i2\pi \mathbf{u} \cdot \boldsymbol{\theta}) d^2\boldsymbol{\theta}}{\int I(\boldsymbol{\theta}) \exp(-i2\pi \mathbf{u} \cdot \boldsymbol{\theta}) d^2\boldsymbol{\theta}}, \quad (33)$$

which is a quantity whose magnitude is at most the size of the primary beam in radians. This dependence on the primary beam means that $\mathbf{h}_0^\alpha \cdot (\delta\mathbf{b}_{ij})/\lambda$ can be made to be small in different ways depending on the instrument one is considering.

(i) For wide field instruments, the primary beamwidth is on the order of 1 rad, so one requires $|\delta\mathbf{b}_{ij}| \ll \lambda$.

(ii) For instruments with narrow primary beams such as the Very Large Array (VLA), we have $|\mathbf{h}| \ll 1$, so the deviations $\delta\mathbf{b}_{ij}$ from perfect redundancies need not be small compared to the wavelength.

If the conditions listed above are satisfied, then calibrating our radio telescope is similar to the zeroth-order case, in the sense that it is once again tantamount to solving equation (32). Here, however, we are *not* guaranteed to have enough equations to solve for all the relevant variables. In addition to the $2N$ antenna calibration parameters (η s and φ s), we now potentially have up to $3N(N - 1)$ numbers to solve for – the $\ln c_0^\alpha$ term contributes up to $N(N - 1)$ numbers, since as we discussed above α can run to $N(N - 1)/2$ in a worst-case scenario, and specifying a single $\ln c_0^\alpha$ requires two numbers: a real part and an imaginary part; the \mathbf{h}_0^α term contributes up to $2N(N - 1)$ numbers – twice that required to specify the $\ln c_0^\alpha$ terms, because \mathbf{h}_0^α is the logarithmic gradient of c_0^α in the uv plane, which is two-dimensional. The $\delta\mathbf{b}_{ij}$ s do not need to be solved for since they can be computed from the layout of the array. Putting this all together, we see that in general there can be up to $3N(N - 1)$ unknowns.

To properly solve equation (32), one must therefore find ways to reduce the range of the index α so that the number of equations exceeds the number of unknowns. In the next few subsections, we will consider different scenarios in which this is the case.

3.1 Zeroth order: perfect point sources or perfect redundancies

The zeroth-order solution to equation (32) involves setting up situations where the $\mathbf{h}_0^\alpha \cdot (\delta\mathbf{b}_{ij})/\lambda$ term can be set to zero. There are two ways to accomplish this.

(i) *Design an array with perfect redundancies, so that $\delta\mathbf{b}_{ij} = 0$.* The form of \mathbf{h}_0^α is then irrelevant (which is another way of saying that we can calibrate with any sky signal), and the only requirement is for the number of *unique* baselines to be small enough that the maximum possible α is modest and there are only a small number of c_0^α terms to fit for. This is precisely the redundant calibration algorithms discussed in Sections 2.2 to 2.8.

(ii) *Have some knowledge of the sky signal used for calibration, so that c_0^α and \mathbf{h}_0^α are known.* For example, if one uses a phased array to image a point calibrator source, then c_0^α becomes a real constant independent of α , while the \mathbf{h}_0^α s all vanish. This is the point source calibration discussed in Section 2.1.

3.2 First order: unknown sources and near perfect redundancies

If one is unable to satisfy the conditions listed in Section 3.1, it may still be possible to calibrate by solving for c_0^α and h_0^α as well as all the calibration parameters simultaneously. For example, taking the real part of equation (32) gives

$$\ln |c_{ij}| = (\eta_i + \eta_j) + \ln |c_0^\alpha| + \operatorname{Re}(h_{0u}^\alpha) \frac{\delta b_{ij}^u}{\lambda} + \operatorname{Re}(h_{0v}^\alpha) \frac{\delta b_{ij}^v}{\lambda}, \quad (34)$$

which can be written as a matrix system:

$$\begin{pmatrix} \ln |c_{12}| \\ \ln |c_{23}| \\ \ln |c_{34}| \\ \ln |c_{45}| \\ \ln |c_{56}| \\ \vdots \\ \ln |c_{29}| \\ \ln |c_{19}| \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 & 0 & \dots & \frac{\delta b_{12}^u}{\lambda} & 0 & \dots & \frac{\delta b_{12}^v}{\lambda} & 0 & \dots \\ 0 & 1 & \dots & 1 & 0 & \dots & \frac{\delta b_{23}^u}{\lambda} & 0 & \dots & \frac{\delta b_{23}^v}{\lambda} & 0 & \dots \\ 0 & 0 & \dots & 1 & 0 & \dots & \frac{\delta b_{34}^u}{\lambda} & 0 & \dots & \frac{\delta b_{34}^v}{\lambda} & 0 & \dots \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 & \frac{\delta b_{45}^u}{\lambda} & \dots & 0 & \frac{\delta b_{45}^v}{\lambda} & \dots \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 & \frac{\delta b_{56}^u}{\lambda} & \dots & 0 & \frac{\delta b_{56}^v}{\lambda} & \dots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots \\ 1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \ln |c_0^1| \\ \ln |c_0^2| \\ \vdots \\ \operatorname{Re}(h_{0u}^1) \\ \operatorname{Re}(h_{0u}^2) \\ \vdots \\ \operatorname{Re}(h_{0v}^1) \\ \operatorname{Re}(h_{0v}^2) \\ \vdots \end{pmatrix}, \quad (35)$$

where for illustrative purposes we have chosen a nine-antenna system with the first three baselines almost redundant. This set of equations can again be solved using equation (13), provided the design of the interferometer is such that the baselines can be grouped into a relatively small number of near-redundant clusters. Fundamentally, the method here is the same as it has been in all previous sections – we have simply written the interferometer response as a linear function of the calibration parameters and the sky signal, and have found ways to reduce the number of parameters (whether by mathematical approximation, interferometer design or careful selection of calibration source) so that the equations are solvable. In the first-order expansion considered in this section, if r denotes the number of b_0^α 's with two or more baselines clustered around them and l denotes the number of completely isolated baselines, the calibration can be solved for provided our array satisfies

$$\frac{N(N-1)}{2} > N + 3r + l, \quad (36)$$

where as usual N is the number of antenna elements in the array. The left-hand side is simply the number of measured complex correlations, while the right-hand side counts the number of complex unknowns we need to solve for: N complex antenna gains, r complex visibilities at each cluster centre, $2r$ complex numbers quantifying deviations in the complex visibilities due to departures from perfect redundancy in the two directions of the uv plane and l complex visibilities for the isolated baselines. For the $s \times s$ regular square arrays simulated in Section 2, the condition is satisfied if $s \geq 4$.

Like before, however, one must be cognizant of possible degeneracies in the matrix system. In particular, consider a cluster comprising of one or two baselines. In fitting for this cluster's visibility, we have one or two inputs (i.e. correlations) but need to solve for three parameters: c_0^α , h_{0u}^α and h_{0v}^α . Our matrix system thus becomes singular. The same problem arises if either δb_{ij}^u or δb_{ij}^v are identical for a large number of baselines in a particular cluster. For instance, if all but one of the baselines in a cluster are perturbed in precisely the same way, then as far as the cluster goes we effectively only have two different baselines and cannot fit for three parameters. These degeneracies can be dealt with in a similar fashion as before: one simply imposes extra constraints on the system of equations, like requiring that h_{0u}^α and h_{0v}^α be zero for single-baseline clusters.

Simulation results for a 4×4 square array with Gaussian antenna position errors are shown in Fig. 13, where we show the rms errors in the recovered gains $\Delta\eta$ as a function of the spread σ_b in the Gaussian from which the position errors are drawn. The simulations were run with precisely the same parameters as those in Section 2, except with zero instrumental noise, so any non-zero error is due purely to position errors (i.e. to the non-perfect redundancy of baselines). Plotted using the solid circles are the results that one obtains if one simply ignores the fact that the baselines are not perfectly redundant, i.e. if one applies a zeroth-order algorithm based on equation (6). The errors are clearly linear in σ_b . Correcting for baseline errors to first order by implementing an algorithm based on equation (31) gives errors that are smaller and quadratic in σ_b . It is also clear from the plot that the wider the primary beam of an instrument, the greater the errors in the calibration. Formally, this is because the beamwidth is directly proportional to the size of the linear baseline error correction term, as we noted above. This result can also be understood intuitively by considering the effect that a primary beam has on uv visibilities. A narrow primary beam acts as a wide convolution kernel on the uv plane, which effectively 'averages' together visibilities over a wide neighbourhood. This smooths the visibilities on the uv plane, making the errors incurred by baseline errors more amenable to a perturbative correction.

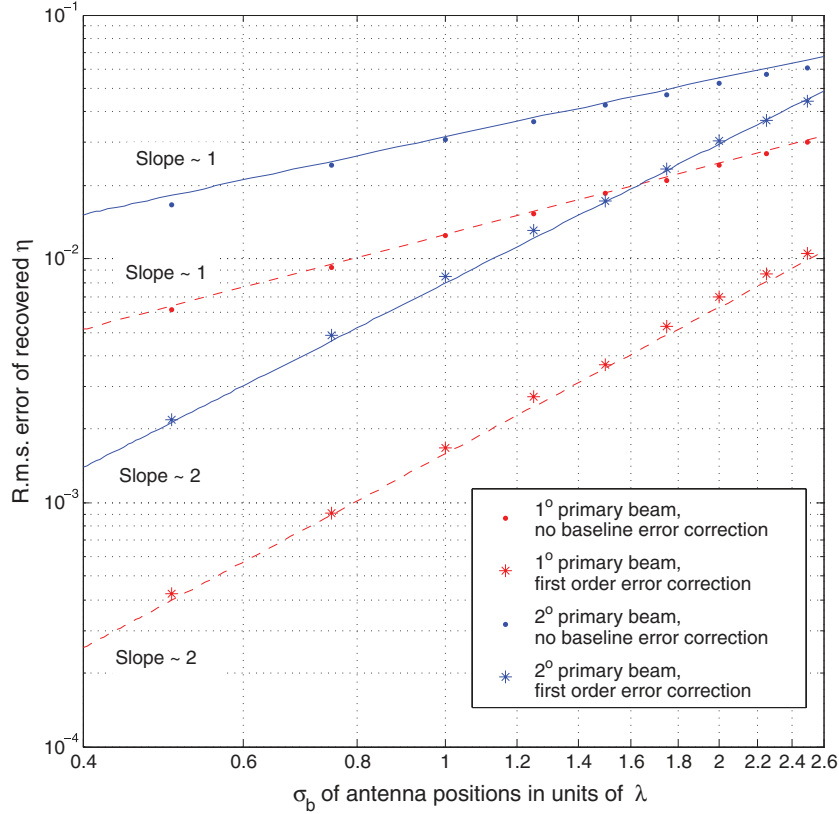


Figure 13. Rms errors in recovered η for one antenna in a noiseless 4×4 square array as a function of antenna position spread σ_b . Shown in red dashed lines are the results for a primary beam with width 1° , while the results for a primary beam with width 2° are shown using the blue solid lines. The solid circles denote results from an algorithm that does not correct for baseline position errors, and the stars denote results from an algorithm that corrects these errors to first order.

In principle, there is no need to stop at the first-order correction. One could, for instance, correct baseline errors to second order by using the following equation as the basic measurement equation:

$$c_{ij} \approx g_i^* g_j c_0^\alpha \left(1 + \mathbf{h}_0^\alpha \cdot \frac{\delta \mathbf{b}_{ij}}{\lambda} + \frac{\delta \mathbf{b}_{ij}^t}{\lambda} \cdot \mathbf{j}_0^\alpha \cdot \frac{\delta \mathbf{b}_{ij}}{\lambda} \dots \right), \quad (37)$$

where $\mathbf{h}_0^\alpha \equiv \nabla \ln \tilde{I}|_{b=b_0^\alpha}$ as before, and \mathbf{j}_0^α is a symmetric matrix of second derivatives of \tilde{I} evaluated at \mathbf{b}_0^α and normalized by $\tilde{I}(\mathbf{b}_0^\alpha/\lambda)$. It is important to note, however, the somewhat counterintuitive fact that for a realistic array (i.e. one with instrumental noise), the errors in the recovered gain parameters may *increase* as one corrects to increasing order in baseline position error $\delta \mathbf{b}_{ij}$. This is because the higher the order that one corrects to, the greater the number of parameters that one must solve for. The number of correlation measurements, however, remains the same as before [with $N(N-1)/2$ of them], so the system of equations becomes less constrained. This makes the calibration algorithm more susceptible to noise, which can negate the benefit that one obtains by correcting baseline position errors to higher order. Put another way, when we introduce a large number of parameters in an attempt to correct for baseline position errors, we run the risk of overfitting the instrumental noise instead of averaging it down using redundant baselines.

In running the baseline error simulations, we find the errors in recovered φ to be much worse than those in the recovered η s. This is likely due to the fact that a small perturbation in antenna position can result in large changes in phases, which (as explained in Section 2.6) can cause problems in a logarithmic implementation.

The reader should also note that because quantities like \mathbf{h}_0^α are dependent on the sky signal, we find that the calibration errors induced by baseline position errors cannot be easily captured by simple formulae analogous to equation (21). An extreme example of this was already discussed above in Section 3.1, where we saw that the correction terms for baseline errors approach zero as the sky becomes increasingly dominated by a single bright point source. The values on the vertical axis of Fig. 13 are thus not generically applicable; in general, one must simulate the relevant array arrangement and sky signal to estimate the errors incurred by antenna position errors.

3.3 First order: non-coplanar arrays

The methods described above can also be used to calibrate radio interferometers even if there are slight deviations from planarity. Suppose the antenna elements of our array are all almost (but not quite) located at a height of $z = 0$. The slight deviations from perfect coplanarity of

the antenna elements result in slight deviations from coplanarity of the baselines, which we denote δb_z^{ij} . With these deviations, equation (27) becomes

$$c(\mathbf{b}) = \int_{\text{sky}} g_i^* g_j \frac{J(\boldsymbol{\theta})B(\boldsymbol{\theta})}{\sqrt{1-\theta_x^2-\theta_y^2}} \exp\left(-\frac{i2\pi\mathbf{b}\cdot\boldsymbol{\theta}}{\lambda} - \frac{i2\pi}{\lambda}\delta b_z^{ij}\sqrt{1-\theta_x^2-\theta_y^2}\right) d^2\boldsymbol{\theta}. \quad (38)$$

Like before, this equation can be manipulated into a form conducive to calibration if one (or both) of the following conditions are met.

(i) *Narrow primary beam.* With a narrow primary beam, one has $\sqrt{1-\theta_x^2-\theta_y^2} \approx 1 - (\theta_x^2/2) - (\theta_y^2/2) \dots \approx 1$. The part of the expression corresponding to the non-coplanar correction thus factors out of the integral, giving

$$c(\mathbf{b}) = \exp\left(-\frac{i2\pi}{\lambda}\delta b_z^{ij}\right) \int_{\text{sky}} g_i^* g_j \frac{J(\boldsymbol{\theta})B(\boldsymbol{\theta})}{\sqrt{1-\theta_x^2-\theta_y^2}} \exp\left(-\frac{i2\pi\mathbf{b}\cdot\boldsymbol{\theta}}{\lambda}\right) d^2\boldsymbol{\theta} = \exp\left(-\frac{i2\pi}{\lambda}\delta b_z^{ij}\right) g_i^* g_j \tilde{I}\left(\frac{\mathbf{b}}{\lambda}\right), \quad (39)$$

where we have adopted the notation defined in equation (28). This illustrates the fact that so long as the primary beam is narrow, deviations from coplanarity simply results in phase shifts in the measured correlations. Aside from these phase shifts, equation (39) looks completely identical to equation (29), and thus the calibration of a non-coplanar array is no harder than the calibration of a planar array when the primary beam is narrow. One simply defines adjusted correlations $\tilde{c} \equiv c \exp[-i(2\pi/\lambda)\delta b_z^{ij}]$ and the ensuing analysis proceeds like before, with no additional computational cost.

(ii) *Near coplanar array.* If the array is close to coplanar ($\delta b_z^{ij} \ll 1$), the calibration can still be performed, but this time at a slight increase in computational cost. With near coplanarity, equation (38) can be Taylor expanded in δb_z :

$$c(\mathbf{b}) = \int_{\text{sky}} g_i^* g_j \frac{J(\boldsymbol{\theta})B(\boldsymbol{\theta})}{\sqrt{1-\theta_x^2-\theta_y^2}} \exp\left(-\frac{i2\pi\mathbf{b}\cdot\boldsymbol{\theta}}{\lambda}\right) \left(1 - \frac{i2\pi}{\lambda}\delta b_z^{ij}\sqrt{1-\theta_x^2-\theta_y^2}\right) d^2\boldsymbol{\theta} \quad (40)$$

$$= g_i^* g_j \tilde{I}\left(\frac{\mathbf{b}}{\lambda}\right) - i g_i^* g_j \frac{2\pi}{\lambda} \delta b_z^{ij} \tilde{K}\left(\frac{\mathbf{b}}{\lambda}\right), \quad (41)$$

where \tilde{I} is defined as before in equations (27) and (28) and

$$\tilde{K}(\mathbf{k}) \equiv \int J(\boldsymbol{\theta})B(\boldsymbol{\theta}) \exp(-i2\pi\mathbf{k}\cdot\boldsymbol{\theta}) d^2\boldsymbol{\theta}. \quad (42)$$

Perturbing this equation around perfectly redundant baselines using exactly the same methods as those employed between equations (30a) and (31), one obtains

$$c(\mathbf{b}) = g_i^* g_j \left[c_0^\alpha \left(1 + \mathbf{h}_0^\alpha \cdot \frac{\delta \mathbf{b}_{ij}}{\lambda}\right) - i2\pi \frac{\delta b_z^{ij}}{\lambda} d_0^\alpha \right], \quad (43)$$

where $d_0^\alpha \equiv \tilde{K}(\mathbf{b}_0^\alpha/\lambda)$. Note that whereas \tilde{I} is expanded to *first* order in $\delta \mathbf{b}$, we are only required to keep the zeroth-order term for \tilde{K} since any linear term in $\delta \mathbf{b}$ would be multiplied by δb_z and end up a second-order term. Proceeding as before, the analogue of equation (32) takes the form

$$\ln c_{ij} \approx (\eta_i + \eta_j) + i(\varphi_j - \varphi_i) + \ln c_0^\alpha + \mathbf{h}_0^\alpha \cdot \frac{\delta \mathbf{b}_{ij}}{\lambda} + i2\pi q_0^\alpha \frac{\delta b_z^{ij}}{\lambda}, \quad (44)$$

where $q_0^\alpha \equiv \ln(d_0^\alpha/c_0^\alpha)$. From here, we can once again form a linear system of equations that can be solved to yield the calibration parameters, the only difference being that we must solve for yet one more parameter per cluster of baselines.

4 CONCLUSIONS

Redundant calibration schemes calibrate radio interferometers by taking advantage of the fact that – once calibrated – redundant sets of baselines should yield identical visibility measurements regardless of what the sky looks like. In this paper we have performed simulations to examine the error properties of redundant calibration, and have found that the usual logarithmic implementations are statistically biased. The linearized implementation introduced in Section 2.7, on the other hand, is both unbiased and computationally feasible, being far less computationally intensive than the standard initial step of computing the signal correlations.

Though for many of our simulations we deliberately chose unfavourable S/Ns to exaggerate the effects of instrumental noise, the errors are seen to scale inversely with S/N, which means that equation (21) and Fig. 9 predict that redundant baseline calibration can yield high precision calibration parameters in a realistic application. Moreover, the fact that the linearized method of Section 2.7 is unbiased means that even if the S/N were unfavourable, one could simply build larger arrays and average over long time periods, in principle suppressing calibration errors to arbitrarily low levels.

We have also shown that both point source calibration and redundant calibration can be considered special cases within a generalized framework of algorithms, all of which find ways to reduce the number of calibration and visibility parameters so that there are few enough of them to be solved for using the $N(N-1)/2$ measurement equations. This can be done either by making a priori assumptions about the calibration sources, or by taking advantage of baseline redundancy. If one has many more constraints than unknowns, non-exact redundancy and non-coplanarity can also be taken into account, which may be essential as the precision requirements for calibration become more

and more stringent. That redundant baseline calibration appears able to satisfy such requirements while being computationally feasible is encouraging for large radio arrays, suggesting that calibration will not limit the vast scientific potential of 21-cm tomography.

ACKNOWLEDGMENTS

We wish to thank Judd Bowman, Michiel Brentjens, Joshua Dillon, Jacqueline Hewitt, Miguel Morales, Ue-Li Pen, Kris Sigurdson, Christopher Williams, Kristian Zarb Adami and Leo Stein for useful discussions. This work was supported by NSF grants AST-0506556, AST-0907969, AST-0708534, AST-6920223 and PHY-0855425 and NASA NN-6056J406 in addition to fellowships from the David and Lucile Packard Foundation, the Research Corporation and the John and Catherine MacArthur Foundation.

REFERENCES

- Barkana R., Loeb A., 2005, *ApJ*, 626, 1
- Bhatnagar S., Cornwell T. J., Golap K., Uson J. M., 2008, *A&A*, 487, 419
- Bowman J. D., Morales M. F., Hewitt J. N., 2007, *ApJ*, 661, 1
- Carilli C. L., Rawlings S., 2004, *Science with the Square Kilometer Array*. Elsevier, Amsterdam
- Chang T., Pen U., Peterson J. B., McDonald P., 2008, *Phys. Rev. Lett.*, 100, 091303
- Cornwell T., Fomalont E. B., 1999, in Taylor G. B., Carilli C. L., Perley R. A., eds, *ASP Conf. Ser. Vol. 180, Synthesis Imaging in Radio Astronomy II*. Astron. Soc. Pac., San Francisco, p. 187
- Datta A., Bhatnagar S., Carilli C. L., 2009, *ApJ*, 703, 1851
- de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, *MNRAS*, 388, 247
- Furlanetto S. R., Sokasian A., Hernquist L., 2004a, *MNRAS*, 347, 187
- Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004b, *ApJ*, 613, 16
- Furlanetto S. R. et al., 2009a, in *Astro2010: The Astronomy and Astrophysics Decadal Survey, Cosmology from the Highly-Redshifted 21 cm Line*, p. 82
- Furlanetto S. R. et al., 2009b, in *Astro2010: The Astronomy and Astrophysics Decadal Survey, Astrophysics from the Highly-Redshifted 21 cm Line*, p. 83
- Garrett M. A., 2009, preprint (arXiv:0909.3147)
- Hinshaw G. et al., 2003, *ApJS*, 148, 63
- Iliev I. T., Shapiro P. R., Ferrara A., Martel H., 2002, *ApJ*, 572, L123
- Ishiguro M., 1974, *A&AS*, 15, 431
- Lidz A., Zahn O., McQuinn M., Zaldarriaga M., Hernquist L., 2008, *ApJ*, 680, 962
- Liu A., Tegmark M., Zaldarriaga M., 2009, *MNRAS*, 394, 1575
- Loeb A., Zaldarriaga M., 2004, *Phys. Rev. Lett.*, 92, 211301
- Lonsdale C. J. et al., 2009, *IEEE Proc.*, 97, 1497
- McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, *ApJ*, 653, 815
- Madau P., Meiksin A., Rees M. J., 1997, *ApJ*, 475, 429
- Mao Y., Tegmark M., McQuinn M., Zaldarriaga M., Zahn O., 2008, *Phys. Rev. D*, 78, 023529
- Morales M. F., 2005, *ApJ*, 619, 678
- Morales M. F., Matejek M., 2009, *MNRAS*, 400, 1814
- Morales M. F., Wyithe J. S. B., 2009, preprint (arXiv:0910.3010)
- Noordam J. E., de Bruyn A. G., 1982, *Nat*, 299, 597
- Parsons A. R. et al., 2010, *AJ*, 139, 1468
- Pearson T. J., Readhead A. C. S., 1984, *ARA&A*, 22, 97
- Peterson J. B., Bandura K., Pen U. L., 2006, preprint (astro-ph/0606104)
- Ramesh R., Subramanian K. R., Sastry C. V., 1999, *A&AS*, 139, 179
- Rau U., Bhatnagar S., Voronkov M. A., Cornwell T. J., 2009, *IEEE Proc.*, 97, 1472
- Santos M. G., Cooray A., 2006, *Phys. Rev. D*, 74, 083517
- Tegmark M., 1997, *ApJ*, 480, L87
- Tegmark M., Zaldarriaga M., 2009a, *Phys. Rev. D*, 79, 083530
- Tegmark M., Zaldarriaga M., 2009b, preprint (arXiv:0909.0001)
- Tozzi P., Madau P., Meiksin A., Rees M. J., 2000a, *ApJ*, 528, 597
- Tozzi P., Madau P., Meiksin A., Rees M. J., 2000b, *Nuclear Phys. B Proc. Suppl.*, 80, C509
- Wieringa M. H., 1992, *Exp. Astron.*, 2, 203
- Wyithe J. S. B., Loeb A., Geil P. M., 2008, *MNRAS*, 383, 1195
- Yang Y., 1988, *A&A*, 189, 361

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.