Mon. Not. R. Astron. Soc. **000**, 1–17 (2004) Printed 4 May 2004 (MN LATEX style file v2.2)

Estimating the bispectrum of the Very Small Array data

Sarah Smith,[1⋆] Graça Rocha,[1,4,5⋆] Anthony Challinor,[1⋆] Richard A. Battye,[2] Pedro Carreira,[2] Kieran Cleary,[2] Rod D. Davies,[2] Richard J. Davis,[2] Clive Dickinson,[2] Ricardo Genova-Santos,[3] Keith Grainge,[1] Carlos M. Gutiérrez,[3] Yaser A. Hafez,[2] Michael P. Hobson,[1] Michael E. Jones,[1] Rüdiger Kneissl,[1] Katy Lancaster,[1] Anthony Lasenby,[1] J. P. Leahy,[2] Klaus Maisinger,[1] Guy G. Pooley,[1] Nutan Rajguru,[1] Rafael Rebolo,[3,6] José Alberto Rubiño-Martin,[3†] Pedro Sosa Molina,[3] Richard D.E. Saunders,[1] Richard S. Savage,[1‡], Paul Scott,[1] Anže Slosar,[1§] Angela C. Taylor,[1] David Titterington,[1] Elizabeth Waldram[1] and Robert A. Watson[2¶]

[1] *Astrophysics Group, Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE*
[2] *University of Manchester, Jodrell Bank Observatory*
[3] *Instituto de Astrofísica de Canarias, 38200 La Laguna, Tenerife, Spain*
[4] *Department of Physics, Nuclear & Astrophysics Laboratory, University of Oxford, Keble Road, Oxford OX1 3RH*
[5] *Centro de Astrofísica da Universidade do Porto, R. das Estrelas s/n, 4150-762 Porto, Portugal*
[6] *Consejo Superior de Investigaciones Científicas, Spain*

arXiv:astro-ph/0401618 v2 4 May 2004

4 May 2004

ABSTRACT

We estimate the bispectrum of the Very Small Array data from the compact and extended configuration observations released in December 2002, and compare our results to those obtained from Gaussian simulations. There is a slight excess of large bispectrum values for two individual fields, but this does not appear when the fields are combined. Given our expected level of residual point sources, we do not expect these to be the source of the discrepancy. Using the compact configuration data, we put an upper limit of 5400 on the value of $f_{\rm NL}$, the non-linear coupling parameter, at 95 per cent confidence. We test our bispectrum estimator using non-Gaussian simulations with a known bispectrum, and recover the input values.

Key words: cosmology: observations – methods: data analysis – cosmic microwave background

1 INTRODUCTION

Currently-favoured cosmological theories predict that the primordial fluctuations in the cosmic microwave background (CMB) obey Gaussian statistics to a high degree. The majority of inflationary scenarios imply a level of non-Gaussianity that is unlikely to be detectable by any forthcoming experiment (Acquaviva et al. 2003; Maldacena 2003), although non-linear gravitational evolution may result in detectable non-Gaussianity from the initially almost-Gaussian fluctuations (Bartolo, Matarrese & Riotto 2004). Any convincing evidence for a departure from primordial Gaussianity would therefore play a very significant role in constraining theories of inflation. However, the dominant contributions to non-Gaussianity in the CMB are expected to come from secondary effects such as gravitational lensing, reionization, the Sunyaev-Zel'dovich effect, and from the local Universe. These effects are of varying significance, depending on the scale. Coupling between lensing and the Sunyaev-Zel'dovich effect tends to dominate over primordial non-Gaussianity on small scales (Goldberg & Spergel 1999). The effects of dust and gas clouds are very dependent on the region of sky observed. The Very Small Array (VSA) fields were carefully chosen to

⋆ E-mail: sjm84@mrao.cam.ac.uk (SS); graca@mrao.cam.ac.uk (GR); a.d.challinor@mrao.cam.ac.uk (AC)
† Present address: Max-Planck Institut für Astrophysik, Garching, Germany.
‡ Present address: Astronomy Centre, University of Sussex.
§ Present address: Faculty of Mathematics and Physics, University of Ljubljana, Slovenia.
¶ Present address: Instituto de Astrofísica de Canarias.

© 2004 RAS

minimise contamination from Galactic emission and bright radio sources, and the signal from Galactic foregrounds is thought to be much less than the primordial CMB signal, with the amount of contamination smaller on smaller scales (Taylor et al. 2003). The recent results from the Wilkinson Microwave Anisotropy Probe (*WMAP*; Komatsu et al. 2003) have considerably tightened the constraints on primordial non-Gaussianity, whilst detecting a non-Gaussian signal arising from residual point sources. The non-Gaussian signal that was found in the bispectrum of the COBE data (Ferreira, Magueijo & Górski 1998) has not been replicated in the *WMAP* data and is now believed to be a result of systematic errors (Magueijo & Medeiros 2003). The VSA has a dedicated point source subtractor, in order to remove all the sources above a certain minimum flux level, so that the residual contribution from unsubtracted sources is less than the flux sensitivity (Taylor et al. 2003).

Whilst we may therefore have little basis for expecting to detect primordial non-Gaussianity in the VSA data, it is important to test the data nevertheless, if only to validate the assumption of Gaussianity which is made during the estimation of the power spectrum and its errors (Hobson & Maisinger 2002). Currently the published VSA data extend past $\ell = 1400$ (Grainge et al. 2003), in comparison with the current *WMAP* data which do not go beyond $\ell = 900$, so we are probing the fluctuations at higher resolution. Testing the data may also help to ascertain whether point source subtraction has been performed satisfactorily, and to see if there is any contamination by foreground sources.

The VSA data have already been tested for non-Gaussianity using a variety of statistics in the map plane, and in the visibility plane by adopting a non-Gaussian likelihood function. The results are presented in Savage et al. (2004). Here, we use the bispectrum, the three-point statistic in the visibility plane.

In Section 2 we discuss the statistics of CMB temperature fluctuations, and how the level of non-Gaussianity can be measured by the bispectrum. We then give a brief overview of the VSA in Section 3, particularly with reference to the point source subtraction technique, and present an approximate expression for the measured visibility three-point function, taking into account the convolution with the primary beam. An exact expression is derived in Appendix C. In Section 4.1 we describe our method for estimating the bispectrum of the VSA data, which we have tested by producing non-Gaussian simulations as described in Section 4.2. We then discuss a method for comparing the results to Gaussian simulations in Section 4.3 and present our results. In Sections 4.4 and 4.5 we consider the effect of point sources on the bispectrum, and look at the feasibility of detecting them in this way. Finally, in Section 4.6 we investigate the constraints that the VSA data are able to place on primordial non-Gaussianity. In the appendices we discuss optimal cubic bispectrum estimators, and the effect of the primary beam on the observed bispectrum from interferometric data.

2 STATISTICS OF CMB TEMPERATURE FLUCTUATIONS

2.1 Power spectrum

Assuming full sky coverage, the temperature fluctuations of the CMB can be decomposed into spherical harmonics ($Y_{\ell m}$), and hence expressed as

$$\frac{\Delta T}{T}(\hat{\boldsymbol{n}}) = \sum_{\ell,m} a_{\ell m} Y_{\ell m}(\hat{\boldsymbol{n}}). \tag{1}$$

Here, $T = T_0$, the mean temperature of the CMB. Rotational invariance demands that $\langle a_{\ell_1 m_1} a^*_{\ell_2 m_2} \rangle = C_{\ell_1} \delta_{\ell_1 \ell_2} \delta_{m_1 m_2}$, where the brackets denote the ensemble average value. The values of the C_ℓ represent the power spectrum of the CMB, which has been the major focus of attention in CMB studies. The measured values of the C_ℓ will differ slightly from the ensemble-averaged C_ℓ due to instrument noise and cosmic variance; if the $a_{\ell m}$ are each drawn independently from a Gaussian distribution, with mean 0 and variance C_ℓ, then the measured C_ℓ will be drawn from a χ^2 distribution, with an intrinsic variance equal to $\frac{2C_\ell^2}{2\ell+1}$.[1] Therefore, particularly at low ℓ, there will always be an uncertainty as to the true value of the C_ℓ, independent of the technical difficulties of experimental measurement.

2.2 Higher-order statistics

If the temperature fluctuations of the CMB are Gaussian, the power spectrum completely describes the statistics. However, if there is a departure from Gaussianity, higher-order statistics are needed for a full description. In principle there is an infinite number of ways in which the CMB could be non-Gaussian, and therefore the optimal statistic to use depends on the type of non-Gaussianity present. If we are looking for a particular signal we can use a statistic which is tailored for optimal detection of that signal. This method has been used to detect non-Gaussianity arising from point sources in the *WMAP* data (Komatsu et al. 2003). However, if we are not seeking a specific signature, then we require something which is fairly general. The natural follow-on from the power spectrum is the bispectrum, which is the three-point function given by

$$B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \equiv \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle. \tag{2}$$

The bispectrum gives a scale-dependent measure of skewness (Santos et al. 2003). The ensemble-average bispectrum will be zero in the case of Gaussian fluctuations. However, for a given realisation, it will be non-zero, even neglecting instrumental noise and resolution, owing to cosmic variance.

The ensemble-averaged bispectrum is constrained by the assumption of rotational invariance in a similar way to the power spectrum. This means that all the information is contained in the dependence of the bispectrum on the values of ℓ, and not in its dependence on m. It can be shown that

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{\ell_1 \ell_2 \ell_3}, \tag{3}$$

[1] In general the $a_{\ell m}$ are complex, and the variance of the real and imaginary parts is $C_\ell/2$, but since they satisfy the relation $a_{\ell m} = (-1)^m a^*_{\ell -m}$ they have $2\ell + 1$ degrees of freedom.

where the matrix represents the Wigner 3-j symbol (Rotenberg et al. 1959). Using the relation

$$\sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}^2 = 1, \quad (4)$$

we find that we can estimate the bispectrum as

$$\widehat{B}_{\ell_1\ell_2\ell_3} = \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}. \quad (5)$$

If the CMB is Gaussian, then the cosmic variance of the bispectrum can be shown to be

$$\left\langle \widehat{B}^2_{\ell_1\ell_2\ell_3} \right\rangle \quad (6)$$
$$= C_{\ell_1} C_{\ell_2} C_{\ell_3} \left(1 + 2\delta_{\ell_1\ell_2}\delta_{\ell_2\ell_3} + \delta_{\ell_1\ell_2} + \delta_{\ell_2\ell_3} + \delta_{\ell_3\ell_1}\right)$$

The variance is twice as large if two ℓs are the same, or six times as large if all the ℓs are equal. This variance is enhanced if we only observe a portion of the sky (sample variance). In real experiments the measured signal will have a contribution arising from instrumental noise, which will be an additional source of variance. It is usually reasonable to assume that the noise is Gaussian, but it is conceivable that this could contribute to a non-zero bispectrum.

In general, we can define any n-point function in harmonic space in a similar way to the power spectrum and bispectrum, and use these as a test for non-Gaussianity (Hu 2002). However, definite detection of a signal arising from true non-Gaussianity at higher orders is progressively more unlikely as n increases.

2.3 Flat-sky approximation

If we are only observing a small patch of sky then we can use the flat-sky approximation (Hu 2000). Instead of decomposing the temperature fluctuation into spherical harmonics, we instead use Fourier modes, so that the temperature fluctuations can now be expressed as

$$\frac{\Delta T}{T}(\hat{\boldsymbol{x}}) = \int a(\boldsymbol{u}) \exp(2\pi i \boldsymbol{u}\cdot\hat{\boldsymbol{x}}) \mathrm{d}^2\boldsymbol{u}. \quad (7)$$

Translational and rotational invariance demand that

$$\langle a(\boldsymbol{u}) a(\boldsymbol{u}')^* \rangle = C(|\boldsymbol{u}|)\delta^2(\boldsymbol{u}-\boldsymbol{u}'),$$

and there is a straightforward correspondence with the full sky power spectrum at large ℓ : $C(|\boldsymbol{u}|) \approx C_\ell \mid_{l=2\pi|\boldsymbol{u}|}$. Similarly, assuming rotational and parity invariance,

$$\langle a(\boldsymbol{u}_1) a(\boldsymbol{u}_2) a(\boldsymbol{u}_3) \rangle = B(\ell_1,\ell_2,\ell_3)\delta^2(\boldsymbol{u}_1+\boldsymbol{u}_2+\boldsymbol{u}_3), \quad (8)$$

with the correspondence $\ell_i = 2\pi\,|\boldsymbol{u}_i|$. Isotropy demands that the bispectrum is zero unless the three u-vectors sum to zero, so the bispectrum is specified by just two vectors, hence its name. Isotropy and parity invariance demand that the permutation-symmetric $B(\ell_1,\ell_2,\ell_3)$ depends only on the lengths of $\boldsymbol{u}_1$,$\boldsymbol{u}_2$ and $\boldsymbol{u}_3$. At large ℓ we can write $B(\ell_1,\ell_2,\ell_3) \approx b_{\ell_1\ell_2\ell_3}$, where $b_{\ell_1\ell_2\ell_3}$ is the reduced bispectrum (Komatsu & Spergel 2001):

$$B_{\ell_1\ell_2\ell_3} = \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} b_{\ell_1\ell_2\ell_3}.$$

The size of the VSA primary beam in the compact configuration implies that the flat-sky approximation is not entirely accurate. However, the correction factor is small, as described in Hobson & Maisinger (2002). The correction can be made be redefining the primary beam, and this is done in the power spectrum analysis. Our current bispectrum calculation does not directly take this effect into account, but we anticipate the associated error in the approximation to be small.

3 THE VERY SMALL ARRAY

The VSA is a 14-element interferometer situated on Mount Teide in Tenerife. It has the ability to make observations anywhere in the frequency range 26–36 GHz with an observing bandwidth of 1.5 GHz. During the first observing season (September 2000 – September 2001) the antennae, of FWHM $4\overset{\circ}{.}6$, were arranged in a compact configuration, with a maximum baseline of 1.5 m, and observations were made at 34 GHz. In this configuration the VSA was sensitive to angular scales corresponding to $\ell \sim 150$–900. The VSA was then reconfigured to form the extended array, with antennae of FWHM $2\overset{\circ}{.}05$ separated by a maximum distance of 2.5 m, sensitive to the range $\ell \sim 300$–1400. It has been making observations in this configuration since October 2001, at 33–34 GHz.

The compact array spent most of its time observing three separate regions of sky (labelled VSA1, VSA2 and VSA3). Within each region it made overlapping (mosaiced) observations of two or three fields (denoted by either no suffix, or the suffixes A, B or -OFF). Details of the fields are given in Taylor et al. (2003), and the power spectrum calculated from these observations is presented in Scott et al. (2003).

The extended array made mosaiced observations of smaller fields within the same three regions of the sky (labelled with the suffixes E, F and G), details of which can be found in Grainge et al. (2003). Here, we have studied the compact and extended array data that were used to compute the CMB power spectrum presented in Grainge et al. (2003). Since completing the analysis here there has been a new release of results from further extended array observations (Dickinson et al. 2004). Further details of the VSA observational technique can be found in Watson et al. (2003).

3.1 Interferometer measurements

An interferometer samples the Fourier transform of the sky intensity multiplied by the primary beam. The fact that an interferometer measures directly in Fourier space makes interferometric data particularly suited to analysis in Fourier space. The visibility measured by the interferometer can be expressed as

$$V(\boldsymbol{u}) = \int \Delta I(\hat{\boldsymbol{x}})\; A(\hat{\boldsymbol{x}})\, e^{2\pi i \boldsymbol{u}\cdot\hat{\boldsymbol{x}}}\, \mathrm{d}^2\hat{\boldsymbol{x}} + N(\boldsymbol{u}), \quad (9)$$

where $\Delta I(\hat{\boldsymbol{x}})$ is the intensity fluctuation, $A(\hat{\boldsymbol{x}})$ the primary beam and $N(\boldsymbol{u})$ the noise on baseline $\boldsymbol{u}$. To a good approximation the beamshape can be modelled as Gaussian with $A(\boldsymbol{x}) = \exp(-|\boldsymbol{x}|^2/2\sigma^2)$ (Hobson & Maisinger 2002), which is the shape that we have assumed when simulating the observations. In Appendix B we describe how this leads to the approximate expression for the variance of the visibilities

$$\langle V(\boldsymbol{u}) V^*(\boldsymbol{u}) \rangle \approx \pi\sigma^2 f^2 C(u) + \sigma^2_{\boldsymbol{u}}, \quad (10)$$

where

$$f = T_0 \frac{\partial B(\nu, T)}{\partial T}\bigg|_{T=T_0} \tag{11}$$

is the conversion to specific intensity and $\sigma_{\boldsymbol{u}}^2$ is the variance of the noise. Similarly, for the three-point function we obtain

$$\langle V(\boldsymbol{u}_1)V(\boldsymbol{u}_2)V(\boldsymbol{u}_3)\rangle \approx f^3 \frac{2}{3}\pi\sigma^2 B(u_1, u_2, u_3), \tag{12}$$

for $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = 0$. This approximation has been used to convert the measured three-point function to a bispectrum in units of μK^3 for the purposes of plotting. In Appendix C we derive an exact expression for a general case with a Gaussian beam which it is necessary to use when trying to quantify a true level of non-Gaussianity and considering a particular non-Gaussian signal.

The sample time per visibility measurement by the VSA is typically 64 seconds. Each field was observed for ∼100–200 hours, resulting in a very large number of individual visibility measurements. To reduce the data to a manageable level, the measurements are binned into square cells on the uv plane, using a maximum-likelihood method as described in Hobson & Maisinger (2002). Adjacent cells are still highly correlated as the cell width is chosen to be less than the width of the aperture function (the Fourier transform of the primary beam) so as not to lose information.

3.2 Source subtraction

The VSA regions were carefully chosen in order to have relatively low dust contamination (using the dust maps of Schlegel, Finkbeiner & Davis 1998), low Galactic free-free and synchrotron emission (as predicted by the 408-MHz all-sky survey of Haslam et al. 1982) and to avoid bright radio sources (Taylor et al. 2003). Choosing to observe at the higher end of the VSA frequency range reduces the signal from Galactic foregrounds (Taylor et al. 2003). The compact array fields were chosen so as to contain no sources brighter than 0.5 Jy. In order to eliminate contamination from point sources, which would otherwise contribute to excess power, they are observed and directly subtracted from the data. The source subtraction strategy used was first to survey the fields with the Ryle telescope in Cambridge at 15 GHz prior to observation with the VSA. Then, simultaneous with VSA observations, a single-baseline interferometer, with a dish separation of 9 m, situated alongside the VSA, is used to monitor the sources identified in the 15-GHz survey. For the compact array, it was calculated that all sources brighter than 80 mJy needed to be subtracted. Therefore the survey with the Ryle telescope sought to identify all sources above 20 mJy at 15 GHz. The power of the CMB is more sensitive to sources at higher values of ℓ, and so for the extended array the fields were chosen, using the previous survey, to contain no sources brighter than 100 mJy (Grainge et al. 2003). The sensitivity of the Ryle survey was increased to 10 mJy, and all sources brighter than 20 mJy subtracted from the measured visibilities.

4 CALCULATION OF THE BISPECTRUM

4.1 Method of estimating the bispectrum

Our starting point for estimation of the bispectrum of the VSA data is equation (8). The delta function tells us that we should be interested only in sets of three vectors in the uv plane that form a closed triangle. The lengths of the sides of the triangle give us the values of ℓ_1, ℓ_2 and ℓ_3. It follows that we can make an estimate of $B(\ell_1, \ell_2, \ell_3)$ by finding from the data all possible triangles with these sides.

There are two additional factors to consider:

(i) The finite beamwidth means that the measured visibility on baseline $\boldsymbol{u}$ is a convolution with the aperture function (the Fourier transform of the primary beam), as shown by equation (B5), so the visibilities are 'smeared' over a region of diameter a few wavelengths on the uv plane. Therefore it is not desirable to form separate bispectrum estimates using nearby visibility points. It is preferable to form individual estimates from 'bands' of points on the uv plane, else adjacent estimates will be highly correlated. This also means that we can slightly relax the requirement that the three vectors must form a closed triangle, as long as the deviation from a closed triangle is significantly less than the size of the aperture function.

(ii) The quantity that we are interested in is the ensemble-averaged value of $B(\ell_1, \ell_2, \ell_3)$, however, since we are observing only a small region of our Universe, we want to average over as many triangles as is reasonable in order to reduce the variance of the estimate.

In a general non-Gaussian scenario we would expect $B(\ell_1, \ell_2, \ell_3)$ to vary reasonably slowly with ℓ; if it fluctuates too rapidly then the signal will be 'washed out' both by the convolution and by combining nearby points into one estimate. Predictions for the primordial bispectrum (see Section 4.6) show a similar number of peaks to the power spectrum (if we fix two values of ℓ and vary the third), however, the bispectrum fluctuates between positive and negative values.

Therefore, following the method used in Santos et al. (2003) for the frequentist estimator, we divide the uv plane into concentric annuli, of width Δ. To form a bispectrum estimate, three annuli are selected, and we set ℓ_1, ℓ_2 and ℓ_3 equal to the radii of the midpoints of the annuli. The data are binned on a square grid in the uv plane. This binning shifts the points very slightly and hence is equivalent to relaxing the requirement of exact triangles. We find all the possible triangles formed from vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ where $\boldsymbol{u}_i$ points to the centre of a cell containing a data point in annulus i. We can express our estimator as:

$$\hat{B}(\ell_1, \ell_2, \ell_3) = \frac{\sum_{\boldsymbol{u}_i \in A_i} \frac{V(\boldsymbol{u}_1)V(\boldsymbol{u}_2)V(\boldsymbol{u}_3)}{\sigma^2_{\boldsymbol{u}_1\boldsymbol{u}_2\boldsymbol{u}_3}}}{F \sum \frac{1}{\sigma^2_{\boldsymbol{u}_1\boldsymbol{u}_2\boldsymbol{u}_3}}}. \tag{13}$$

where $F = f^3 \frac{2}{3}\pi\sigma^2$, the prefactor in equation (12), which accounts for the conversion to flux density and the effect of the primary beam. The σ^2 are used to weight each individual triangle of vectors, and are estimated as explained in Section 4.1.2. A_i denotes annulus i. The visibilities are complex, but the resulting bispectrum estimates are real since $a(-\boldsymbol{u}) =$

$a^*(\boldsymbol{u})$, by virtue of the fact that the temperature field is purely real.[2]

Our method for estimating the bispectrum from interferometer data mirrors that used for the *MAXIMA* data (Santos et al. 2003). However, the *MAXIMA* data are obtained in real space and so it is necessary first to perform a Fourier transform. We begin with the data in the uv plane, so we do not have to concern ourselves with window functions. However, we have to deal with non-uniform uv plane coverage, which results in large variations in the noise on each cell. There is also a lot of variation in the number of triangles of vectors which form each bispectrum estimate.

4.1.1 Choice of Δ

A natural scale for Δ, the width of the annuli, is set by the width of the aperture function. Increasing Δ increases the number of triangles that form each bispectrum estimate, but they correspond to a wider spread of the underlying values of ℓ. This loss of $\Delta\ell$ resolution runs the risk of 'washing out' any oscillatory signal that may be present. However, if Δ is too small then the estimates have large variances resulting from there being few triangles forming each estimate. In addition, adjacent estimates would be highly correlated. Therefore we chose a value of Δ which is large enough to encompass the width of the aperture function, but with the relative size compared to the aperture function slightly smaller for the case of the extended array, since it has a larger aperture function. Fig. 1 shows how particular bispectrum estimates vary with the width of the annuli. For the compact array we chose $\Delta = 16\ \lambda$, and for the extended array, $\Delta = 32.4\ \lambda$. This is equivalent to 1.46 and 1.31 times the FWHM of the aperture function for the compact and extended arrays respectively.

4.1.2 Weighting the bispectrum estimates

We can achieve a near-optimal estimator by weighting each visibility with its inverse variance. In the absence of primary beam effects, this estimator would be the optimal cubic estimator of the bispectrum (see Appendix A). We thus weight each individual triangle as $\frac{1}{\sigma^2_{\boldsymbol{u}_1\boldsymbol{u}_2\boldsymbol{u}_3}}$ where

$$\sigma^2_{\boldsymbol{u}_1\boldsymbol{u}_2\boldsymbol{u}_3} = \left(\widetilde{C}_{\ell_1} + \sigma^2_{\boldsymbol{u}_1}\right)\left(\widetilde{C}_{\ell_2} + \sigma^2_{\boldsymbol{u}_2}\right)\left(\widetilde{C}_{\ell_3} + \sigma^2_{\boldsymbol{u}_3}\right), \qquad (14)$$

is the product of the variances of the visibilities. Here, $\widetilde{C}_{\ell_i}$ is given by $\pi\sigma^2 f^2 C(\boldsymbol{u}_i)$ as shown in equation (10).

We can make a rough estimate of the variance of the bispectrum by ignoring correlations due to the primary beam, and neglecting the fact that some visibilities are used more than once in the bispectrum estimate. Under the assumption of a Gaussian signal, each triangle is the product of three independent visibilities, and nearly all triangles in a given estimate $\hat{B}$ are also independent. It follows that

[2] Although the signal contribution to the visibilities satisfies $a^*(\boldsymbol{u}) = a(-\boldsymbol{u})$, the noise is generally uncorrelated between visibilities. In our analysis we impose the symmetry $a^*(\boldsymbol{u}) = a(-\boldsymbol{u})$ on the noise too by averaging the data $a(\boldsymbol{u})$ and $a^*(-\boldsymbol{u})$ to form the visibility at $\boldsymbol{u}$.

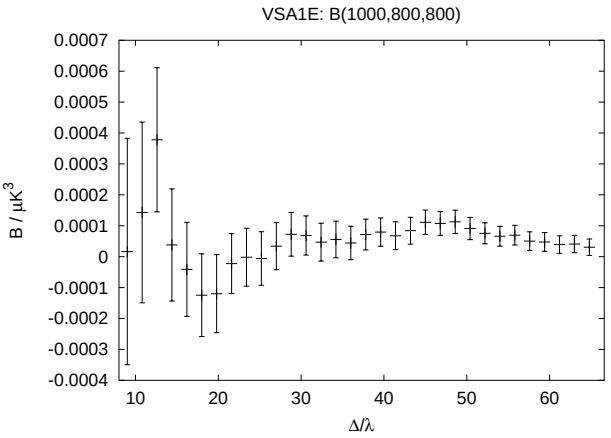


Figure 1. Illustration of the variation of a bispectrum estimate with Δ, the width of the annuli, for the extended array field VSA1E, with estimated variance calculated according to equation (15). For small values, there are few triangles and so the estimate is large, and fluctuates rapidly. The bispectrum appears to converge towards zero as we increase Δ, and varies smoothly with Δ. We wish to chose the smallest value of Δ for which the bispectrum estimate is changing reasonably slowly, and for which adjacent bispectrum estimates are uncorrelated.

$$\sigma^2_{\hat{B}} \approx \frac{1}{F^2 \sum_{\boldsymbol{u}_i \in A_i} \frac{1}{\sigma^2_{\boldsymbol{u}_1\boldsymbol{u}_2\boldsymbol{u}_3}}}. \qquad (15)$$

In practice, we compute errors and assess the statistical significance of our results using Gaussian simulations which properly take account of correlations, however, we can use equation (15) as a way of checking the calculation of the weights as given by equation (14).

4.2 Simulations

4.2.1 Gaussian Simulations

In order to assess the statistical significance of our results, we have performed Gaussian simulations of the VSA data from a simulated Gaussian sky. We extract visibilities at the same uv positions as our data, and add noise which is drawn from a Gaussian distribution with the same variance as the noise on the data. For each simulation we compute the bispectrum, with the same code as that employed on the real data. From the suite of 15000 simulations we estimate the variance in each bispectrum estimate, and their covariances (which arises from a combination of sample variance and instrumental noise).

The power spectrum for the simulations is drawn from a flat ΛCDM model, with the same parameters as model A in Slosar et al. (2003), using all CMB data. Model A was the simplest and had the highest evidence.

Fig. 2 shows the distribution of some bispectrum estimates obtained from the Gaussian simulations, together with the value calculated from the real data. The estimates which are formed from more triangles tend to have a smaller variance as would be expected. The majority of the distributions fit well with a Gaussian curve with the same variance, although there are exceptions (which tend to be when some of the ℓs are low) such as the distribution of $B(773, 366, 366)$.

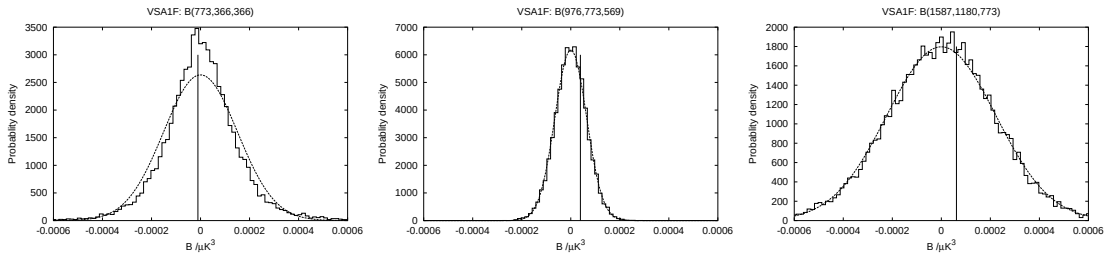


Figure 2. Distribution of bispectrum estimates for VSA1F obtained after 15000 simulations, and comparison with a Gaussian distribution with the same variance. The measured value from the real data is shown by a vertical line.

4.2.2 Non-Gaussian Simulations

It is desirable to test the bispectrum estimator on simulated data with a known bispectrum to check for bias. In order to do this, we adopt a probability distribution function (pdf) derived from the Hilbert space of a linear harmonic oscillator, developed by Rocha et al. (2001). This exercise helps us to assess the performance of our code to compute the bispectrum, and could also be a useful alternative hypothesis in testing for non-Gaussianity. We give here a brief account of the procedure followed to simulate these non-Gaussian CMB maps; a more detailed description will be presented in Rocha et al. (in preparation).

We start by drawing the values of the CMB temperature fluctuations, $\Delta T(\hat{\boldsymbol{x}}_p)$, independently in each real-space pixel from our non-Gaussian pdf. Since the pdf is based on the wavefunctions of the eigenstates of a linear harmonic oscillator, it takes the form of a Gaussian multiplied by the square of a (possibly finite) series of Hermite polynomials where the coefficients α_n are used as non-Gaussian qualifiers. These amplitudes α_n can be written as series of cumulants (Contaldi, Bean & Magueijo 1999), and can be independently set to zero without mathematical inconsistency (Rocha et al. 2001). Hence we should regard α_n as non-perturbative generalisations of cumulants. Let x represent a general random variable, within a set of variables which are assumed to be independent. The most general probability density for the fluctuations in x is thus:

$$P(x) = |\psi(x)|^2 = e^{-\frac{x^2}{2\sigma_0^2}} \left| \sum_n \alpha_n C_n H_n\left(\frac{x}{\sqrt{2}\sigma_0}\right)\right|^2, \qquad (16)$$

where the H_n are the Hermite polynomials, and the quantity σ_0^2 is the variance associated with the (Gaussian) probability distribution for the ground state $|\psi_0|^2$. The C_n are fixed by normalising the individual states. The only constraint on the amplitudes α_n is:

$$\sum |\alpha_n|^2 = 1. \qquad (17)$$

This is a simple algebraic expression which can be eliminated explicitly by setting $\alpha_0 = \sqrt{1 - \sum_1^\infty |\alpha_n|^2}$.

We consider here the situation in which all α_n are set to zero, except for the real part of α_3 (and consequently α_0). The reason for this is that such a quantity reduces to the skewness in the perturbative regime. The imaginary part of α_3 is only meaningful in the non-perturbative regime (and can be set to zero independently without inconsistency; Rocha et al. 2001). Hence we are considering a pdf of the form:

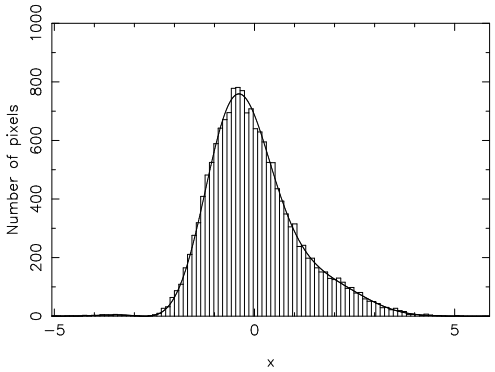


Figure 3. Non-Gaussian pdf given by equation (18) with $\alpha_1 = \alpha_2 = 0$, $\alpha_3 = 0.2$ and $\sigma_0 = 1$.

$$P(x) = \frac{e^{-x^2/(2\sigma_0^2)}}{\sqrt{2\pi}\sigma_0}\left[\alpha_0 + \frac{\alpha_3}{\sqrt{48}} H_3\left(\frac{x}{\sqrt{2}\sigma_0}\right)\right]^2, \qquad (18)$$

with $\alpha_0 = \sqrt{1-\alpha_3^2}$. This allows us to obtain a centred distribution with $\mu_1 = 0$, where μ_n is the nth moment around the origin defined as:

$$\mu_n = \langle x^n \rangle = \int_{-\infty}^{\infty} x^n P(x) \mathrm{d}x. \qquad (19)$$

The first, second and third moments of our pdf are related to α_3 and σ_0 as follows (Contaldi & Magueijo 2001):

$$\begin{aligned} \mu_1 &= 0, \\ \mu_2 &= \sigma_0^2\left(1 + 6\alpha_3^2\right), \\ \mu_3 &= \left(2\sigma_0^2\right)^{\frac{3}{2}} \sqrt{3\left[\alpha_3^2\left(1-\alpha_3^2\right)\right]}. \end{aligned} \qquad (20)$$

Therefore we have generated a centred distribution with a fixed variance and skewness. For our purposes here we considered the distribution with $\alpha_3 = 0.2$ and $\sigma_0 = 1$ plotted in Fig. 3.

The space of possible distribution functions is constrained due to restricting the set of parameters to two parameters only. This implies that we cannot generate distributions with any given variance and skewness. However, in general our method can generate higher values of the relative skewness (since it can generate any distribution) but for that purpose one needs more parameters α_n (Contaldi & Magueijo 2001).

The maps simulated by drawing the pixel values from this pdf will be, by construction, statistically isotropic. They consist of non-Gaussian white noise with variance given by μ_2. We then Fourier transform the map $\Delta T(\hat{\boldsymbol{x}}_p)$ to get the Fourier modes, $a(\boldsymbol{u})$, which have variance proportional to

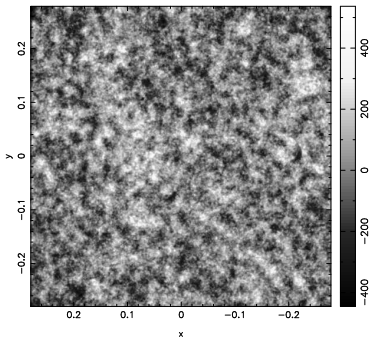


Figure 4. A Non-Gaussian simulation of the sky in real space obtained using the procedure described in Sec. 4.2.2

μ_2. We rescale these Fourier coefficients so that the variance is given by the correct angular power spectrum C_ℓ. We then inverse Fourier transform these coefficients back to real space to produce a new signal map, $\Delta T(\hat{\boldsymbol{x}}_p)$, which now has the appropriate covariance matrix. In Fig. 4 we plot one of these non-Gaussian maps. This new map can then be used as input to simulate the VSA observational strategy and to obtain a set of visibilities as observed by the VSA. For our purposes we are interested in the simplest case, i.e. with no beam convolution and no noise, though these can easily be incorporated if we wish. We output the visibilities on a square grid, and then compute the bispectrum of these simulated visibilities with the same code that we apply to the real data.

By construction, the power spectrum of the non-Gaussian map is C_ℓ and the bispectrum is related to the skewness and variance of our non-Gaussian pdf by:

$$b_{\ell_1\ell_2\ell_3} = \frac{\mu_3}{\mu_2^{3/2}} A_{\rm pix}^{1/2} \sqrt{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \qquad (21)$$

where $A_{\rm pix}$ is the pixel area given by $A_{\rm pix} = L^2/N_{\rm pix}$, for a small patch of the sky of area L^2. For a detailed calculation see Rocha et al. (in preparation). In Fig. 5 we plot the computed values of the bispectrum against the predicted ones. The agreement shows that our bispectrum calculations are indeed correctly obtained.

We note that from the generalised Bayesian analysis using the non-Gaussian likelihood developed by Rocha et al. (2001), and applied to the VSA data by Savage et al. (2004), we concluded that VSA data are mostly consistent with zero α_3, resulting in no evidence for this type of non-Gaussianity.

4.3 Testing for non-Gaussianity

Fig. 6 shows the estimates of $B(\ell, \ell, \ell)$, together with error estimates from Gaussian simulations, from the observations of the three compact array fields and the three extended array fields in the VSA1 region. Little correlation can be seen between the three datasets in each graph. At low values of ℓ noisy estimates are obtained due to the fact that few triangles can be found at low ℓ. The estimated variance of the mean is not dissimilar to the mean itself, indicating that we cannot expect to find a significant deviation from zero. The bispectrum estimates from the extended array are much less noisy than those from the compact array, reflecting the fact that the noise on the visibilities is smaller for the extended

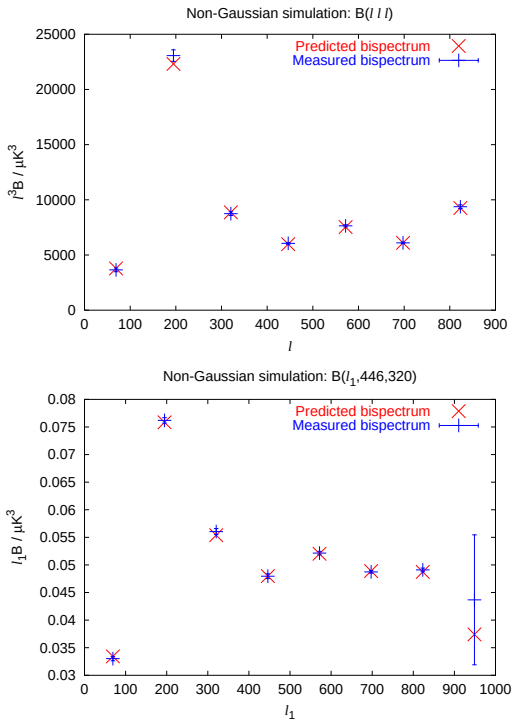


Figure 5. Bispectrum estimates of the non-Gaussian simulations with $\alpha_1 = \alpha_2 = 0$, $\alpha_3 = 0.2$ and $\sigma_0 = 1$ (i.e. with $\mu_3 \sim 0.95$ for a patch of the sky of side $28.^\circ6$ and pixel number $N_{\rm pix} = 128^2$), as compared to the predicted bispectrum obtained using equation (21). The error on the measured bispectrum are due to the finite number of simulations (1865).

array than for the compact array (the mean noise per binned visibility is 4.4–8.8 Jy for the compact array and 0.7–1.1 Jy for the extended array). The noise contributes as $\sigma_{\boldsymbol{u}}^3$ to the bispectrum error.

Figs 7 and 8 show a sample of the bispectrum values computed from the real data, together with the 1 and 2-σ values from the Gaussian simulations. The change in the variance with multipole is due mainly to the variation in the number of triangles of vectors used to form each bispectrum estimate. In addition, the noise on each binned visibility value varies as a result of the differing numbers of raw visibility measurements used to form each binned value. The rough estimates of the errors from equation (15) agree well with the 1-σ values from the simulations for large values of ℓ, tending to be too small by approximately 0–4 per cent. However, for small values of ℓ, equation (15) generally underestimates the variance by a greater amount, the most extreme example being for $B(366, 366, 366)$ where the estimated variance is only 35 per cent of the value obtained from simulations for the case of the VSA1F field. The distribution of this particular bispectrum value is very non-Gaussian, as shown in Fig. 9, and hence it appears that the correlations that we have neglected in equation (15) are significant in this case. Although around 160 different triangles of vectors in the uv plane make up the bispectrum estimate, the effective number of independent triangles is much smaller.

For the 17 fields, there are a total of 1839 estimated bispectrum values. Of these, 99 were found to be greater

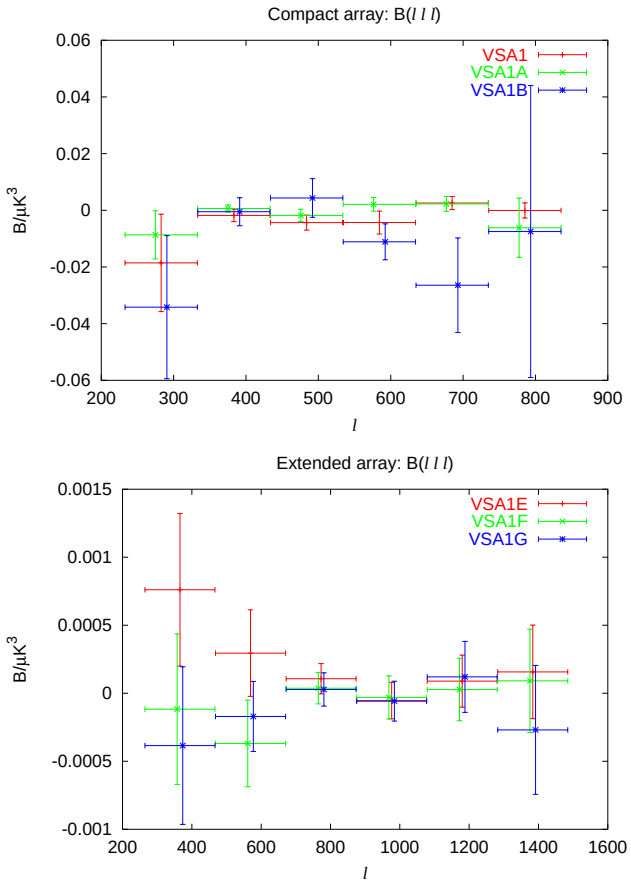


Figure 6. Estimated diagonal component of bispectrum from the compact (top) and extended (bottom) arrays for the region VSA1, with error bars from Gaussian simulations. Some values with very large errors at high and low ℓ have been omitted.

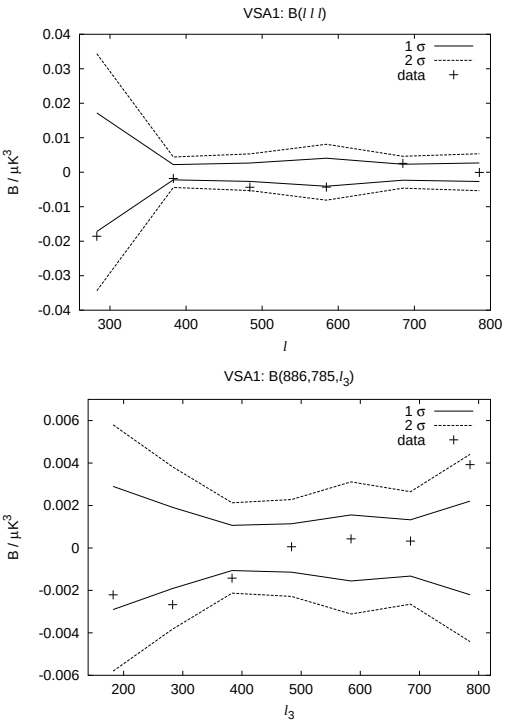


Figure 7. Results from simulations for VSA1 showing variance of bispectrum estimates and data values. The top figure shows the diagonal component of the bispectrum.

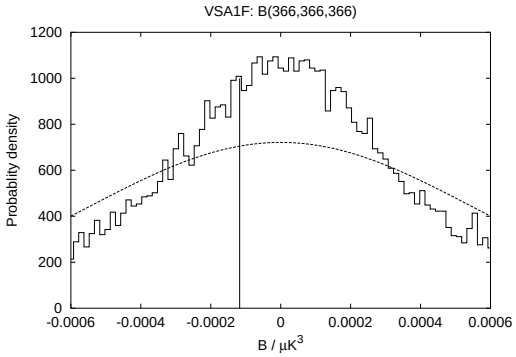


Figure 9. Distribution of the bispectrum estimate $B(366, 366, 366)$ for the field VSA1F, and comparison with a Gaussian distribution with the same variance.

than 2σ in magnitude, compared with the expected value of $\sim$84 if the bispectrum estimates were Gaussian distributed (which we have found to be the case except for when one of the ℓs is small). The VSA3 field had a significant number of large bispectrum estimates (18 out of 176 of modulus $> 2\sigma$, 6 of which were $> 3\sigma$), most of the largest of which appear to have in common at least one value of ℓ=584. There was one bispectrum estimate outside the 4-σ limit, in the field VSA2G. In order to proceed further, we need a way of testing the data as a whole. In the following section we describe one such test that we have applied to the individual fields, as well as to sets of bispectrum estimates formed from the weighted sum of the estimates for individual fields that are in the same region of the sky.

4.3.1 Kolmogorov-Smirnoff test

We wish to perform a quantitative comparison between our data and the simulations, in order to test the null hypothesis, H_0, that the data values were drawn from the same distribution as the Gaussian simulations. Following the method in Komatsu et al. (2002), for each bispectrum estimate $\hat{B}_\alpha$ we compute the quantity

$$P_\alpha = \frac{N\left(\left|\widehat{B}_\alpha^{\rm SIM}\right| < \left|\widehat{B}_\alpha^{\rm VSA}\right|\right)}{N_{\rm Total}}, \qquad (22)$$

where α represents a particular set of $\{\ell_1, \ell_2, \ell_3\}$ and $N_{\rm Total} = 15000$, the total number of simulations. This gives the probability that the magnitude of the bispectrum estimate drawn under H_0 is less than the magnitude of the bispectrum estimate obtained from the data. If H_0 is true, then the distribution of P_α is uniform on the interval [0,1]. If there is a tendency towards a non-zero bispectrum we could (naïvely) expect to obtain more values of P_α close to 1. Fig. 10 shows the resulting cumulative distributions of P_α obtained for the VSA1 observations, in comparison with the straight line expected from a uniform distribution.

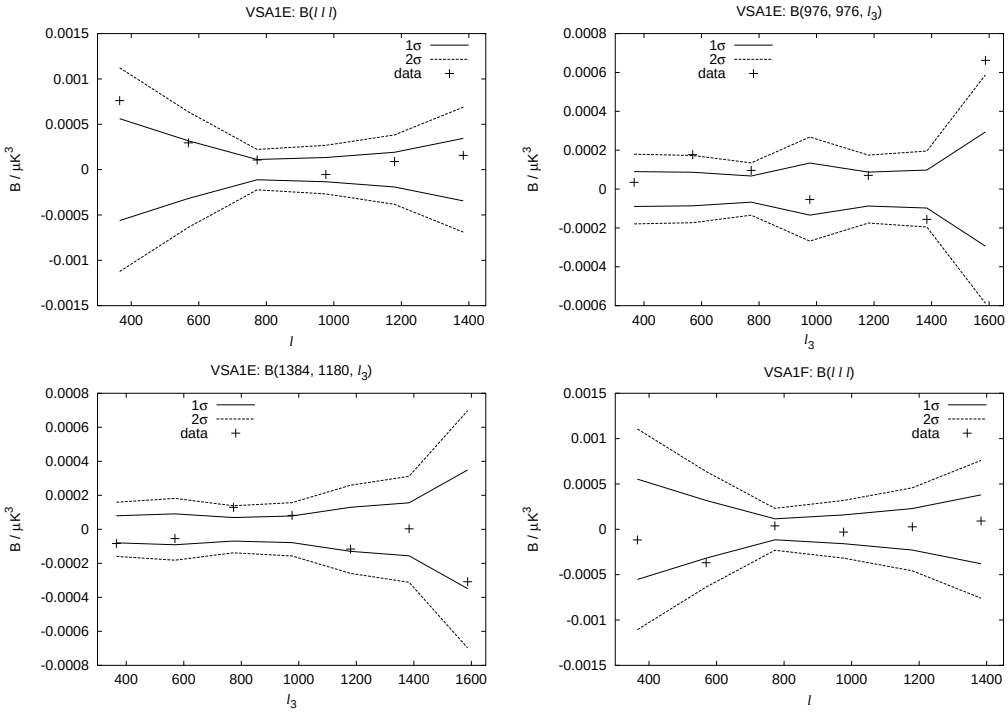


Figure 8. Bispectrum estimates with variance from simulations for the extended array fields VSA1E and VSA1F.

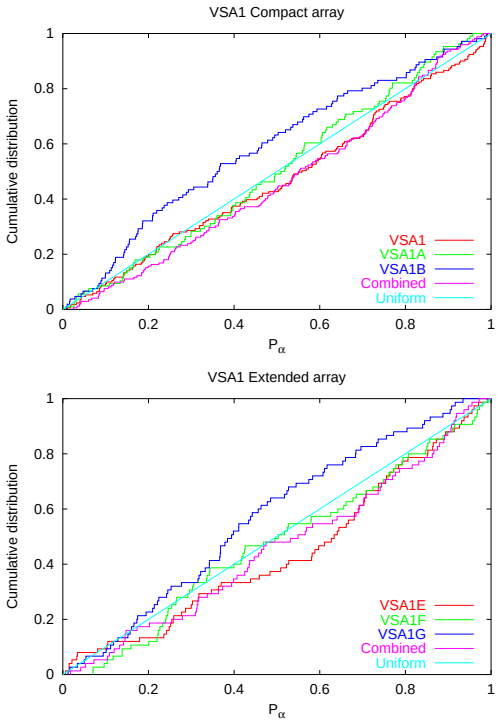


Figure 10. Cumulative distribution of P_α for the VSA1 compact and extended arrays.

The Kolmogorov-Smirnoff test operates by finding the quantity

$$D = \max |F_e(P) - F(P)|, \tag{23}$$

where F_e is the empirical distribution function defined by

$$F_e(P) = \frac{N(P_\alpha < P)}{N}, \tag{24}$$

and $F(P)$ is the continuous cumulative distribution function for P under H_0. If the P_α are independent, identically distributed under H_0 then the distribution of D is universal, depending only on the number of values, N.

Given $D = d$, one can then compute the tail probability p which is given by

$$p(d) = \text{Prob}(D \geq d \mid H_0). \tag{25}$$

In our case, we must calculate $p(d)$ using the distribution of D obtained from simulations, since the assumption that the P_α are independent is invalid: although the formal covariance between different bispectrum estimates is zero apart for the effects of the primary beam, the same visibility point is used to calculate many different bispectrum estimates. Fig. 11 shows the distribution of D obtained from simulations, compared with the distribution which would be obtained if the P_α were independently drawn from a uniform distribution. It can be seen that there is an increase of larger values of D. A very similar distribution is obtained from noise-only simulations, when there is no correlation between any of the visibility points.

Table 1 shows the calibrated results. It should be recognised that this test gives us an indication of how well the simulations match the data, which is dependent on a number of factors other than simply whether the data are Gaussian: whether the power spectrum is accurate and the correct telescope parameters have been used, and whether the noise estimate is correct. For comparison, when we compute the values of P_α by comparing the VSA1F data with Gaussian simulations which have a beam of FWHM $4\overset{\circ}{.}6$ rather than $2\overset{\circ}{.}05$, we obtain $p(d) = 0.006$. Altering the amplitude of the power spectrum by ∼10 per cent was found to change only

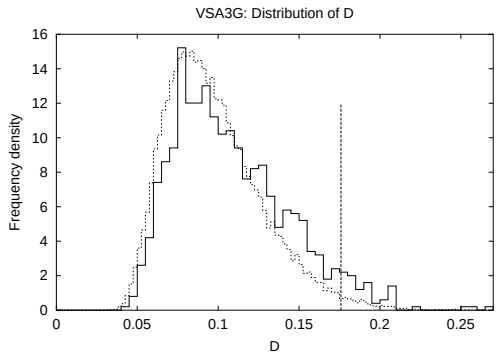


Figure 11. The distribution of D from simulations (solid line) as compared with the distribution obtained if the P_α were independent (dotted line), together with the measured value from the data for the field VSA3G.

Table 1. d-statistic and corresponding calibrated tail probability for the various fields. The number of bispectrum estimates for each field is n. The field labelled 'VSA1 Compact' represents the weighted average bispectrum from the individual VSA1 compact fields, and likewise for the others.

Field	n	d	$p(d)$
VSA1	171	0.076	0.38
VSA1A	106	0.064	0.81
VSA1B	106	0.16	0.03
VSA1 Compact	172	0.086	0.25
VSA1E	75	0.17	0.06
VSA1F	75	0.10	0.52
VSA1G	75	0.15	0.11
VSA1 Extended	75	0.11	0.34
VSA2	173	0.093	0.23
VSA2-OFF	174	0.084	0.29
VSA2 Compact	174	0.069	0.46
VSA2E	76	0.096	0.56
VSA2F	76	0.070	0.87
VSA2G	95	0.107	0.43
VSA2 Extended	95	0.103	0.43
VSA3	176	0.118	0.05
VSA3A	106	0.068	0.76
VSA3B	106	0.050	0.97
VSA3 Compact	176	0.084	0.23
VSA3E	97	0.107	0.45
VSA3F	76	0.090	0.62
VSA3G	76	0.176	0.05
VSA3 Extended	97	0.111	0.39

slightly the distribution of d, acting to alter $p(d)$ by $\sim$0.01. Four out of the 17 individual fields have values of $p(d)$ which are fairly low: VSA1B, VSA1E, VSA3 and VSA3G. For the total number of fields analysed we would expect an average of one to have $p(d) \leq 0.06$ by chance. The lowest value of $p(d)$ is for the VSA1B data, at 3 per cent. This field was observed for a total integration time of only 68 h, in comparison with around 200 hours for the majority of compact array fields (Taylor et al. 2003), and consequently the mean error on the visibilities is approximately twice that of the other fields. Furthermore, the large value of d is caused by an excess of low values of P_α, as can be seen in Fig. 10. This indicates that we should be very hesitant about drawing any conclusions about non-Gaussianity in the VSA1B field. The large value of d for the VSA3G field is also due to an excess of low values of P_α, whereas there is an excess of high values for the VSA3 and VSA1E fields. The VSA3 data have a greater extent in the uv plane than the other two VSA3 compact array fields (hence the high value of n). On eliminating all the bispectrum values which are not calculated for the other two fields, we obtain $n = 106$, $d = 0.149$ and $p(d) = 0.034$, so the significance level is increased slightly, indicating that the large bispectrum values are not concentrated in the region of large ℓ. The VSA2 compact array fields, which have the longest integration times, are consistent with the simulations.

By testing the data in this way, we are making no assumptions about the type of non-Gaussianity that we are looking for. This means that our test is very general, but not very powerful as it is not tailored for optimal detection of any particular non-Gaussian signature. One disadvantage of the test is that all of the P_α values are given equal consideration, regardless of the fact that some bispectrum estimates are more dominated by noise than others. In the following section we consider two tests which are tailored for detecting particular types of non-Gaussianity: that arising from point sources, and from the time of recombination.

4.4 Point sources

4.4.1 Theory

A point source of strength s at position $\boldsymbol{x}_0$ can be described by

$$\Delta I(\hat{\boldsymbol{x}}) = s\,\delta(\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}_0). \tag{26}$$

The contribution this makes to the visibilities is

$$\Delta V(\boldsymbol{u}) = s\,A(\hat{\boldsymbol{x}}_0)\,e^{2\pi i \hat{\boldsymbol{x}}_0 \cdot \boldsymbol{u}}, \tag{27}$$

and therefore the contribution to the bispectrum is

$$\begin{aligned} \Delta V(\boldsymbol{u}_1)\Delta V(\boldsymbol{u}_2)\Delta V(\boldsymbol{u}_3) &= s^3\,A^3(\hat{\boldsymbol{x}}_0)\,e^{2\pi i \hat{\boldsymbol{x}}_0 \cdot (\boldsymbol{u}_1+\boldsymbol{u}_2+\boldsymbol{u}_3)} \\ &= s^3\,A^3(\hat{\boldsymbol{x}}_0). \end{aligned} \tag{28}$$

There will also be cross terms such as $V(\boldsymbol{u}_1)V(\boldsymbol{u}_2)\Delta V(\boldsymbol{u}_3)$ but these will have a mean value of zero. In addition, if we have several point sources there will be additional cross terms introduced. However, on average the contribution to the reduced bispectrum arising from point sources should be uniform (Komatsu & Spergel 2001):

$$B^{\rm ps}(\ell_1, \ell_2, \ell_3) = B^{\rm ps}. \tag{29}$$

So we can form an estimate of the point source contribution simply by calculating the (weighted) mean value of the bispectrum.

4.4.2 Point sources in the VSA data

We compared the estimated value of $B^{\rm ps}$ from the extended array data before and after source subtraction (Taylor et al. 2003; Grainge et al. 2003). The results are shown in Table 2, and illustrated in Fig. 12. The mean value of the bispectrum is altered by $\sim$1–6$\times10^{-6}\mu$K^3 when the point sources are subtracted. Therefore, although subtracting the sources does change the bispectrum estimates, the resulting difference is considerably smaller than the standard deviation on

Table 2. Mean bispectrum values before and after point source subtraction for the three fields in the VSA1 region observed with the extended array.

Field	$\widehat{B}^{\rm ps}/(10^{-5}\mu{\rm K}^3)$ Raw data	$\widehat{B}^{\rm ps}/(10^{-5}\mu{\rm K}^3)$ Sources subtracted	$\sigma/(10^{-5}\mu{\rm K}^3)$
VSA1E	1.19	0.92	1.19
VSA1F	-0.96	-1.61	1.30
VSA1G	1.01	1.09	1.34

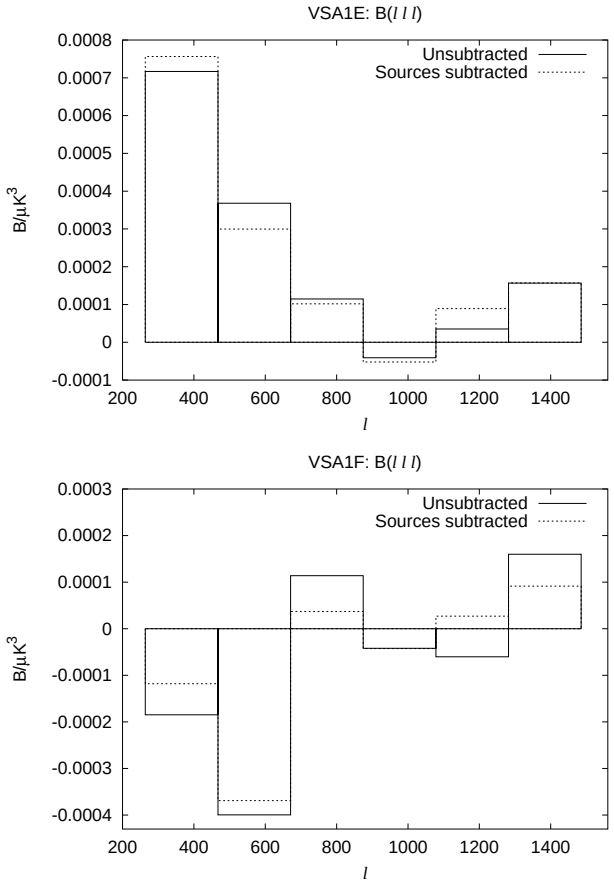


Figure 12. Diagonal component of the bispectrum for the fields VSA1E and VSA1F before and after point source subtraction, illustrating the level of change.

the mean value of the bispectrum which arises from sample variance and noise, and so it does not seem possible to detect point sources in this way. However, if future VSA observations have a much lower noise and probe higher values of ℓ then it may be possible to use the bispectrum to detect the presence of point sources.

4.4.3 Residual point sources

We can estimate the contribution to the bispectrum from residual point sources using the results of the 15-GHz Ryle survey (Waldram et al. 2003). This survey found that the differential source count at 15 GHz could be well approximated by

$$n(S) \approx K \left(\frac{S}{\rm Jy}\right)^{-\beta} {\rm Jy}^{-1}{\rm sr}^{-1}, \qquad (30)$$

where S is the flux density, $\beta = 2.15$ and $K = 51$. Assuming a spectral index $\alpha = 0.55$ we find that, at 34 GHz, $K = 30$.

An individual point source contributes to the three-point visibility function as described by equation (28). Assuming a Poisson distribution of sources, the mean contribution from unsubtracted sources can be expressed as

$$\begin{aligned}\langle V(\boldsymbol{u}_1)V(\boldsymbol{u}_2)V(\boldsymbol{u}_3)\rangle &= \int A^3(\hat{\boldsymbol{x}})\, d^2\hat{\boldsymbol{x}} \int_0^{S_*} S^3 n(S)\, dS \\ &= \frac{2}{3}\pi\sigma^2 K \frac{S_*^{4-\beta}}{4-\beta}, \qquad (31)\end{aligned}$$

for $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = 0$, where S_* is the source subtraction level.

For the compact array, taking $S_* = 80$ mJy we obtain a theoretical contribution from unsubtracted sources of $\sim 4 \times 10^{-4}{\rm Jy}^3$ (equivalent to $4 \times 10^{-6}\mu{\rm K}^3$) and for the extended array, taking $S_* = 20$ mJy we obtain $\sim 6 \times 10^{-6}{\rm Jy}^3$ (equivalent to $3 \times 10^{-7}\mu{\rm K}^3$). Comparing this with the error on the mean value of the bispectrum of $\sim 10^{-4}\mu{\rm K}^3$ for the compact array and $\sim 10^{-5}\mu{\rm K}^3$ for the extended array, we see that we do not expect the bispectrum to detect the presence of the residual sources.

If we take these values and consider the comparison with the predicted bispectrum arising from the coupling between the Sunyaev-Zel'dovich effect and weak lensing effects, as given by Komatsu & Spergel (2001), we find that the SZ-lensing bispectrum will be overwhelmed by residual point sources at the source subtraction level of the extended array for $\ell \gtrsim 200$. At lower values of ℓ we estimate that the SZ-lensing contribution to the bispectrum is approximately four orders of magnitude smaller than the error on our bispectrum estimates. Therefore our data is not sensitive to this effect.

4.5 Simulated point sources

4.5.1 Noise-only simulations

As a preliminary test we subtracted a single point source from an extended-array simulation with no noise or CMB signal. The resulting mean bispectrum[3] was 96 per cent of the theoretical value of $s^3A^3(\hat{\boldsymbol{x}}_0)$, with all the individual bispectrum estimates being very similar in value. The discrepancy is due to the fact that we slightly relax the requirement that $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = 0$ and so the phase in equation (28) is slightly non-zero.

On subtracting two point sources (of flux density 1.0 Jy and 0.89 Jy after attenuation) from the same simulation, we immediately find that each individual bispectrum estimate is very different, varying from -0.048 Jy^3 to -2.4 Jy^3, as a result of the cross-terms. The mean bispectrum value is -1.3 Jy^3 compared to a theoretical value of -1.7 Jy^3. On subtracting the VSA1E source list, 10 sources in total, the mean bispectrum value is $-2.95\times10^{-5}{\rm Jy}^3$ in comparison with the theoretical value of $-3.53\times10^{-5}{\rm Jy}^3$, and there is less

[3] By 'bispectrum' in this section we mean the value of the three-point function as given by equation (12) since it is natural to work in units of Jy when considering point sources.

Table 3. The rms noise, mean bispectrum values and standard deviations for noise-only simulations with point sources added.

rms noise / Jy	$\langle B\rangle/10^{-5}\mathrm{Jy}^3$ $[/10^{-5}\mu\mathrm{K}^3]$	$\sigma/10^{-5}\mathrm{Jy}^3$ $[/10^{-5}\mu\mathrm{K}^3]$
2.2	8.8 [4.4]	390 [19]
0.44	3.59 [0.18]	3.1 [0.15]
0.22	1.89 [0.093]	0.39 [0.019]
0.044	2.13 [0.11]	3.1×10^{-3} $[1.5\times10^{-4}]$
0.0044	2.22 [0.11]	3.1×10^{-6} $[1.5\times10^{-7}]$

variation between individual bispectrum estimates, suggesting that with more sources the cross-terms have a greater tendency to cancel.

In order to ascertain the effect of noise on our ability to detect point sources we used sets of simulated data with no CMB component, just noise, to which we added the point sources subtracted from the VSA1E field. Each simulation had exactly the same *uv* and noise template apart from an overall scaling factor on the noises. The values of the mean bispectra, together with the standard deviation of the mean bispectrum obtained from noise-only simulations, are shown in Table 3. The mean value of the bispectrum is positive in all cases, but a definite detection of point sources is obtained only when the rms noise is less than 0.44 Jy. The point source bispectrum is completely swamped by noise for the case with the highest noise. Fig. 13 illustrates the bispectrum components for each case.

There were 10 sources in the simulations in total, and the sum $\sum_{i=1}^{10} A^3(\hat{\boldsymbol{x}}_i)s_i^3$ is $3.5\times10^{-5}\mathrm{Jy}^3$. This is 1.6 times the mean bispectrum value calculated, weighting according to the variance from simulations. However, if we use an alternative weighting scheme, according to the number of triangles of vectors which are combined to form each bispectrum estimate, we obtain (for the lowest noise case) a mean bispectrum value of $2.6\times10^{-5}\mathrm{Jy}^3$. The cross-terms introduce a lot of variation between the different individual bispectrum estimates and hence cause the final values to depend on the weighting scheme.

The rms noise on each visibility point for the VSA1E data is ∼1 Jy, which is at a level at which the contribution from point sources is swamped by noise. In addition, the CMB itself will increase further the variance on the mean value of the bispectrum. Therefore we cannot expect that we will detect the presence of sources in the extended array data using this method.

For comparison, we also performed the Kolmogorov-Smirnoff test on the point source simulations. This gave a conclusive detection of non-Gaussianity for the cases with a rms of 0.0044 Jy and 0.044 Jy but no detection when the rms noise was 0.22 Jy [the value of $p(d)$ was 0.42] despite the conclusive detection found by computing the mean value of the bispectrum. This illustrates how using a tailored statistic when searching for a particular non-Gaussian signal is a much more powerful method of detection than using a general statistic.

4.5.2 CMB simulation with strong sources

We computed the mean bispectrum from a simulation based on the noise and *uv*-positions of an extended-array field, with both CMB and bright sources. The observed sky is shown in Fig. 14 with the point sources clearly visible in the map. The mean bispectrum value was 0.3 Jy^3, a factor of 1000 greater than the standard deviation from simulations with no point sources. The bispectrum is sensitive to the point sources as (source strength)^3 and so is useful for detecting strong sources but not for weak sources.

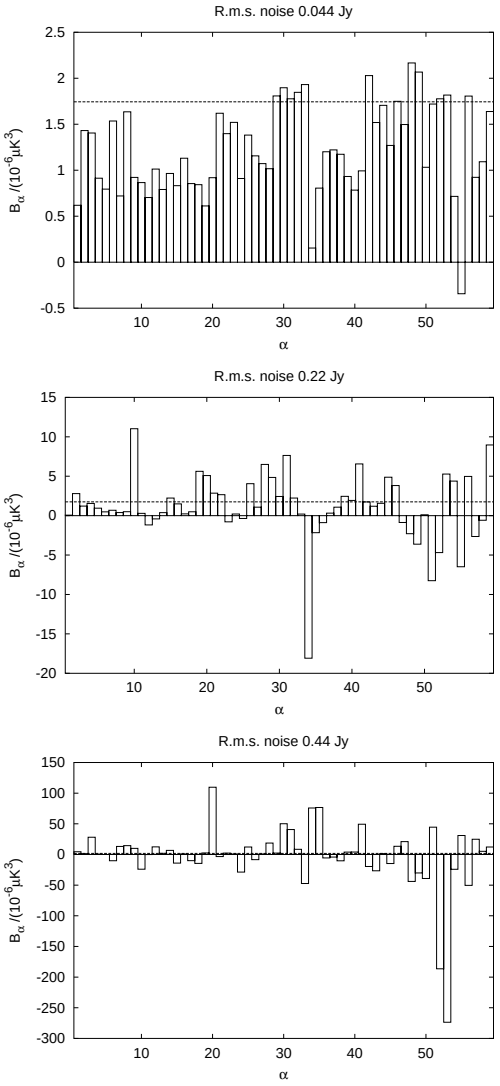


Figure 13. Bispectrum estimates from noise-only simulations with point sources added, with theoretical bispectrum plotted as a horizontal line. α represents a particular set of $\{\ell_1, \ell_2, \ell_3\}$

4.6 Primordial Non-Gaussianity

It has become usual to quantify the primordial non-Gaussianity which arises from weak non-linear evolution on super-Hubble scales by a single parameter f_{NL}, the non-linear coupling parameter:

$$\mathcal{R}(\boldsymbol{x}) = \mathcal{R}_G(\boldsymbol{x}) + f_{\mathrm{NL}}(\mathcal{R}_G^2(\boldsymbol{x}) - \langle\mathcal{R}_G^2(\boldsymbol{x})\rangle), \tag{32}$$

where $\mathcal{R}(\boldsymbol{x})$ is the comoving curvature perturbation as defined according to the convention in Liddle & Lyth (2000),

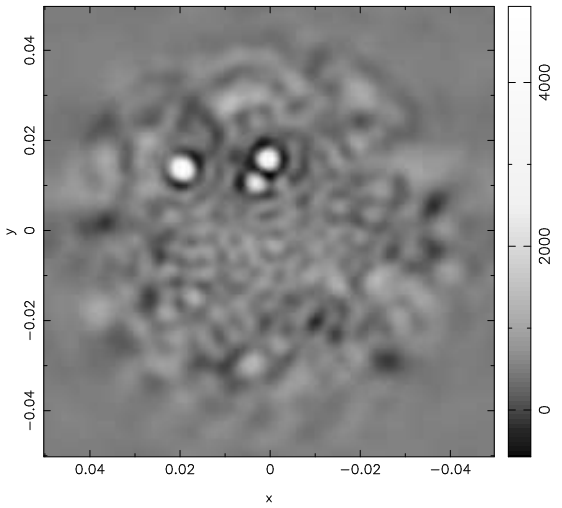


Figure 14. Simulated CMB with point sources, convolved with the primary beam.

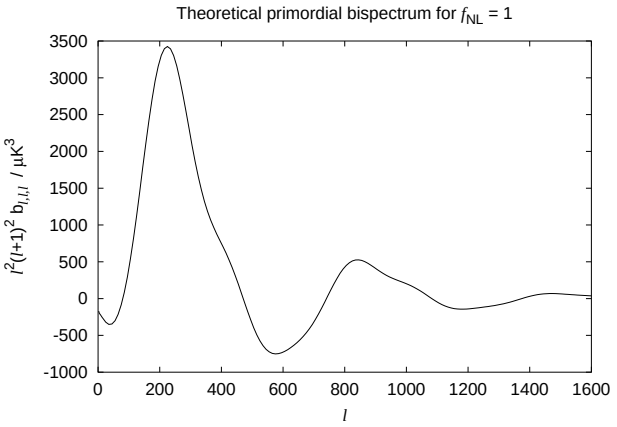


Figure 15. Diagonal component of the theoretical bispectrum for $f_{\rm NL} = 1$.

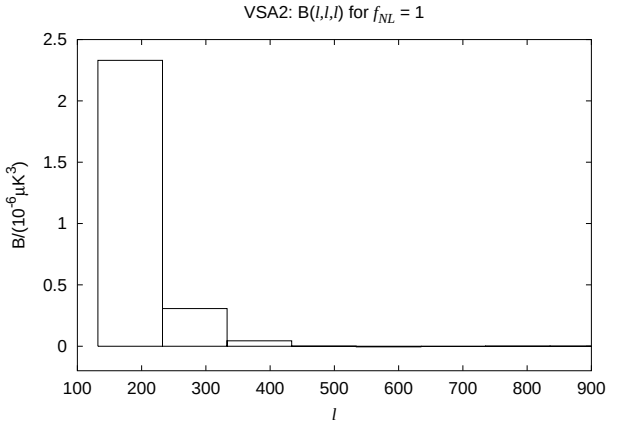


Figure 16. Diagonal component of theoretical measured bispectrum for the VSA2 data with $f_{\rm NL} = 1$

and $\mathcal{R}_G(\boldsymbol{x})$ its Gaussian component[4] (Acquaviva et al. 2003; Maldacena 2003). Even if the fluctuations produced by inflation are perfectly Gaussian, a non-zero $f_{\rm NL}$ will be at second-order in perturbation theory due to the non-linear nature of general relativity (Bartolo et al. 2004). The current distribution of matter, which is highly non-Gaussian, is a result of the non-perturbative non-linear coupling between modes, which becomes increasingly significant after recombination as the perturbations grow. Previous studies have not detected any significant primordial non-Gaussianity, but have simply put upper limits on it. This is not surprising as the predictions from all but the most contrived inflationary models are well below the current upper limits. The recent results from *WMAP* place an upper limit of 89 on the value of $f_{\rm NL}$ (Komatsu et al. 2003), in comparison with predictions from slow-roll inflation which give $|f_{\rm NL}| \sim 10^{-1}$–$10^{-2}$, apart from the non-Gaussianity introduced by the subsequent evolution of the perturbations.

We should therefore only expect to place limits on the value of $f_{\rm NL}$ using the VSA data, but it is interesting to see what limits the data are capable of producing. In addition, the techniques developed to estimate $f_{\rm NL}$ can easily be applied to any other theoretical bispectrum which has its amplitude as the only free parameter.

We calculated the theoretical bispectrum for the case $f_{\rm NL}$=1 using a modified version of CAMB (Lewis, Challinor & Lasenby 2000). The resulting bispectrum has features on the usual acoustic scale of the power spectrum, and its diagonal component is shown in Fig. 15.

In order to relate the theoretical bispectrum to the measured three-point visibility function, we performed the integral described in Appendix C which convolves the bispectrum with the aperture function. For each field, we then estimate the set of Q_α, the theoretical values of the B_α for $f_{\rm NL} = 1$ [where $\alpha = (\ell_1, \ell_2, \ell_3)$] according to equation (13) using the predicted value of $\langle S(\boldsymbol{u}_1)S(\boldsymbol{u}_2)S(\boldsymbol{u}_3)\rangle$ as calculated from the integral. Figure 16 shows the resulting diagonal component of the theoretical measured bispectrum.

We estimate the value of $f_{\rm NL}$ from our bispectrum estimates with the estimator

$$\widehat{f}_{\rm NL} = \frac{\sum_\alpha \frac{\hat{B}_\alpha Q_\alpha}{\sigma_\alpha^2}}{\sum_\beta \frac{Q_\beta^2}{\sigma_\beta^2}}, \tag{33}$$

where the σ_α^2 are the variances of the bispectrum estimates from Gaussian simulations. We obtain these variances from mosaiced simulations to allow us properly to include mosaiced fields. Equation (33) reduces to the optimal linear estimator of $f_{\rm NL}$ from the bispectrum estimates in the limit that these are uncorrelated.

The overall estimate of $f_{\rm NL}$ obtained from the compact array data is 85, with a standard deviation of 2700. For the extended array, we obtain an estimate of -400 with a standard deviation of 3500. The 95 per cent confidence limits are 5400 and 7000 for the compact and extended arrays respectively. These contraints are slightly weaker than those obtained from the *MAXIMA* data using the FFT estimator (Santos et al. 2003), due probably to the fact that

[4] Some authors (Komatsu & Spergel 2001; Santos et al. 2003) use a slightly different definition of $f_{\rm NL}$ based on $\Phi(\boldsymbol{x}) = \frac{2}{3}\mathcal{R}(\boldsymbol{x})$ for adiabatic perturbations in radiation domination. This ignores neutrino anisotropic stress which produces a five per cent correction.

the *MAXIMA* data extend to lower multipoles with slightly higher $\Delta\ell$ resolution, although the VSA sky coverage is slightly greater. [The VSA compact array data cover 101 deg^2 (Taylor et al. 2003); the *MAXIMA* data used had an area of 60 deg^2.] This is also why the compact array data are better at constraining $f_{\rm NL}$ than the extended array data. The bispectrum arising from primordial non-Gaussianity falls quickly with ℓ and so extending the measurements to higher values of ℓ does not significantly improve the constraints (unless the noise is very low). This is why the *WMAP* data, which are cosmic-variance limited at low ℓ, are able to place so much tighter constraints on $f_{\rm NL}$.

5 CONCLUSIONS

Our bispectrum calculations indicate a slightly greater discrepancy with the Gaussian simulations for the fields VSA1B, VSA1E, VSA3, and VSA3G than we would expect if the CMB sky were perfectly Gaussian. However, there is little discernible pattern in the way in which the data deviate from the simulations. A small level of non-Gaussianity is to be expected as there will inevitably be a degree of foreground contamination and unsubtracted point sources, but we have shown that in the case of point sources we do not anticipate to be able to detect the resulting non-Gaussianity with the current level of experimental noise. In the case of point sources, the CMB itself acts as a level of noise that can only be reduced by taking measurements at higher ℓ. We note that Savage et al. (2004) found some evidence for non-Gaussianity in the VSA1 mosaic, which was attributed to point sources or contamination from galactic foregrounds. It is possible that the excess of large bispectrum values in the VSA1E field arises from the same cause. The fact the low values of the tail probability $p(d)$ in a Kolmogorov-Smirnoff test only appear in isolated individual fields indicate that, if the non-Gaussianity which is hinted at is real, it is localised in space.

The limit on the largest scales probed by the VSA means that it can only be used to place weak constraints on the value of the quadratic non-Gaussianity parameter $f_{\rm NL}$. The data are more suited to detecting non-Gaussianity that is present on smaller scales.

ACKNOWLEDGEMENTS

We thank the staff of the Mullard Radio Astronomy Observatory, the Jodrell Bank Observatory and the Teide Observatory for invaluable assistance in the commissioning and operation of the VSA. The VSA is supported by PPARC and the IAC. AC acknowledges a Royal Society University Research Fellowship. GR acknowledges a Leverhulme Fellowship at the University of Cambridge. RSS, SS, NR and KL acknowledge support by PPARC studentships. CD thanks PPARC for funding a post-doctoral research associate position for part of the work. We thank J. Magueijo for useful discussions, and M. Kuntz for valuable comparisons on material presented in Section 4.6.

REFERENCES

Acquaviva V., Bartolo N., Matarrese S., Riotto A., 2003, Nucl. Phys. B667, 119
Bartolo N., Matarrese S., Riotto A., JCAP 01 (2004) 003
Contaldi C. R., Magueijo J., 2001, Phys. Rev. D, 63, 103512
Contaldi C., Bean R., Magueijo J., 1999, Phys. Lett., B468, 189
Dickinson C., et al., submitted to MNRAS, preprint (astro-ph/0402498)
Ferreira P.G., Magueijo J., Górski K.M., 1998, ApJ, 503, L1
Goldberg D. M., Spergel D. N., 1999, Phys. Rev. D, 59, 103002
Grainge K., et al., 2003, MNRAS, 341, L23
Haslam C. G. T., Stoffel H., Salter C. J., Wilson W. E., 1982, A&AS, 47, 1
Heavens A. F., 1998, MNRAS, 299, 805
Hinshaw G., et al., 2003, ApJS, 148, 135
Hobson M. P., Maisinger K., 2002, MNRAS, 334, 569
Hu W., 2000, Phys. Rev. D, 62 043007
Hu W., 2002, Phys. Rev. D, 64 083005
Komatsu E., Spergel D.N., 2001, Phys. Rev. D, 63, 063002
Komatsu E., Wandelt B.D., Spergel D.N., Banday A.J., Gorski K.M., 2002, ApJ, 566, 19
Komatsu E., et al., 2003, ApJS, 148, 119
Lewis A., Challinor A., Lasenby A., 2000, ApJ538, 473
Liddle A.R., Lyth D.H., 2000, Cosmological inflation and large-scale structure, Cambridge University Press
Magueijo J., Medeiros J., submitted to MNRAS, preprint (astro-ph/0311096)
Maldacena J., 2003, JHEP 0305, 013
Mather J.C., et al., 1994, ApJ, 420, 439
Rocha G., Magueijo J., Hobson M. P., Lasenby A., 2001, Phys. Rev. D64, 063512
Rotenberg M., Bivins R., Metropolis N., Wooten Jr. J.K., 1959, The Wigner 3-j and 6-j Symbols, The Technology Press, Massachusetts Institute of Technology, Massachusetts
Santos M. G., et al., 2003, MNRAS, 341, 623
Savage R., et al., 2004, MNRAS, 349, 973
Schlegel D. J., Finkbeiner D. P., Davis M.,1998, ApJ, 500, 525
Scott P. F., et al., 2003, MNRAS, 341, 1076.
Slosar A., et al., 2003, MNRAS, 341, L29
Taylor A. C., et al., 2003, MNRAS, 341, 1066
Waldram E. M., Pooley G. G., Grainge, K. J. B., Jones, M. E., Saunders, R. D. E., Scott, P. F., Taylor, A. C., 2003, MNRAS, 342, 915
Watson R. A, et al., 2003, MNRAS, 341, 1057

APPENDIX A: OPTIMAL CUBIC BISPECTRUM ESTIMATORS

We have estimated the bispectrum using a simple cubic estimator with each visibility weighted by the inverse of its variance (see equation 13). This estimator is not optimal, but for interferometer data, where the signal and noise covariances are nearly diagonal, it is close to optimal as we shall argue in this appendix.

We begin by reviewing the calculation of the optimal cubic estimator, first given by Heavens (1998) and discussed further by Santos et al. (2003). For simplicity, consider real data $\{a_i\}$ whose three-point function is related to the bispectrum B_α that we wish to estimate by

$$\langle a_i a_j a_k \rangle = \sum_\alpha Q^\alpha_{ijk} B_\alpha. \tag{A1}$$

Here, α denotes an (ordered) triplet $\ell_1 \geq \ell_2 \geq \ell_3$ and Q^α_{ijk} is totally symmetric in i, j and k. We construct an estimator $\hat{y}_\alpha$ which is cubic in the $\{a_i\}$,

$$\hat{y}_\alpha = \sum_{ijk} E^\alpha_{ijk} a_i a_j a_k, \tag{A2}$$

and which is unbiased,

$$\langle \hat{y}_\alpha \rangle = \sum_{ijk} E^\alpha_{ijk} \sum_\beta Q^\beta_{ijk} B_\beta \quad \Rightarrow \quad \sum_{ijk} E^\alpha_{ijk} Q^\beta_{ijk} = \delta_{\alpha\beta}. \tag{A3}$$

Only the symmeterised part of E^α_{ijk} enters the estimator $\hat{y}_\alpha$ but we do not impose total symmetry at this stage as enforcing this complicates the variational procedure that follows. Following Heavens (1998), we construct the variance of $\hat{y}_\alpha$ under the assumption that the data are Gaussian (so the estimator is optimised for only weakly non-Gaussian data), to find

$$\mathrm{var}(\hat{y}_\alpha) = \sum_{ijk} \sum_{i'j'k'} E^\alpha_{ijk} E^\alpha_{i'j'k'} (C_{ii'} C_{jj'} C_{kk'} + \mathrm{perms}), \tag{A4}$$

where $C_{ij} \equiv \langle a_i a_j \rangle$ is the (symmetric) covariance of the data. Unlike Heavens (1998), we have not imposed symmetry of E^α_{ijk}, so we cannot simplify equation (A4) in the manner that he does (see his equation 16). We now vary E^α_{ijk} to minimise the variance of $\hat{y}_\alpha$, enforcing the constraints on the mean (equation A3) with a set of Lagrange multipliers $\lambda_{\alpha\beta}$, to find

$$6 \sum_{ijk} E^\alpha_{(ijk)} (2 C_{ii'} C_{jj'} C_{kk'} + 3 C_{ij} C_{k(i'} C_{j'k')}) - \sum_\beta \lambda_{\alpha\beta} Q^\beta_{i'j'k'} = 0. \tag{A5}$$

Note that only the symmetric part $E^\alpha_{(ijk)}$ enters this expression and so only that part is constrained as expected. We can now solve for a symmetric E^α_{ijk} and $\lambda_{\alpha\beta}$ to find

$$E^\alpha_{ijk} = \sum_\beta \lambda_{\alpha\beta} \sum_{i'j'k'} Q^\beta_{i'j'k'} \left(\frac{1}{12} C^{-1}_{ii'} C^{-1}_{jj'} C^{-1}_{kk'} - \frac{1}{4(4+N)} C^{-1}_{i'j'} C^{-1}_{k'(i} C^{-1}_{jk)} \right) \tag{A6}$$

where

$$\lambda^{-1}_{\alpha\beta} = \frac{1}{12} \sum_{ijk} \sum_{i'j'k'} Q^\alpha_{ijk} Q^\beta_{i'j'k'} \left(C^{-1}_{ii'} C^{-1}_{jj'} C^{-1}_{kk'} - \frac{3}{4+N} C^{-1}_{i'j'} C^{-1}_{k'i} C^{-1}_{jk} \right). \tag{A7}$$

In these expressions, N is the number of data points and is typically very large. Our equations (A6) and (A7) correct the results given as equations (21) and (25) in Heavens (1998). The error in the latter analysis arises because symmetry of E^α_{ijk} is assumed prior to the variation, but is then not enforced during it. However, the terms in error in Heavens (1998) are suppressed by $1/N$ for large N, and so the differences from the results derived here are small for $N \gg 1$.

For interferometer data, the covariance matrix C_{ij} is close to diagonal with correlations limited roughly to the extent of the aperture function. Multiplication by C^{-1}_{ij} reduces to inverse variance weighting the visibilities if the effect of the primary beam is neglected. Furthermore, in this limit Q^α_{ijk} enforces $\boldsymbol{u}_i + \boldsymbol{u}_j + \boldsymbol{u}_k = 0$, so that equation (A6), with $N \gg 1$, reduces to the heuristically-weighted estimator of equation (13). In practice, the finite extent of the primary beam makes the heuristic estimator somewhat sub-optimal. The necessary developments to compute the optimal estimator are given in Appendix C, where we construct Q^α_{ijk}. However, given the need to perform large-volume Monte-Carlo simulations to assess properly the significance of our results, employing the optimal estimator would have been prohibitively slow for the analysis in this paper.

APPENDIX B: INTERFEROMETER MEASUREMENTS

The actual values recorded by the VSA are flux density measurements in Janskys. The receivers are calibrated primarily by observations of Jupiter (Taylor et al. 2003), with the primary beam $A(\boldsymbol{x})$ normalised so that $A(0) = 1$. The intensity fluctuations of the CMB can be connected to the temperature fluctuations by

$$\Delta I(\hat{\boldsymbol{x}}, \nu) \approx \left. \frac{\partial B(\nu, T)}{\partial T} \right|_{T=T_0} \Delta T_{\mathrm{cmb}}(\hat{\boldsymbol{x}}). \tag{B1}$$

The visibility seen by the interferometer can be expressed as

$$S(\boldsymbol{u}) = \int \Delta I(\hat{\boldsymbol{x}})\; A(\hat{\boldsymbol{x}})\; e^{2\pi i \boldsymbol{u}.\hat{\boldsymbol{x}}}\; \mathrm{d}^2\hat{\boldsymbol{x}}. \tag{B2}$$

The measured visibility $V(\hat{\boldsymbol{u}})$ is related to this by

$$V(\boldsymbol{u}) = S(\boldsymbol{u}) + N(\boldsymbol{u}), \tag{B3}$$

where $N(\boldsymbol{u})$ is the noise on baseline $\boldsymbol{u}$. We can relate this to the values of $a(\boldsymbol{u})$ by

$$S(\boldsymbol{u}) = \left.\frac{\partial B(\nu,T)}{\partial T}\right|_{T=T_0} T_0\; a(\boldsymbol{u}) \star \tilde{A}(\boldsymbol{u}) \tag{B4}$$

$$= f a(\boldsymbol{u}) \star \tilde{A}(\boldsymbol{u}) \tag{B5}$$

where $f = 94 \times 10^6 \mathrm{Jy\,sr^{-1}}$ for $\nu = 34$ GHz and a CMB temperature T_0 of 2.726K (Mather et al. 1994). If we consider the variance of the measured signal we obtain

$$\begin{aligned}
\langle V(\boldsymbol{u})V^*(\boldsymbol{u})\rangle &= f^2 \int \mathrm{d}^2\boldsymbol{u}_1 \int \mathrm{d}^2\boldsymbol{u}_2\; \tilde{A}(\boldsymbol{u}-\boldsymbol{u}_1)\tilde{A}^*(\boldsymbol{u}-\boldsymbol{u}_2)\,\langle a(\boldsymbol{u}_1)a^*(\boldsymbol{u}_2)\rangle + \sigma_{\boldsymbol{u}}^2 \\
&= f^2 \int \mathrm{d}^2\boldsymbol{u}_1 \int \mathrm{d}^2\boldsymbol{u}_2\; C(u_1)\delta(\boldsymbol{u}_1-\boldsymbol{u}_2)\tilde{A}(\boldsymbol{u}-\boldsymbol{u}_1)\tilde{A}^*(\boldsymbol{u}-\boldsymbol{u}_2) + \sigma_{\boldsymbol{u}}^2 \\
&\approx f^2 C(u) \int \mathrm{d}^2\boldsymbol{u}_1\, |A(\boldsymbol{u}_1)|^2 + \sigma_{\boldsymbol{u}}^2,
\end{aligned}$$

if we make the approximation that $C(u)$ is constant over the width of the aperture function. (We use u_i to denote $|\boldsymbol{u}_i|$.) Here, $\sigma_{\boldsymbol{u}}^2$ is the variance of the noise, which is statistically independent of the visibility measurement.

If $A(\boldsymbol{x}) = \exp(-|\boldsymbol{x}|^2/2\sigma^2)$ then we find that

$$\langle V(\boldsymbol{u})V^*(\boldsymbol{u})\rangle \approx \pi\sigma^2 f^2 C(u) + \sigma_{\boldsymbol{u}}^2. \tag{B6}$$

Similarly, if we consider

$$\begin{aligned}
\langle V(\boldsymbol{u}_1)V(\boldsymbol{u}_2)V(\boldsymbol{u}_3)\rangle &= f^3 \int \mathrm{d}^2\boldsymbol{u}_1' \int \mathrm{d}^2\boldsymbol{u}_2' \int \mathrm{d}^2\boldsymbol{u}_3' \tilde{A}(\boldsymbol{u}_1-\boldsymbol{u}_1')\tilde{A}(\boldsymbol{u}_2-\boldsymbol{u}_2')\tilde{A}(\boldsymbol{u}_3-\boldsymbol{u}_3')B(u_1',u_2',u_3')\delta^2(\boldsymbol{u}_1'+\boldsymbol{u}_2'+\boldsymbol{u}_3') \\
&\approx f^3 B(u_1,u_2,u_3) \int \mathrm{d}^2\hat{\boldsymbol{x}} \int \mathrm{d}^2\boldsymbol{u}_1' \int \mathrm{d}^2\boldsymbol{u}_2' \int \mathrm{d}^2\boldsymbol{u}_3' \\
&\quad \times \exp[2\pi i\hat{\boldsymbol{x}}.(\boldsymbol{u}_1'+\boldsymbol{u}_2'+\boldsymbol{u}_3')]\tilde{A}(\boldsymbol{u}_1-\boldsymbol{u}_1')\tilde{A}(\boldsymbol{u}_2-\boldsymbol{u}_2')\tilde{A}(\boldsymbol{u}_3-\boldsymbol{u}_3') \\
&= f^3 B(u_1,u_2,u_3) \int \mathrm{d}^2\boldsymbol{x} A^3(\hat{\boldsymbol{x}}) \exp[2\pi i\hat{\boldsymbol{x}}.(\boldsymbol{u}_1+\boldsymbol{u}_2+\boldsymbol{u}_3)],
\end{aligned} \tag{B7}$$

assuming that the bispectrum varies only slowly over the width of the aperture function, and that the noise is Gaussian. Taking the above form for $A(\boldsymbol{x})$ and the case $\boldsymbol{u}_1+\boldsymbol{u}_2+\boldsymbol{u}_3 = 0$ we obtain

$$\langle V(\boldsymbol{u}_1)V(\boldsymbol{u}_2)V(\boldsymbol{u}_3)\rangle \approx f^3 \frac{2}{3}\pi\sigma^2 B(u_1,u_2,u_3). \tag{B8}$$

APPENDIX C: VISIBILITY THREE-POINT FUNCTION

In this appendix we derive an integral expression for the three-point function of the interferometer visibilities. For the special case of a Gaussian beam, we are able to reduce the integral further to obtain a closed-form expression for the quantity Q_{ijk}^{α} introduced in Appendix A.

Our starting point is the first equality in equation (B7) which expresses the three-point function of the visibilities as an integral over the bispectrum:

$$\langle S(\boldsymbol{u}_1)S(\boldsymbol{u}_2)S(\boldsymbol{u}_3)\rangle = \int_{-\infty}^{\infty} \mathrm{d}^2\boldsymbol{u}_1' \int_{-\infty}^{\infty} \mathrm{d}^2\boldsymbol{u}_2' \int_{-\infty}^{\infty} \mathrm{d}^2\boldsymbol{u}_3' B(u_1',u_2',u_3')\tilde{A}(\boldsymbol{u}_1-\boldsymbol{u}_1')\tilde{A}(\boldsymbol{u}_2-\boldsymbol{u}_2')\tilde{A}(\boldsymbol{u}_3-\boldsymbol{u}_3')\delta^2(\boldsymbol{u}_1'+\boldsymbol{u}_2'+\boldsymbol{u}_3'). \tag{C1}$$

We can divide the region of integration into six separate regions, $u_1' \geq u_2' \geq u_3'$ and its permutations. Since $B(u_1',u_2',u_3')$ is invariant under permutations of the $\boldsymbol{u}_i'$, as is $\delta^2(\boldsymbol{u}_1'+\boldsymbol{u}_2'+\boldsymbol{u}_3')$, we can permute the $\boldsymbol{u}_i$ instead of the $\boldsymbol{u}_i'$ to find

$$\langle S(\boldsymbol{u}_1)S(\boldsymbol{u}_2)S(\boldsymbol{u}_3)\rangle = \int_{u_1'\geq u_2'\geq u_3'} \mathrm{d}^2\boldsymbol{u}_1'\mathrm{d}^2\boldsymbol{u}_2'\mathrm{d}^2\boldsymbol{u}_3' B(u_1',u_2',u_3') \sum_{\substack{\mathrm{perms} \\ \{\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3\}}} \tilde{A}(\boldsymbol{u}_1-\boldsymbol{u}_1')\tilde{A}(\boldsymbol{u}_2-\boldsymbol{u}_2')\tilde{A}(\boldsymbol{u}_3-\boldsymbol{u}_3')\delta^2(\boldsymbol{u}_1'+\boldsymbol{u}_2'+\boldsymbol{u}_3'). \tag{C2}$$

Transforming to polar coordinates, the delta function restricts the lower bounds of u_2' and u_3' by the triangle inequality, hence

$$\langle S(\boldsymbol{u}_1)S(\boldsymbol{u}_2)S(\boldsymbol{u}_3)\rangle = \int_0^{\infty} u_1'\mathrm{d}u_1' \int_{u_1'/2}^{u_1'} u_2'\mathrm{d}u_2' \int_{u_1'-u_2'}^{u_2'} u_3'\mathrm{d}u_3' B(u_1',u_2',u_3')$$

$$\left(\int_0^{2\pi} \mathrm{d}\phi_1' \int_0^{2\pi} \mathrm{d}\phi_2' \int_0^{2\pi} \mathrm{d}\phi_3' \sum_{\substack{\mathrm{perms} \\ \{\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3\}}} \tilde{A}(\boldsymbol{u}_1 - \boldsymbol{u}_1')\tilde{A}(\boldsymbol{u}_2 - \boldsymbol{u}_2')\tilde{A}(\boldsymbol{u}_3 - \boldsymbol{u}_3')\delta^2(\boldsymbol{u}_1' + \boldsymbol{u}_2' + \boldsymbol{u}_3') \right). \quad \text{(C3)}$$

For every u_1', u_2' and u_3' within the domain of integration, the delta function fixes $\phi_2' = \phi_1' + \psi$, where the two values of $\psi \in [-\pi/2, \pi/2]$ are given by the cosine rule $2u_1'u_2'\cos\psi = u_3'^2 - u_2'^2 - u_1'^2$, and $\phi_3' = \phi_{\boldsymbol{u}_1'+\boldsymbol{u}_2'} + \pi$ closes the triangle. Performing the (trivial) integrations over ϕ_2' and ϕ_3', we find that

$$\begin{aligned} \langle S(\boldsymbol{u}_1)S(\boldsymbol{u}_2)S(\boldsymbol{u}_3)\rangle \quad = \quad & \int_0^\infty u_1'\mathrm{d}u_1' \int_{u_1'/2}^{u_1'} u_2'\mathrm{d}u_2' \int_{u_1'-u_2'}^{u_2'} u_3'\mathrm{d}u_3' \frac{2B(u_1', u_2', u_3')}{\sqrt{4u_1'^2u_2'^2 - (u_3'^2 - u_1'^2 - u_2'^2)^2}} \\ & \left(\int_0^{2\pi} \mathrm{d}\phi_1' \sum_{\substack{\mathrm{perms} \\ \{\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3\}}} \sum_{\pm\psi} \tilde{A}(\boldsymbol{u}_1 - \boldsymbol{u}_1')\tilde{A}(\boldsymbol{u}_2 - \boldsymbol{u}_2')\tilde{A}(\boldsymbol{u}_3 + \boldsymbol{u}_1' + \boldsymbol{u}_2') \right), \end{aligned} \quad \text{(C4)}$$

with ϕ_2' fixed to $\phi_1' + \psi$, and the additional summation is over the two ($\pm$) values of ψ. This result is consistent with equation (C2) of Santos et al. (2003) if we transform their result to Fourier space and replace their weight function $w(\boldsymbol{x})$ by the primary beam of the interferometer.

To make further progress, we specialise to a Gaussian primary beam. As noted in Section 3, this is a good approximation for the VSA and was assumed for all of the analyses in this paper. For a Gaussian beam with dispersion σ, normalised to unity at its peak, the aperture function is $\tilde{A}(\boldsymbol{u}) = 2\pi\sigma^2 \exp(-2\pi^2\sigma^2u^2)$. If we write $\boldsymbol{u}_2' = (u_2'/u_1')\mathbf{R}_\psi\boldsymbol{u}_1'$, where $\mathbf{R}_\psi$ is a right-handed rotation through ψ, then dot products $\boldsymbol{u}_i \cdot \boldsymbol{u}_2'$ can be written as $(u_2'/u_1')(\mathbf{R}_\psi^{-1}\boldsymbol{u}_i) \cdot \boldsymbol{u}_1'$. In this manner, the argument of the exponential arising from the product of the three aperture functions in equation (C4) involves

$$(\boldsymbol{u}_1 - \boldsymbol{u}_1')^2 + (\boldsymbol{u}_2 - \boldsymbol{u}_2')^2 + (\boldsymbol{u}_3 + \boldsymbol{u}_1' + \boldsymbol{u}_2')^2 = \Sigma^2 + \Sigma'^2 - 2\boldsymbol{u}_1' \cdot \boldsymbol{U}, \quad \text{(C5)}$$

where $\Sigma^2 \equiv u_1^2 + u_2^2 + u_3^2$ and $\Sigma'^2 \equiv u_1'^2 + u_2'^2 + u_3'^2$, and the vector $\boldsymbol{U} \equiv \boldsymbol{u}_1 - \boldsymbol{u}_3 + (u_2'/u_1')\mathbf{R}_\psi^{-1}(\boldsymbol{u}_2 - \boldsymbol{u}_3)$. The integration over ϕ_1' can now be performed to give a modified Bessel function $I_0(x)$, and our result for the three-point function reduces to

$$\langle S(\boldsymbol{u}_1)S(\boldsymbol{u}_2)S(\boldsymbol{u}_3)\rangle = \int_0^\infty u_1'\mathrm{d}u_1' \int_{u_1'/2}^{u_1'} u_2'\mathrm{d}u_2' \int_{u_1'-u_2'}^{u_2'} u_3'\mathrm{d}u_3' \frac{2(2\pi)^4\sigma^6 B(u_1', u_2', u_3')e^{-2\pi^2\sigma^2(\Sigma^2+\Sigma'^2)}}{\sqrt{4u_1'^2u_2'^2 - (u_3'^2 - u_1'^2 - u_2'^2)^2}} \sum_{\substack{\mathrm{perms} \\ \{\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3\}}} \sum_{\pm\psi} I_0(4\pi^2\sigma^2u_1'U), \quad \text{(C6)}$$

where $U = |\boldsymbol{U}|$. From this expression, we can easily read off the complex version of Q_{ijk}^α introduced in Appendix A. If we are only interested in triangle configurations, $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = 0$, then it is possible to simplify the product $u_1'U$ further:

$$u_1'U = 3(u_1^2u_1'^2 + u_2^2u_2'^2 + u_3^2u_3'^2) - \frac{1}{2}\Sigma^2\Sigma'^2 \pm \Delta\Delta', \quad (\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = 0), \quad \text{(C7)}$$

where $\Delta \equiv \sqrt{4u_1^2u_2^2 - (u_3^2 - u_1^2 - u_2^2)}/4$ is the area of the triangle formed from $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$, and Δ' is the area of the triangle formed from the primed vectors. The $\pm$ in equation (C7) arises from the two values of ψ. (Whether $\pm$ corresponds to $\pm\psi$ or $\mp\psi$ depends on the orientation of the unprimed triangle, but this is irrelevant for the three-point function since both cases are summed over.)

Finally we note the symmetry properties of the three-point function of the visibilities. For an azimuthally-symmetric primary beam, $\langle S(\boldsymbol{u}_1)S(\boldsymbol{u}_2)S(\boldsymbol{u}_3)\rangle$ is invariant under reflections, rotations and permutations. For triangle configurations, these symmetries ensure that $\langle S(\boldsymbol{u}_1)S(\boldsymbol{u}_2)S(\boldsymbol{u}_3)\rangle$ depends only on the lengths u_1, u_2 and u_3, as is apparent in equation (C7).