ORCA – Online Research @ Cardiff

NON-GAUSSIANITY OF SECONDARY ANISOTROPIES FROM ACTPOL AND PLANCK

William R. Coulton,[1] Simone Aiola,[1] Nicholas Battaglia,[2,3] Erminia Calabrese,[4] Steve K. Choi,[1] Mark J. Devlin,[5]
Patricio A. Gallardo,[6] J. Colin Hill,[3,7] Adam D. Hincks,[8,9] Johannes Hubmayr,[10] John P. Hughes,[11] Arthur Kosowsky,[12,13]
Thibaut Louis,[14] Mathew S. Madhavacheril,[2] Loïc Maurin,[15] Sigurd Naess,[3] Federico Nati,[5] Michael D. Niemack,[6]
Lyman A. Page,[1] Bruce Partridge,[16] Blake D. Sherwin,[17,18] David N. Spergel,[2,3] Suzanne T. Staggs,[1] Alexander Van Engelen,[19]
and Edward J. Wollack[20]

[1]*Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544, USA*

[2]*Department of Astrophysical Sciences, Princeton University,Peyton Hall, Princeton, NJ 08544, USA*

[3]*Center for Computational Astrophysics, Flatiron Institute,162 5th Avenue, 10010, New York, NY, USA*

[4]*School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, CF24 3AA, UK*

[5]*Department of Physics and Astronomy, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA, USA 19104*

[6]*Department of Physics, Cornell University, Ithaca, NY 14853, USA*

[7]*School of Natural Sciences, Institute for Advanced Study, Olden Lane, Princeton, NJ 08540*

[8]*Department of Physics, University of Rome "La Sapienza", Piazzale Aldo Moro 5, I-00185 Rome, Italy*

[9]*Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada V6T 1Z4*

[10]*NIST Quantum Devices Group, 325 Broadway, Mailcode 817.03, Boulder, CO 80305, USA*

[11]*Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ USA 08854-8019*

[12]*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, USA 15260*

[13]*Pittsburgh Particle Physics, Astrophysics, and Cosmology Center, University of Pittsburgh, Pittsburgh PA 15260*

[14]*Laboratoire de l'Accélérateur Linéaire, Univ. Paris-Sud,CNRS/IN2P3, Université Paris-Saclay, Orsay, France*

[15]*Instituto de Astrofísica and Centro de Astro-Ingeniería, Facultad de Física, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile*

[16]*Department of Physics and Astronomy, Haverford College, Haverford, PA, USA 19041*

[17]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA*

[18]*Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720*

[19]*Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, ON, Canada M5S 3H8*

[20]*NASA/Goddard Space Flight Center, Greenbelt, MD, USA 20771*

ABSTRACT

Most secondary sources of cosmic microwave background anisotropy (radio sources, dusty galaxies, thermal Sunyaev Zel'dovich distortions from hot gas, and gravitational lensing) are highly non-Gaussian. Statistics beyond the power spectrum are therefore potentially important sources of information about the physics of these processes. We use data from the Atacama Cosmology Telescope and the Planck satellite to constrain the amplitudes of a set of theoretical bispectrum templates from the thermal Sunyaev-Zeldovich (tSZ) effect, dusty star-forming galaxies (DSFGs), gravitational lensing, and radio galaxies. We make a strong detection of radio galaxies (13σ) and have hints of non-Gaussianity arising from the tSZ effect (3.2σ), DSFGs (3.8σ), from cross-correlations between the tSZ effect and DSFGs (4.2σ) and from cross-correlations among the tSZ effect, DSFGs and radio galaxies (4.6σ). These results suggest that the same halos host radio sources, DSFGs, and have tSZ signal. At current noise levels, strong degeneracies exist between the various sources; upcoming data from Advanced ACT, SPT-3G, Simons Observatory, and CMB-S4 will be able to separate the components. With these caveats, we use the tSZ bispectrum measurement to constrain the amplitude of matter fluctuations to be $\sigma_8 = 0.79 \pm 0.07$.

Corresponding author: William R. Coulton
wcoulton@princeton.edu

1. INTRODUCTION

Secondary sources of cosmic microwave background (CMB) anisotropy occur as photons propagate from the surface of last scattering to us. Lensing distortions arise as the paths of CMB photons are deflected by intervening matter (Blanchard & Schneider 1987). The thermal Sunyaev Zel'dovich (tSZ) effect is induced by the inverse Compton scattering of CMB photons as they travel through ionized gas, distorting the frequency spectrum of the photons (Sunyaev & Zeldovich 1970). The additional kinematic Sunyaev-Zel'dovich (kSZ) effect produces a near-blackbody temperature shift if the ionized gas possesses a bulk velocity along the line of sight (Sunyaev & Zeldovich 1972). Radio galaxies and dusty star-forming galaxies (DSFGs) are the two dominant populations of extragalactic emissive sources at millimeter wavelengths. Radio galaxies have large levels of synchrotron emission, primarily from active galactic nuclei, and in dusty star-forming galaxies the emission arises from thermal dust emission.

Many of these secondary sources have been intensely studied. The power spectrum of the tSZ effect was measured by Planck Collaboration (2016e) and its amplitude on small angular scales has been measured (e.g. Dunkley et al. 2013; George et al. 2015; Planck Collaboration 2016b). The thermal SZ effect in galaxy clusters has been studied through cluster number counts (e.g. Staniszewski et al. 2009; Hasselfield et al. 2013; de Haan et al. 2016; Planck Collaboration 2016g) and through cross-correlations with numerous other tracers (e.g. Planck Collaboration 2016f; Hill & Spergel 2014; Hojjati et al. 2017; Van Waerbeke et al. 2014). The gas properties in these clusters have been studied at low redshifts through X-ray measurements (Arnaud et al. 2010). The dusty star-forming galaxies have been measured through their number counts with *Herschel* and *Spitzer* (e.g. Glenn et al. 2010; Berta et al. 2011b; Béthermin et al. 2010) and power spectrum (Planck Collaboration 2014f) as well as through cross-correlations with lensing and other tracers (e.g. Planck Collaboration 2014d, 2016f; van Engelen et al. 2015). Similarly the radio galaxies' number counts have been measured (de Zotti et al. 2010) and hints of their correlation with the tSZ effect have been seen (Gralla et al. 2014; Gupta et al. 2017). Finally the kSZ effect has been studied through a wide variety of different methods and different data sets (e.g. Hand et al. 2012; De Bernardis et al. 2017; Planck Collaboration 2016k; Soergel et al. 2016; Planck Collaboration 2017; Schaan et al. 2016; Hill et al. 2016). All of these measurements have provided information about the astrophysics of these sources, but the non-Gaussian nature of these sources means that we can gain complementary information by analyzing the higher order correlation functions (e.g. Wilson et al. 2012; Hill & Sherwin 2013; Crawford et al. 2014; Hill et al. 2014; Cooray et al. 2000; Timmons et al. 2017).

Here we focus on the bispectrum from secondary sources. The bispectrum is the harmonic-space transform of the spatial three-point correlation function, which is zero for a purely Gaussian random field. Numerous observed and hypothesized sources give a nonzero bispectrum contribution in the microwave sky. These fall into four categories: primordial sources, secondary sources, second-order gravitational terms and galactic foregrounds. Primordial non-Gaussianity offers another handle on constraining the physics of the early universe and could provide the ability to differentiate between many classes of inflationary theories, though so far it has been found to be consistent with zero (Komatsu et al. 2011; Planck Collaboration 2014e, 2016d); see Liguori et al. (2010); Yadav & Wandelt (2010); Chen (2010) for reviews of primordial non-Gaussianity. Second-order gravitational terms arise from non-linearities in the transfer function from the initial perturbations to the late time temperature fluctuations in CMB (Pettinari et al. 2013; Bartolo et al. 2006, 2007). We do not discuss these here as they are below the sensitivity of our experiments but they will be important for future, more sensitive measurements (Pettinari et al. 2014). Galactic foreground sources include thermal dust emission and synchrotron emission; these are typically most important on the largest scales (degree scales). As we restrict our analysis to small angular scales ($\ell > 200$), these sources of non-Gaussianity are not considered in this work (Gold et al. 2011). We focus on investigating the bispectrum from CMB secondary sources.

The large scale DSFG bispectrum has been measured in Planck Collaboration (2014f) as has the large scale bispectrum from the tSZ effect (Planck Collaboration 2016e) and the bispectrum from correlations between the Integrated Sachs-Wolfe and gravitational lensing (Planck Collaboration 2016d). The skewness of the tSZ effect was examined in Wilson et al. (2012) and Hill & Sherwin (2013) and the one point one point probability density function (PDF) of the tSZ effect has been studied in Hill et al. (2014). The bispectrum from the kSZ effect and galaxies has been probed by pairwise estimators (Hand et al. 2012; De Bernardis et al. 2017; Soergel et al. 2016; Planck Collaboration 2016k), by combining stacked clusters with their reconstructed velocities (Schaan et al. 2016), and through projected field estimators (Hill et al. 2016). Crawford et al. (2014) explored the bispectrum from secondary sources constraining the small-scale tSZ effect, DSFG and Poisson contributions to the bispectrum. Our work builds on aspects of these previous works; in particular we extend these analyses to include the one-halo contributions to the cross-correlations among the tSZ effect, DSFGs and radio galaxies, such as tSZ-tSZ-DSFG and radio-DSFG-tSZ bispectra. Our sensitivity to these terms is enhanced compared to Crawford et al. (2014) due to our inclusion of cross frequency bispectra. By jointly constraining these sources we hope to clearly identify the contributions arising from the differing components.

Bispectrum measurements of secondary sources could provide a wealth of astrophysical information. One halo tSZ bispectrum measurements probe gas pressure profiles on small scales, which provide insight into galaxy cluster processes such as AGN

feedback (Battaglia et al. 2012a). Cross-correlations among the tSZ effect, radio galaxies and DSFGs can disentangle their contributions to the temperature power spectrum measurements, which could lead to a better understanding of the observed deficit of small-scale tSZ power (Planck Collaboration 2016e). The cross-bispectra from the tSZ and optical weak-lensing convergence maps can help constrain hydrostatic mass bias (Nelson et al. 2014; Rasia et al. 2006), a limiting systematic in tSZ cluster cosmology, as well as constraining cosmological parameters (Bhattacharya et al. 2012; Crawford et al. 2014; Wilson et al. 2012). Beyond the tSZ effect, the small-scale DSFG bispectrum can constrain the spatial distribution of DSFGs and the cross-bispectrum from the tSZ and dusty galaxies or radio galaxies probes the masses of haloes these galaxies occupy. The bispectrum provides a flux-weighted count of the total number of these galaxies which can be combined with other measures of galaxy counts to constrain their population properties. DSFGs are a tracer of star formation (Elbaz et al. 2007) and bispectrum measurements can elucidate the interplay between cluster physics and star formation. Bispectrum measurements can help constrain the power spectrum of the kSZ effect (Reichardt et al. 2012), and joint measurements of the kSZ bispectrum and power spectrum can probe cluster thermodynamics and the distribution of baryons (Battaglia et al. 2017; Schaan et al. 2016). Finally, bispectrum measurements can be used to characterize extragalactic biases to CMB lensing power spectra (van Engelen et al. 2014; Osborne et al. 2014).

In this work we analyze microwave intensity data from the Atacama Cosmology Telescope Polarimeter (ACTPol) at 148 GHz over a sky area of $\sim 550 \deg^2$ (Thornton et al. 2016; Louis et al. 2017), combined with *Planck* 100 GHz and 217 GHz intensity data (Planck Collaboration 2014a) in the same region. The ACTPol data have an effective angular resolution of 1.4 arcmin, compared to Planck's effective resolution of 9.7 arcmin (100 GHz) and 5.0 arcmin (217 GHz). We use the Planck 2015 cosmology (Planck Collaboration 2016c).

In section 2 of this paper we review the Komatsu, Spergel and Wandelt (KSW) estimator (Komatsu et al. 2005) and its flat-sky limit and section 3 provides an overview of our secondary CMB anisotropy templates. Sections 4 and 5 briefly describe the data sets used here and the analysis pipeline applied to the data.. We present our results and conclusions in sections 6 and 7. Appendix A derives the correspondence between the full-sky and flat-sky bispectrum estimators, while Appendix B contains a detailed derivation of the bispectrum templates for the major non-Gaussian microwave components. The details of our pipeline validation tests are described in Appendix C. Appendix D contains the calculation of the non-Gaussian contribution to our estimator errors and Appendix E describes how to calculate the DSFG n point functions used in calculating the non-Gaussian errors.

2. BISPECTRUM ESTIMATORS IN THE FLAT-SKY APPROXIMATION

The ACTPol and *Planck* experiments measure fluctuations in the specific intensity, $\Delta I_\nu(\boldsymbol{n})$, of CMB across the sky at a frequency ν. These intensity fluctuations are then related to a temperature fluctuations $\Delta T(\boldsymbol{n})$ via

$$\Delta T(\boldsymbol{n}) = \frac{k_B T^2 c^2 (e^{h\nu/(k_B T)} - 1)^2}{2h^2 \nu^4} \Delta I_\nu(\boldsymbol{n}), \tag{1}$$

where k_B is the Boltzmann's constant, T is the CMB temperature, h is the Planck's constant and c is the speed of light. In this work we focus on small regions of sky, so a flat-sky approximation is valid. In this regime the temperature anisotropies are decomposed as Fourier modes:

$$\frac{\Delta T(\boldsymbol{n})}{T} = \int \frac{d\ell^2}{4\pi^2} a_{\boldsymbol{\ell}} \, e^{i\boldsymbol{n} \cdot \boldsymbol{\ell}}. \tag{2}$$

This is analogous to full-sky analyses where the temperature fluctuations are expanded into spherical harmonics, $Y_{\ell,m}$:

$$\frac{\Delta T(\boldsymbol{n})}{T} = \sum_\ell \sum_{-\ell < m < \ell} Y_{\ell,m}(\boldsymbol{n}) a_{\ell,m}. \tag{3}$$

The flat-sky approximation is accurate to better than 1% for $\ell > 200$ (Loverde & Afshordi 2008). We will focus our analysis on the bispectrum, which is equal to the ensemble average of three $a_{\boldsymbol{\ell}}^X$ (where the superscript X denotes maps from different frequencies or telescope arrays):

$$B^{(X_1, X_2, X_3)}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) = \left\langle a_{\boldsymbol{\ell}_1}^{X_1} a_{\boldsymbol{\ell}_2}^{X_2} a_{\boldsymbol{\ell}_3}^{X_3} \right\rangle. \tag{4}$$

Under the assumption of rotational invariance, the flat-sky bispectrum can be expressed as (Spergel & Goldberg 1999; Hu 2000)

$$B^{X_1, X_2, X_3}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) = 4\pi^2 \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) b_{\ell_1 \ell_2 \ell_3}^{X_1 X_2 X_3}, \tag{5}$$

where $b_{\ell_1 \ell_2 \ell_3}^{X_1 X_2 X_3}$ is the reduced bispectrum. The assumption of rotational invariance is justified as the bispectra considered in this work are from extragalactic sources. The full sky bispectrum has a corresponding form for rotational invariance; see Appendix A for a more detailed discussion of the correspondance.

In the case of weak non-Gaussianity, we can obtain an optimal (minimum-variance) estimator for the amplitude A of a non-Gaussian component with a given reduced bispectrum, $b^{\text{th}}_{\ell_1\ell_2\ell_3}$. In full-sky analyses the optimal estimator can be written as (Babich et al. 2004; Creminelli et al. 2006; Senatore et al. 2010; Verde et al. 2013)

$$\hat{A} = \frac{1}{\mathcal{N}} \sum_{X_i,} \sum_{X_i'} \sum_{\ell_i,m_i} \sum_{\ell_i',m_i'} \mathcal{G}^{\ell_1\ell_2\ell_3}_{m_1m_2m_3} b^{X_1X_2X_3,\,\text{th}}_{\ell_1\ell_2\ell_3} \left\{ \left[(C^{-1}_{\ell_1m_1,\ell_1'm_1'})^{X_1X_1'} a^{X_1'}_{\ell_1'm_1'} (C^{-1}_{\ell_2m_2,\ell_2'm_2'})^{X_2X_2'} a^{X_2'}_{\ell_2'm_2'} (C^{-1}_{\ell_3m_3,\ell_3'm_3'})^{X_3X_3'} a^{X_3'}_{\ell_3'm_3'} \right] \right.$$
$$\left. - 3 \left[(C^{-1}_{\ell_1m_1,\ell_2m_2})^{X_1X_2} (C^{-1}_{\ell_3m_3,\ell_3'm_3'})^{X_3X_3'} a^{X_3'}_{\ell_3'm_3'} + \text{ cyclic} \right] \right\} , \tag{6}$$

where X parameterizes the different sky maps, $(C^{-1})^{X_iX_j}$ is the inverse of the covariance matrix $C^{X_iX_j} = \left\langle a^{X_i}_{\ell_1m_1} a^{X_j}_{\ell_2m_2} \right\rangle$, $\mathcal{N}$ is the normalization and $\mathcal{G}^{\ell_1\ell_2\ell_3}_{m_1m_2m_3}$ is the Gaunt integral over the product of three spherical harmonics (see equation A8 for the Gaunt integral definition). The normalization is chosen such that a unity amplitude is returned for an input map with a reduced bispectrum equal to the theoretical reduced bispectrum. The covariance matrix includes two components:

$$C^{\text{Total}}_{\ell,m,\ell',m'} = \text{w}_{\ell,m}\text{w}_{\ell',m'} C^{\text{Signal}}_{\ell,m,\ell',m'} + \text{N}_{\ell,m,\ell',m'}. \tag{7}$$

The $\text{w}_{\ell,m}$ carries the information of the beam profile and the pixelization function, $C^{\text{Signal}}_{\ell,m,\ell',m'}$ is the covariance of the sky signal and $N_{\ell,m,\ell',m'}$ is the noise covariance.

For the flat-sky limit, we start from the distribution for Gaussian fluctuations:

$$P(a^{\nu_1}, a^{\nu_2}, ...) \propto \exp\left[-\frac{1}{2} \sum_{X_1,X_2} \int \text{d}^2\boldsymbol{\ell}_1 \text{d}^2\boldsymbol{\ell}_2 \, a^{X_1}_{\boldsymbol{\ell}_1} (C^{-1}_{\ell_1,\ell_2})^{X_1X_2} a^{*X_2}_{\boldsymbol{\ell}_2} \right], \tag{8}$$

where the $X_i \in \{\nu_1, \nu_2...\}$, and analogously

$$C^{\text{Total}}_{\ell,\ell'} = \text{w}_\ell \text{w}_{\ell'} C^{\text{Signal}}_{\ell,\ell'} + \text{N}_{\ell,\ell'}, \tag{9}$$

hereafter $C = C^{\text{Total}}$. Assuming that the amount of non-Gaussianity is small, we can then perform an Edgeworth expansion (Taylor & Watts 2001; Juszkiewicz et al. 1995; Bernardeau & Kofman 1995) giving

$$P(a^{\nu_1}, a^{\nu_2}, ...) \propto \left(1 - \frac{1}{6} \sum_{X_i} \int \prod_i \text{d}^2\boldsymbol{\ell}_i \left\langle a^{X_1}_{\boldsymbol{\ell}_1} a^{X_2}_{\boldsymbol{\ell}_2} a^{X_3}_{\boldsymbol{\ell}_3} \right\rangle \frac{\partial}{\partial a^{*X_1}_{\boldsymbol{l}_1}} \frac{\partial}{\partial a^{*X_2}_{\boldsymbol{l}_2}} \frac{\partial}{\partial a^{*X_3}_{\boldsymbol{l}_3}} + ... \right) \exp\left[-\frac{1}{2} \sum_{X_i} \int \text{d}^2\boldsymbol{\ell}_1 \text{d}^2\boldsymbol{\ell}_2 a^{X_1}_{\boldsymbol{\ell}_1} (C^{-1}_{\ell_1,\ell_2})^{X_1X_2} a^{*X_2}_{\boldsymbol{\ell}_2} \right]$$
$$\propto \left(1 - \frac{1}{6} \sum_{X_i} \sum_{X_i'} \int \prod_i \text{d}^2\boldsymbol{\ell}_i \left\langle a^{X_1}_{\boldsymbol{\ell}_1} a^{X_2}_{\boldsymbol{\ell}_2} a^{X_3}_{\boldsymbol{\ell}_3} \right\rangle \left((C^{-1}_{\ell_1,\ell_1'})^{X_1X_1'} a^{X_1'}_{\boldsymbol{\ell}_1'} (C^{-1}_{\ell_2,\ell_2'})^{X_2X_2'} a^{X_2'}_{\boldsymbol{\ell}_3} (C^{-1}_{\ell_3,\ell_3'})^{X_3X_3'} a^{X_3'}_{\boldsymbol{\ell}_3'} \right. \right.$$
$$\left. \left. - 3\big((-C^{-1}_{\ell_1,\ell_2})^{X_1X_2} (C^{-1}_{\ell_3,\ell_3'})^{X_3X_3'} a^{X_3'}_{\boldsymbol{\ell}_3'} + ... \text{ cyclic}\big) \right) \right) \exp\left[-\frac{1}{2} \sum_{X_i} \int \text{d}^2\boldsymbol{\ell}_1 \text{d}^2\boldsymbol{\ell}_2 \, a^{X_1}_{\boldsymbol{\ell}_1} (C^{-1}_{\ell_1,\ell_2})^{X_1X_2} a^{*X_2}_{\boldsymbol{\ell}_2} \right]. \tag{10}$$

Then we impose rotational invariance on the bispectrum (Hu 2000)

$$\left\langle a^{X_1}_{\boldsymbol{\ell}_1} a^{X_2}_{\boldsymbol{\ell}_1} a^{X_3}_{\boldsymbol{\ell}_1} \right\rangle = f_{\text{NL}} 4\pi^2 b^{X_1X_2X_3,\,\text{th}}_{\ell_1\ell_2\ell_3} \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3), \tag{11}$$

where $\ell = |\boldsymbol{\ell}|$ and obtain the optimal estimator as

$$\hat{A} = \frac{1}{\mathcal{N}} \bar{A} = \frac{1}{\mathcal{N}} \sum_{X_i} \sum_{X_i'} \int \prod_i \text{d}^2\boldsymbol{\ell}_i \, 4\pi^2 b^{X_1X_2X_3,\,\text{th}}_{\ell_1\ell_2\ell_3} \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) \left((C^{-1}_{\ell_1,\ell_1'})^{X_1X_1'} a^{X_1'}_{\boldsymbol{\ell}_1} (C^{-1}_{\ell_2,\ell_2'})^{X_2X_2'} a^{X_2'}_{\boldsymbol{\ell}_2} (C^{-1}_{\ell_3,\ell_3'})^{X_3X_3'} a^{X_3'}_{\boldsymbol{\ell}_3} \right.$$
$$\left. - 3\big((C^{-1}_{-\boldsymbol{\ell}_1,\ell_2})^{X_1X_2} (C^{-1}_{\ell_3,\ell_3'})^{X_3X_3'} a^{X_3'}_{\boldsymbol{\ell}_3'} + ... \text{ cyclic}\big) \right), \tag{12}$$

where $\mathcal{N}$ is the normalization term as before, $\bar{A}$ is the unnormalized template amplitude, and $\hat{A}$ is the estimated template amplitude. For simplicity, we use the following notation to denote inverse covariance filtered maps:

$$C^{-1X}_{\boldsymbol{\ell}} a \equiv \sum_{X'} \int \text{d}^2\ell' (C^{-1}_{\ell,\ell'})^{X,X'} a^{X'}_{\boldsymbol{\ell}'}. \tag{13}$$

The last term in equation 12, which depends only on one $a_{\boldsymbol{\ell}}^X$, is known as the linear term. We calculate the linear term and the estimator normalization by ensemble averages, which is described in section 2.2. In the optimal case the variance of this estimator is related to the normalization by:

$$\mathrm{Var}(\hat{A}) = \frac{1}{\mathcal{N}}. \tag{14}$$

In our analysis we apply a real space mask, $M(\boldsymbol{n})$, and this alters this formula to

$$\mathrm{Var}(\hat{A}) \propto \frac{1}{\mathcal{N}} \tag{15}$$

where the proportionality constant depends on $f_{\mathrm{sky}}^{(n)} = 1/(N_{\mathrm{pix}}) \int \mathrm{d}^2\boldsymbol{n}\, M(\boldsymbol{n})^n$ and N_{pix} is the total number of pixels. We will fit multiple templates simultaneously so we will use an extended version of this estimator:

$$\hat{A}_i^{\mathrm{joint}} = \sum_j \mathcal{N}^{-1}{}_{i,j} \bar{A}_j \tag{16}$$

where $\mathcal{N}_{i,j}$ is a generalized normalization constant and the covariance of the estimators is given by:

$$\mathrm{Cov}(\hat{A}^i, \hat{A}^j) \propto \mathcal{N}^{-1}{}_{i,j}, \tag{17}$$

where the proportionality constant depends on $f_{\mathrm{sky}}^{(n)}$. Whilst this estimator is derived in the limit of weak non-Gaussianity, it is unbiased for large non-Gaussianity but not optimal. For large non-Gaussianity there could be additional contributions to the template covariances and these are discussed in Appendix D. Hereafter the covariance matrix in 17 is called the 'Gaussian' covariance and the covariance containing the contributions described in Appendix D is called the 'non-Gaussian' covariance.

2.1. KSW estimator

The Komatsu, Spergel and Wandelt (KSW) estimator (Komatsu et al. 2005) offers the ability to compute the amplitude of separable or nearly separable bispectra in an efficient manner. The technique uses the separability to significantly reduce the computational cost. Consider the following reduced bispectrum:

$$b_{\ell_1,\ell_2,\ell_3} = \mathrm{w}_{\ell_1}\mathrm{w}_{\ell_2}\mathrm{w}_{\ell_3} \int dr \left(\alpha_{\ell_1}(r)\beta_{\ell_2}(r)\gamma_{\ell_3}(r) + \alpha_{\ell_2}(r)\beta_{\ell_3}(r)\gamma_{\ell_1}(r) + 4\,\text{permutations of } \ell_1,\, \ell_2 \text{ and } \ell_3 \right) \tag{18}$$

where w_ℓ is again the window function with pixelation window and the integral over α, β and γ is the theoretical signal bispectrum. Many types of primordial bispectra have reduced bispectra with this structure. We have assumed that we have only one map so that, for simplicity, we can drop the X_i superscripts. Inserting this form into our estimator and rearranging we attain the following:

$$\hat{A} \propto \int dr \int d^2\mathbf{n} \Big(\int d^2\boldsymbol{\ell}_1\, \mathrm{w}_{\ell_1}\alpha_{\ell_1}(r)C_{\ell_1}^{-1}ae^{-i\mathbf{n}\cdot\boldsymbol{\ell}_1} \int d^2\boldsymbol{\ell}_2\, \mathrm{w}_{\ell_2}\beta_{\ell_2}(r)C_{\ell_2}^{-1}ae^{-i\mathbf{n}\cdot\boldsymbol{\ell}_2} \int d^2\boldsymbol{\ell}_3\, \mathrm{w}_{\ell_3}\gamma_{\ell_3}(r)Ca_{\boldsymbol{\ell}_3}^{-1}e^{-i\mathbf{n}\cdot\boldsymbol{\ell}_3} + ...$$

$$+ \left\langle \int d^2\boldsymbol{\ell}_1\, \mathrm{w}_{\ell_1}\alpha_{\ell_1}(r)C_{-\boldsymbol{\ell}_1}^{-1}ae^{-i\mathbf{n}\cdot\boldsymbol{\ell}_1} \int d^2\boldsymbol{\ell}_2\, \mathrm{w}_{\ell_2}\beta_{\ell_2}(r)C_{\ell_2}^{-1}ae^{-i\mathbf{n}\cdot\boldsymbol{\ell}_2} \right\rangle \int d^2\boldsymbol{\ell}_3\, \mathrm{w}_{\ell_3}\gamma_{\ell_3}(r)C_{\ell_3}^{-1}ae^{-i\mathbf{n}\cdot\boldsymbol{\ell}_3} + ... \Big), \tag{19}$$

where $\ell \equiv |\boldsymbol{\ell}|$. Note that we also used the relation $C_{\ell,\ell'}^{-1} = \left\langle C_\ell^{-1}aC_{\ell'}^{-1}a \right\rangle$ to reformulate the linear term. The separation of the integrals allows an efficient calculation of the estimator, and this method can be extended for templates with double integrals as we will describe in the next section.

When applied to real data we discretise the integrals over ℓ's and the resulting sums are efficiently calculated via fast Fourier transforms, for which we use the FFTW package (Frigo et al. 2005).

2.2. Estimator Normalization and Linear Term

We compute the estimator's normalization and linear term using the ensemble average method described in Smith & Zaldarriaga (2011). We find the linear term is negligible for this work and this is discussed below. We outline this method here and refer the reader to the original paper for more details First we generate sets of Gaussian simulations with power spectrum C_ℓ^{Total}. To achieve this we use the CAMB (Lewis et al. 2000) code to generate the underlying CMB power spectrum and apply the beam and pixel window function, as has been presented in Hasselfield et al. (2013) for ACTPol and in Planck Collaboration (2014b) for *Planck*. We add secondary sources to our CAMB power spectrum with amplitudes as determined in Dunkley et al. (2013). For *Planck*,

the noise power is estimated from simulations (Planck Collaboration 2016i). For the ACTPol maps we have four splits of the data with uncorrelated noise, created by using data taken on every fourth night (see also Section 4). From these, we estimate the noise power, assuming it is diagonal, via the following equation:

$$N_{\boldsymbol{\ell}} \approx \frac{1}{4} \sum_i a_{\boldsymbol{\ell}}^i a_{\boldsymbol{\ell}}^{i*} - \frac{1}{12} \sum_{i \neq j} a_{\boldsymbol{\ell}}^i a_{\boldsymbol{\ell}}^{j*}. \tag{20}$$

where $a_{\boldsymbol{\ell}}^i$ is the power in the i$^{\text{th}}$ split of the data. We assume that the noise is uncorrelated between the different arrays. We apply the same masks as used in the data to our simulations, when masking is required. The real noise is anisotropic and we model the anisotropy in our simulations by weighting the noise simulations in real space by the square root of the hits map.

For each Gaussian simulation, we then calculate the following quantity:

$$\nabla_{\boldsymbol{\ell}}^i T(a^X) = \frac{1}{2} \sum_{X_1, X_2} \int \int d^2 \ell_1 d^2 \ell_2 4\pi^2 \delta^{(2)}(\boldsymbol{\ell} + \boldsymbol{\ell}_1 + \boldsymbol{\ell}_2) b^{i,X,X_1,X_2}(\ell_1, \ell_2, \ell_3) a_{\boldsymbol{\ell}_1}^{*X_1} a_{\boldsymbol{\ell}_2}^{*X_2}, \tag{21}$$

where the index i labels the type of non-Gaussianity. We then use this quantity to calculate the normalization:

$$\mathcal{N}_{i,j} \propto \int \int d^2 \ell_1 d^2 \ell_2 \left(\frac{1}{3} \left\langle \nabla_{\boldsymbol{\ell}_1}^i T(C^{-1} a^{X_1}) C_{\boldsymbol{\ell}_1, \boldsymbol{\ell}_2}^{X_1, X_2} (\nabla_{\boldsymbol{\ell}_2}^j T(C^{-1} a^{X_2}))^* \right\rangle - \frac{1}{3} \left\langle \nabla_{\boldsymbol{\ell}_1}^i T(C^{-1} a^{X_1}) \right\rangle C_{\boldsymbol{\ell}_1, \boldsymbol{\ell}_2}^{X_1, X_2} \left\langle (\nabla_{\boldsymbol{\ell}_2}^j T(C^{-1} a^{X_2}))^* \right\rangle \right) \tag{22}$$

and the linear term:

$$\int \prod_i d^2 \boldsymbol{\ell}_i \, 4\pi^2 \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) b^{i,X_1,X_2,X_3}(\ell_1, \ell_2, \ell_3) C^{-1}{}_{-\boldsymbol{\ell}_1, \boldsymbol{\ell}_2}^{X_1, X_2} C_{\boldsymbol{\ell}_3}^{-1}{}^{X_3} a = \int d^2 \ell_3 \left\langle \nabla_{\boldsymbol{\ell}_3}^i T(C^{-1} a_3^X) \right\rangle C_{\boldsymbol{\ell}_3}^{-1}{}^{X_3} a. \tag{23}$$

The normalization proportionality constant depends on $f_{\text{sky}}^{(n)}$. We use this method to calculate the normalization as it was found to be more efficient than a more direct approach.

The linear term of the estimator is only important when the covariance matrix has large off diagonal terms. This can be seen by considering the case when $C_{\boldsymbol{\ell}, \boldsymbol{\ell}'}^{-1 \, X, X'}$ is diagonal ($\boldsymbol{\ell} = \boldsymbol{\ell}'$). In this case the contribution from the linear term vanishes. In our pipeline we take steps to reduce the impact of mode coupling and inhomogeneous noise, which are the main sources of off-diagonal covariance terms. This means that the main contributions to the linear term will be from configurations with one very long leg (small ℓ). As discussed in section 5 we mask the lowest ℓ modes to avoid ground contamination and in doing so we strongly suppress the linear term. For the templates used here we found that the linear term contributed negligibly to our results and that when it was included we required many more simulations for the ensemble averages to converge than if we neglected the linear term. After verifying it was negligible for all the templates we proceeded to neglect its contribution.

3. BISPECTRUM TEMPLATES

In this work we consider non-Gaussianity from lensing-induced anisotropies, the tSZ effect, DSFGs and radio galaxies. The kSZ effect is too small to be constrained with the data used in this work; the form of its bispectrum template is described in Appendix B.6 for completeness. In this section we present a brief overview of the structure of the bispectrum templates that are used in this analysis and provide the details of these calculations in Appendix B.

3.1. *Poisson point source templates*

Unclustered point sources in our maps follow Poisson statistics. Their reduced bispectrum is a flux weighted count of the number of point sources and has the form

$$b_{\ell_1, \ell_2, \ell_3}^{\text{Point Sources}} = k_\nu^3 \int_0^{S_c} dS_\nu \, S_\nu^3 \frac{dn}{dS d\Omega}, \tag{24}$$

where S_ν is the flux density, $dn/dS d\Omega$ is the number count per steradian per flux interval and k_ν is the conversion from flux to map temperature. Radio galaxies and dusty star forming galaxies (DSFGs) both have a Poisson point source contribution.

3.2. *Thermal Sunyaev Zel'dovich, dusty star forming galaxy and radio galaxy templates*

We construct templates for non-Gaussianity from auto and cross correlations from the tSZ effect, DSFG and radio galaxies with the halo model. Our analysis is focused on the smallest scales and so we only include the one halo term. Here we briefly
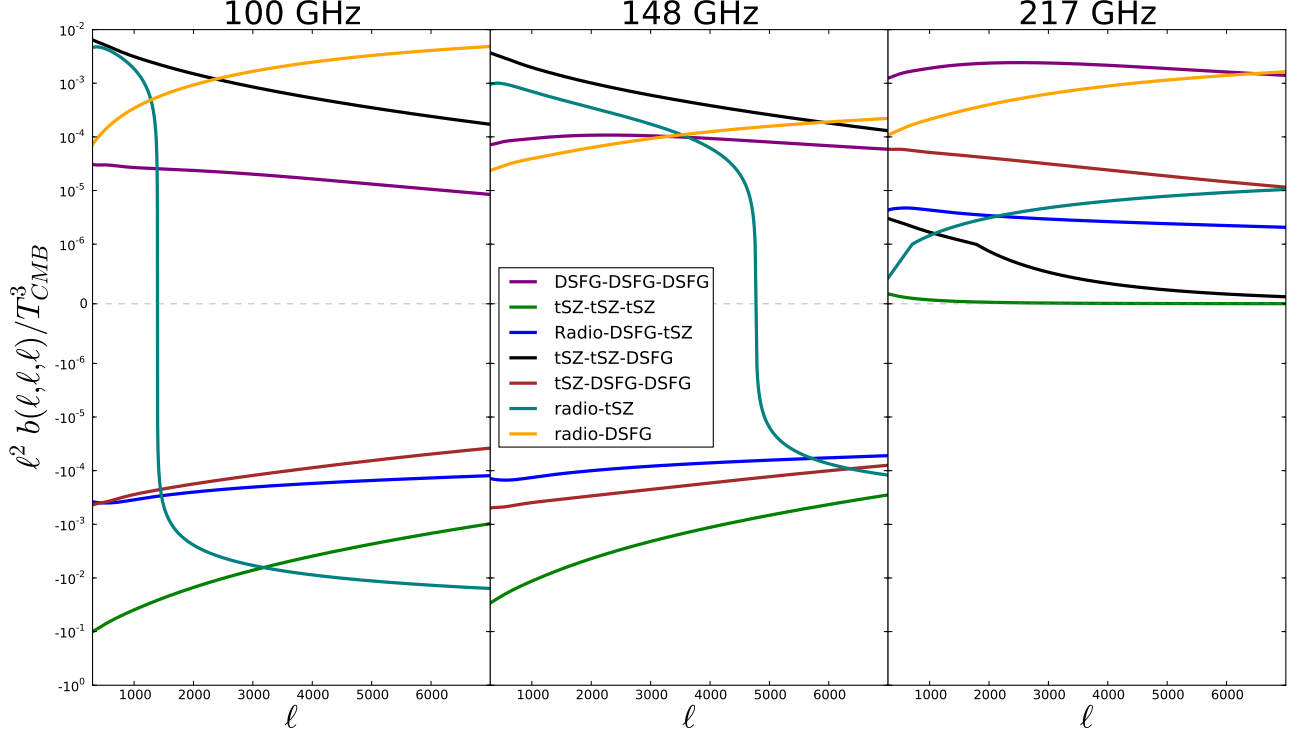
Figure 1. The one-halo bispectrum templates with $l_1 = l_2 = l_3$ for the tSZ, DSFGs and radio galaxies for three frequencies: 100, 148 and 217 GHz. The DSFG term contains the two and three galaxy term but not the one galaxy term; the one galaxy term is a Poisson term with trivial scale dependence and so is excluded. Similarly we exclude the radio point source term. The radio-tSZ is composed of two physical terms: the radio-tSZ-tSZ and the radio-radio-tSZ terms. The feature in the radio-tSZ template arises as the two contributions have different signs below 220 GHz and are dominant on different scales; the radio-tSZ-tSZ term is dominant on the largest scales and is positive at all frequencies and the radio-radio-tSZ term is dominant on the smallest scales and is negative at frequencies below 220 GHz. Similarly the radio-DSFG term is composed of two physical terms: the radio-radio-DSFG and the radio-DSFG-DSFG terms. The scale is linear between 10^{-6} and 10^6 and logarithmic outside this range. This leads to kinks in the curves that cross the log-linear transition.
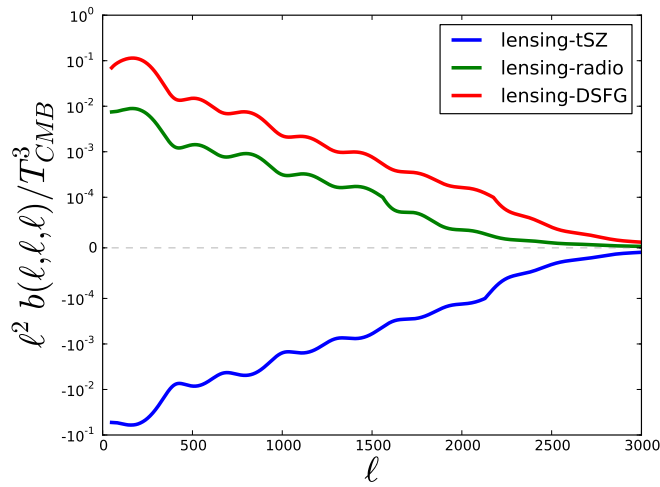

Figure 2. The lensing cross secondary source bispectrum templates with $l_1 = l_2 = l_3$ at 148 GHz. We exclude the ISW-lensing template here as it has been well studied before. The scale is linear between 10^{-4} and 10^4 and logarithmic otherwise.

overview the general prescription for calculating one halo bispectrum templates. Our discussion is based on that of Hill & Pajer (2013) and Lacasa et al. (2014).

In the halo model it is assumed that all the matter is contained in dark matter halos. We then assume that each quantity of interest is distributed within each cluster with profile $X^i(\boldsymbol{x}, M)$, where X^i could be the pressure of the halo gas or the distribution of radio galaxies and M is the cluster virial mass. We also make the assumption that these profiles are all spherically symmetric. The contribution to the 3D X^i field at point $\boldsymbol{x}$ in space from all clusters is then

$$X^i(\boldsymbol{x}) = \sum_{c_i} X^i(\boldsymbol{x} - \boldsymbol{x}_{c_i}, M_{c_i}), \tag{25}$$

where $\boldsymbol{x}_{c_i}$ is the location of halo c_i. We project this 3D field along the line of sight to find the sky signal in direction $\boldsymbol{n}$ as

$$X^i(\boldsymbol{n}) = \int \mathrm{d}t \, c X^i(\boldsymbol{x}(t, \boldsymbol{n})) = \int \mathrm{d}\chi a(\chi) X^i(\boldsymbol{x}(\chi, \boldsymbol{n})), \tag{26}$$

where c is the speed of light, t is the age of the universe at a given epoch, χ is the comoving distance, a is the scale factor. The three point function of three X^i fields is then

$$\left\langle X^i(\boldsymbol{n}) X^j(\boldsymbol{m}) X^k(\boldsymbol{p}) \right\rangle = \int \mathrm{d}\chi_i \mathrm{d}\chi_j \mathrm{d}\chi_k a(\chi_i) a(\chi_j) a(\chi_k) \times$$
$$\left\langle \sum_{c_i, c_j, c_k} X^i\left(\boldsymbol{x}(\chi_i, \boldsymbol{n}) - \boldsymbol{x}_{c_i}, M_{c_i}\right) X^j\left(\boldsymbol{x}(\chi_j, \boldsymbol{m}) - \boldsymbol{x}_{c_j}, M_{c_j}\right) X^k\left(\boldsymbol{x}(\chi_k, \boldsymbol{p}) - \boldsymbol{x}_{c_k}, M_{c_k}\right) \right\rangle \tag{27}$$

The contributions to the bispectrum can be decomposed into three parts: the one, two and three halo terms. The three halo term arises from correlations among three different halos, i.e. $c_i \neq c_j \neq c_k$. The two halo component arises when two of the quantities are located in same halo, e.g. $c_i = c_j \neq c_k$. The one halo term arises from intra halo correlations, i.e. $c_i = c_j = c_k$. The one halo term is

$$\left\langle X^i(\boldsymbol{n}) X^j(\boldsymbol{m}) X^k(\boldsymbol{p}) \right\rangle = \int \mathrm{d}\chi_i \mathrm{d}\chi_j \mathrm{d}\chi_k a(\chi_i) a(\chi_j) a(\chi_k) \times$$
$$\left\langle \sum_{c_i} \int \mathrm{d}^3 \boldsymbol{x}' \, \mathrm{d}M \, \delta(M - M_{c_i}) \delta^{(3)}(\boldsymbol{x}' - \boldsymbol{x}_{c_i}) X^i\left(\boldsymbol{x}(\chi_i, \boldsymbol{n}) - \boldsymbol{x}', M\right) X^j\left(\boldsymbol{x}(\chi_j, \boldsymbol{m}) - \boldsymbol{x}', M\right) X^k\left(\boldsymbol{x}(\chi_k, \boldsymbol{p}) - \boldsymbol{x}', M\right) \right\rangle \tag{28}$$

Using the definition of the halo mass function:

$$\frac{\mathrm{d}n(M, z)}{\mathrm{d}M} = \left\langle \sum_{c_i} \delta(M - M_{c_i}) \delta^{(3)}(\boldsymbol{x}' - \boldsymbol{x}_{c_i}) \right\rangle \tag{29}$$

leads to

$$\left\langle X^i(\boldsymbol{n}) X^j(\boldsymbol{m}) X^k(\boldsymbol{p}) \right\rangle = \int \mathrm{d}\chi_i \mathrm{d}\chi_j \mathrm{d}\chi_k a(\chi_i) a(\chi_j) a(\chi_k) \times$$
$$\int \mathrm{d}^3 \boldsymbol{x}' \, \mathrm{d}M \, \frac{\mathrm{d}n}{\mathrm{d}M} \left\langle X^i\left(\boldsymbol{x}(\chi_i, \boldsymbol{n}) - \boldsymbol{x}', M\right) X^j\left(\boldsymbol{x}(\chi_j, \boldsymbol{m}) - \boldsymbol{x}', M\right) X^k\left(\boldsymbol{x}(\chi_k, \boldsymbol{p}) - \boldsymbol{x}', M\right) \right\rangle. \tag{30}$$

Fourier expanding the 3D X^i fields and using the translational property we have

$$\left\langle X^i(\boldsymbol{n}) X^j(\boldsymbol{m}) X^k(\boldsymbol{p}) \right\rangle = \int \mathrm{d}\chi_i \mathrm{d}\chi_j \mathrm{d}\chi_k a(\chi_i) a(\chi_j) a(\chi_k) \times$$
$$\int \mathrm{d}M \, \frac{\mathrm{d}n}{\mathrm{d}M} \int \frac{\mathrm{d}\boldsymbol{k}_i^3 \mathrm{d}\boldsymbol{k}_j^3 \mathrm{d}\boldsymbol{k}_k^3}{(2\pi)^9} \left\langle \tilde{X}^i(\boldsymbol{k}_i, M) \tilde{X}^j(\boldsymbol{k}_j, M) \tilde{X}^k(\boldsymbol{k}_k, M) (2\pi)^3 \delta^{(3)}(\boldsymbol{k}_i + \boldsymbol{k}_j + \boldsymbol{k}_k) e^{i\boldsymbol{k}_i \cdot \boldsymbol{x}(\chi_i, \boldsymbol{n})} e^{i\boldsymbol{k}_j \cdot \boldsymbol{x}(\chi_j, \boldsymbol{m})} e^{i\boldsymbol{k}_k \cdot \boldsymbol{x}(\chi_k, \boldsymbol{p})} \right\rangle \tag{31}$$

Using the results from Appendix A.1 with the assumption that the X^i profiles are spherically symmetric, we can show that the reduced bispectrum is

$$b_{\ell_1, \ell_2, \ell_3}^{i, j, k} = \int \mathrm{d}\chi_i \mathrm{d}\chi_j \mathrm{d}\chi_k a(\chi_i) a(\chi_j) a(\chi_k) \times$$
$$\int \mathrm{d}x x^2 \prod_a \mathrm{d}k_a k_a^2 J_{\ell_a + \frac{1}{2}}(k_a \chi_a) J_{\ell_a + \frac{1}{2}}(k_a x) \frac{1}{k_a \sqrt{x \chi_a}} \int \mathrm{d}\ln M \, \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{X}^i(k_i, M) \tilde{X}^j(k_j, M) \tilde{X}^k(k_k, M), \tag{32}$$

where $k_i = |\boldsymbol{k}_i|$. Using the Limber approximation we have the final form (see Appendix A.1 for the details)

$$
\begin{aligned}
b^{i,j,k}_{\ell_1,\ell_2,\ell_3} &= \int \mathrm{d}\chi \frac{a(\chi)a(\chi)a(\chi)}{\chi^4} \int \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{X}^i(k_i, M)\tilde{X}^j(k_j, M)\tilde{X}^k(k_k, M) \\
&= \int \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z\mathrm{d}\Omega} \frac{a(\chi)a(\chi)a(\chi)}{\chi^6} \int \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{X}^i\left(\frac{\ell_1+1}{\chi(z)}, M\right) \tilde{X}^j\left(\frac{\ell_2+1}{\chi(z)}, M\right) \tilde{X}^k\left(\frac{\ell_3+1}{\chi(z)}, M\right).
\end{aligned} \tag{33}
$$

With this formalism we construct eight templates to characterize the non-Gaussianity from auto and cross correlations among the tSZ effect, DSFG and radio galaxies. The equatorial slice for these bispectra is plotted for three different frequencies in figure 1. The tSZ-tSZ-tSZ template describes non-Gaussianty from the tSZ effect only; the DSFG template describes the Poisson and clustered components of the DSFGs. The tSZ-tSZ-DSFG and tSZ-DSFG-DSFG templates model the cross correlations between the tSZ effect and DSFGs. The radio galaxy template describes the Poisson contributions of the radio galaxies. The radio-tSZ template contains two terms, the radio-radio-tSZ and radio-tSZ-tSZ terms, and arises from cross correlations between radio galaxies and the tSZ effect. The radio-DSFG template, which similarly contains both the radio-radio-DSFG and radio-DSFG-DSFG terms, describes the cross correlations between the DSFGs and radio galaxies. Finally the radio-DSFG-tSZ template describes the cross correlations among the three effects. Extending the results shown in Lacasa et al. (2014) we find that all the one-halo bispectrum terms are largely insensitive to the configuration but display strong scale dependence. Motivated by this we plot only equatorial slices, where $\ell_1 = \ell_2 = \ell_3$, of our templates to show the scale dependence.

3.3. *Lensing cross secondary sources templates*

Bispectra from lensing-induced anisotropies have been well studied (Spergel & Goldberg 1999; Goldberg & Spergel 1999; Lewis et al. 2011). We briefly summarize the origin of this non-Gaussianity here. We can decompose the anisotropies into contributions from the early universe and late-time sources:

$$
\Delta T(\boldsymbol{n}) = \Delta T^P(\boldsymbol{n} + \nabla\phi) + \Delta T^s(\boldsymbol{n}), \tag{34}
$$

where ΔT^P is the unlensed CMB fluctuations, ϕ is the lensing potential, and ΔT^S encodes the contributions from late time sources such as radio galaxies or clusters via the tSZ effect. If we expand in the lensing potential to first order we find

$$
\Delta T(\boldsymbol{n}) = \Delta T^P(\boldsymbol{n}) + \nabla\phi \cdot \nabla\Delta T^P(\boldsymbol{n}) + \Delta T^s(\boldsymbol{n}). \tag{35}
$$

From this we see that we can get non-vanishing bispectra arising from the terms $\left\langle \Delta T^P(\boldsymbol{n})\nabla\phi \cdot \nabla\Delta T^P(\boldsymbol{n})\Delta T^s(\boldsymbol{n}) \right\rangle$ as the secondary sources are correlated with the structures that cause the lensing. More precisely these lead to the following reduced bispectrum

$$
b^{\text{lensing-sec.}}_{\ell_1,\ell_2,\ell_3} = -\boldsymbol{\ell}_1 \cdot \boldsymbol{\ell}_2 C^{TT}_{\ell_1} C^{\phi S}_{\ell_2} + 5\,\text{permutations}, \tag{36}
$$

where $C^{\phi S}$ is the cross-correlation between the lensing potential and the secondary source. We consider four lensing bispectrum templates arising from cross-correlations with the tSZ effect, radio galaxies, DSFG and the integrated Sachs Wolfe effect. In figure 2 we show the equatorial slice through the bispectrum for several of the lensing templates considered in this work. The shape of the lensing templates is largely insensitive to frequency so we display only a single frequency here. The shape of the ISW-lensing template has been well studied before (Lewis et al. 2011) and so is not shown here. The other lensing templates have a similar spatial dependence so we plot only the equatorial configuration here.

4. DATA SETS

The analysis in this paper is focused on using the small-scale information from the ACTPol experiment. Several non-Gaussian sources, such as the Poisson contributions from DSFG and radio galaxies, have similar template shapes and cannot be distinguished with the current 148GHz ACTPol maps. The different spectral behavior of the sources enables their differentiation with multi-frequency data. For this purpose we use data from the *Planck Experiment*. Through combining ACTPol and *Planck* 100 GHz and 217 GHz data we use the small-scale information from ACTPol and the multifrequency data from *Planck*. Most of our constraining power will come from bispectrum configurations involving one *Planck* map and two ACTPol maps. Figure 3 shows that our estimator will be most sensitive to scales of $1000 < \ell < 2000$ for the *Planck* maps and $2000 < \ell < 5000$ for the ACTPol maps.
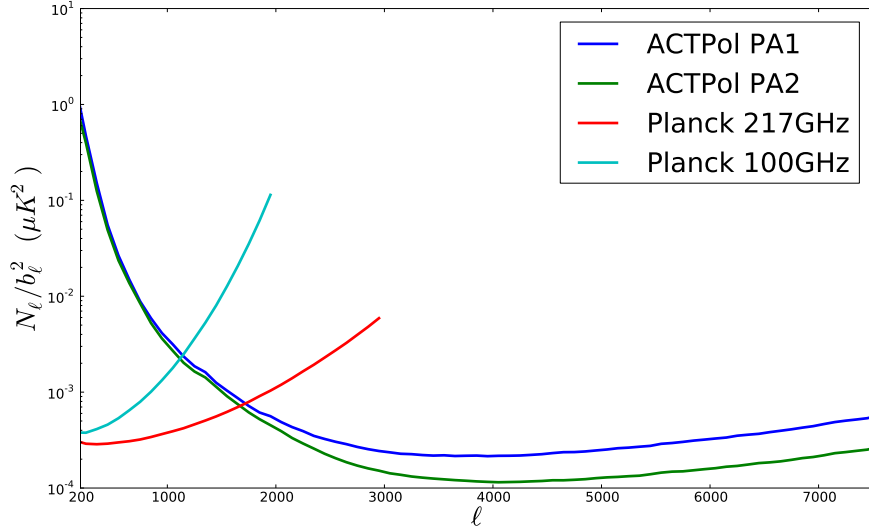
Figure 3. The noise power spectra divided by the beam function squared for the four maps used in this analysis. The beams for the *Planck* 100 GHz and 217 GHz maps are treated as zero for $\ell > 2000$ and $\ell > 3000$ respectively.

4.1. *ACTPol data sets*

We use data in the 'D56' field, a patch of sky on the equator with coordinates $-7.2° < \text{dec} < 4°$ and $352° < \text{RA} < 41°$. This is part of the data described in Louis et al. (2017); in particular we use only the wide field and not the deep fields. In this work we use the data from the two arrays (called PA1 and PA2 hereafter) which observed the sky at 148 GHz. The ACT experiment has a full width at half maximum (FWHM) of 1.4′ at 148 GHz. As was shown in Louis et al. (2017) non-white atmospheric noise, from atmospheric temperature brightness fluctuations, dominates the largest scales. On the smallest scales the noise is approximately white at 31 μK-arcmin and 25 μK-arcmin for PA1 and PA2 respectively. We mask all the point sources whose fluxes were measured to be above 30 mJy with discs of radius 5′ and perform no masking of clusters or of galactic dust.

4.2. *Planck Data sets*

We use only the *Planck* 100 and 217 GHz maps as the lower frequencies have too limited sensitivity to the small-scales (due to the > 5′ beam) and the higher frequencies are obscured by dust in this region. Instead of using the full information of the Planck maps we remain in the flat-sky regime and use only the portion of the *Planck* data that overlaps with the ACTPol D56 region. The ACTPol and Planck maps we used as input do not come with the same pixelization or coordinate system. ACTPol uses equatorial coordinates in equal-area coordinates pixelization (CEA), while *Planck* uses HEALPix-pixelized galactic coordinates. Before cross-correlating the maps, we reprojected *Planck* onto the ACTPol pixels by first expanding it in spherical harmonic coefficients and rotating these coefficients from galactic to equatorial coordinates using healpy, and then evaluating these coefficients on the ACTPol pixels using libsharp (Reinecke & Seljebotn 2013).

We mask point sources that were detected in the ACTPol 148 GHz map with fluxes above 269 mJy and 152 mJy for the 100 GHz and 217 GHz maps respectively with discs of radii 12.5′ and 6.5′. These masking levels are *Planck*'s 90% completeness limit for these frequencies (Planck Collaboration 2016h). We perform no masking of clusters or of galactic dust. The *Planck* experiment's beam FWHM are 9.7′, and 5.0′ at 100 GHz and 217 GHz respectively. The small scale *Planck* noise is white with levels of 77.4 μK-arcmin and 46.8 μK-arcmin at 100 GHz and 217 GHz respectively (Planck Collaboration 2016a). To avoid issues of nearly singular matrices during our inverse covariance filtering we add uncorrelated noise to the *Planck* maps with a power spectrum equal to 10% of the primary CMB (with the appropriate beam and window functions). Without this the largest scale modes of the *Planck* maps would be very highly correlated leading to potential difficulties numerical instabilities.

5. ANALYSIS PIPELINE

Full inverse-variance weighting of these maps is a computationally costly process and whilst methods exists for performing this (Elsner & Wandelt 2013; Smith et al. 2009), the *Planck* team has shown (Planck Collaboration 2014e) that near optimality can

be obtained using approximate methods that assume the covariance matrix is diagonal. In this work we will use the approximate method described in Planck Collaboration (2014e).

First we apply a point source mask that masks the brightest point sources as stated in sections 4.1 and 4.2. We then use a similar method to Planck Collaboration (2014e), in which we fill in masked point source pixels with the average of their neighbors and iterate until the solution converges. Convergence is typically attained within 1000 iterations and 2000 iterations are used for our final analysis. This infilling procedure reduces mode coupling and allows the approximation of a diagonal covariance matrix without loss of optimality.

Next we apply our real space mask. This mask is based on the square root of the smoothed hit counts map ($H_s(\boldsymbol{n})$) with the weighting of each pixel, M($\boldsymbol{n}$), given by:

$$M(\boldsymbol{n}) \propto \frac{H_s(\boldsymbol{n})}{H_s(\boldsymbol{n}) + 15000} F(\boldsymbol{n}), \tag{37}$$

where $F(\boldsymbol{n})$ is a smoothed top hat mask, we mask all pixels with fewer than 5500 counts in either of the ACTPol arrays and then convolve this mask with a cosine squared of width 25′. This mask serves two purposes: it down-weights noisy and infrequently observed regions of the sky, thereby reducing the influence of inhomogeneous noise, and it serves as an edge taper that reduces mode coupling. The hits map was smoothed with a Gaussian of FWHM 9.5′ to limit leakage. On the scale of our smoothed hits-map mask, the power in the map for heavily observed regions (hit counts $\gtrsim 15,000$) is dominated by the CMB. In these regions of the maps our mask is approximately uniform to prevent overweighting these regions.

The value of 15000 counts was chosen to prevent over weighting of the heavily observed regions; The FFT methods we use assume periodic boundary conditions. To prevent spurious correlations we zero-pad our maps by a factor of two before performing a Fourier transform on the maps. Finally, we Fourier transform our maps and then apply a k-space mask to our maps. For the ACTPol maps, we mask modes with $\ell < 100$ and apply a sine squared filter to modes with $100 < \ell < 500$; these modes are dominated by atmospheric $1/f$ noise. This filter reduces the leakage of power from these modes to higher ℓ modes, which have less power. To avoid complications of the sine squared mask, we apply a mask to our theoretical templates for the ACTPol maps that masks all modes with $\ell < 500$. We then remove all modes with $|\ell_x| < 90$ and $|\ell_y| < 90$, as these are dominated by ground contamination (Das et al. 2011). For the *Planck* maps, we filter modes with $\ell < 200$ by applying a sine squared filter to modes with $0 < \ell < 200$ and then applying a $\ell < 200$ cut to our templates. This is done for two reasons; first the flat-sky approximation becomes less accurate at low ℓ and second the two and three halo terms cannot be ignored if the lowest ℓ modes are used.

Once processed as above, we use the maps to estimate the amplitudes of the theoretical templates described in section 3. We use the KSW implementation of equation 12 to calculate the template amplitudes and equation 16 to jointly fit our templates. The Gaussian errors are calculated using equation 17. Our estimator jointly fits the amplitudes of all the non-Gaussian templates and uses all the possible auto and cross bispectra from the four maps. We use an $\ell_{\max} = 7500$. This cut off was chosen as there is little signal at small scales (higher ℓ) and it is computational expensive to include the smaller scales. The validation of the analysis pipeline is described in Appendix C.

6. TEMPLATE MEASUREMENTS WITH ACTPOL AND *PLANCK*

In table 1 we present the measured template amplitudes and the Gaussian errors for these templates. The "joint fit" results come from simultaneously fitting all of the templates. The Gaussian error characterizes how well we can measure the different types of non- Gaussianity given the finite number of noisy modes in our data set. Our measured amplitudes with the Gaussian error describe the level of non-Gaussianity and its significance in this data set. We find significant amplitudes for the tSZ-tSZ-tSZ, tSZ-tSZ-DSFG, radio galaxy, and radio-DSFG-tSZ templates. When simultaneously measuring multiple templates, the Gaussian covariance represents how well the estimator can differentiate between the types of non-Gaussianity. In figure 4 we present the Gaussian covariance matrix for our measurements; it can be seen that, despite the use of multifrequency information, we still have several highly correlated terms. In particular, we see strong covariances between the radio-DSFG term and the DSFG only term; between the radio-tSZ term and the tSZ only term; and between the tSZ-DSFG-DSFG term and the radio galaxy term. These strong covariances, as well as the others in the dataset, make our measurements particularly sensitive to our models, since inaccurate theoretical models can 'leak' the signal of one template into another, tightly correlated template.

As we have strong detections of these templates it is important to verify whether the Gaussian errors (that are valid in the limit of weak non-Gaussianity) are still dominant and, if not, to include the non-Gaussian contribution to the cosmic variance. Our method for calculating the six point function is described in Appendix D. In table 1 we present our constraints on the amplitudes with the non-Gaussian cosmic variance errors; the measured amplitudes with the full non-Gaussian errors represents our constraints on the population amplitudes of these templates. It can be seen that the errors for some of the terms are effectively doubled by the

Type	Measured A_i	
	Gaussian Errors	Full non-Gaussian Errors
lensing x radio	-0.31 ± 6.26	-0.31 ± 6.37
lensing x tSZ	1.74 ± 1.55	1.74 ± 1.60
lensing x DSFG	0.43 ± 0.41	0.43 ± 0.45
lensing x ISW	47.86 ± 29.22	47.86 ± 28.81
tSZ-tSZ-tSZ	0.80 ± 0.25	0.80 ± 0.65
tSZ-tSZ-DSFG	1.20 ± 0.29	1.20 ± 0.80
tSZ-DSFG-DSFG	-0.96 ± 0.47	-0.96 ± 0.78
radio-DSFG-tSZ	6.03 ± 1.31	6.03 ± 1.83
DSFG-DSFG-DSFG	1.65 ± 0.44	1.65 ± 0.45
radio-tSZ	1.20 ± 0.82	1.20 ± 1.05
radio-DSFG	-0.45 ± 1.40	-0.45 ± 1.65
radio-radio-radio	0.99 ± 0.08	0.99 ± 0.13

Table 1. Template amplitudes obtained from a joint fit using 100 GHz and 217 GHz Planck maps with the ACTPol PA1 and PA2 148 GHz maps and $\ell_{max} = 7500$. Point sources below 269, 152, 30 mJy for frequencies 100, 217, 148 GHz respectively were masked and no masking of clusters was performed. The amplitudes, A_i, are overall scalings of our templates and are dimensionless. The Gaussian errors only indicate the significance of the non-Gaussianity in these data sets. The non-Gaussian error results include the one-halo contributions to the cosmic variance and these errors reflect the statistical uncertainty in the template amplitudes.

inclusion of the six point terms. Figure 6 shows our results with the Gaussian only errors and with the full non-Gaussian errors. In figure 5 we present the covariance matrix of the templates when we include the full six point function. The full covariance deviates significantly from the Gaussian-only covariance matrix further indicating the necessity of including the non-Gaussian components. The source of the non-Gaussian covariances differs fundamentally from those in the Gaussian case: the Gaussian covariances arise due to the similarity of the templates and the difficulty of differentiating between them with our data sets, whereas the non-Gaussian covariances arise from the physical correlations among the sources of these templates. For example, the large covariance that exists between the tSZ-tSZ-tSZ and tSZ-tSZ-DSFG template primarily arises from the Poissonian fluctuations of the number of dark matter halos, as an increase (or decrease) in the number of dark matter halos will increase (decrease) the tSZ-tSZ-tSZ bispectrum and tSZ-tSZ-DSFG bispectrum in an identical and correlated manner.

7. DISCUSSION, CONCLUSIONS AND FUTURE DIRECTIONS

We have applied the KSW estimator to simultaneously measure a wide range of sources of non-Gaussianity in the CMB. These first measurements with ACTPol extend the similar work done by Crawford et al. (2014) with the SPT experiment by examining cross-correlations between the sources. In future work we will have the sensitivity to constrain the bispectrum of the kinetic Sunyaev Zel'dovich effect (see Appendix B.6 for current sensitivities) and thereby fully constrain the bispectrum of known CMB secondary sources.

There have been several previous measurements of non-Gaussianity from secondary anisotropies. In table 2 we show the measurements of the bispectrum from the DSFG and radio galaxies from previous work. We compare our results to measurements of the bispectrum amplitude of radio sources from Komatsu et al. (2009) and Calabrese et al. (2010) for the WMAP Q, V and W bands and to measurements of the DSFGs bispectrum from Crawford et al. (2014). We also show the bispectrum contribution that we would predict from our model as constrained by the measurements. For the WMAP results we assume that the only contribution is from the radio galaxies and use a flux cut of 0.5 Jy at 22 GHz. In both of the analyses of the WMAP data, the authors mask all the point sources detected in Wright et al. (2009) as well radio galaxies identified in external surveys with fluxes greater than 0.5 Jy and 22 GHz (as described in Hinshaw et al. 2007). The comparison to the WMAP Q and V bands is done by extrapolating our model beyond the frequencies measured in this work. To compare our measurements of DSFGs with Crawford et al. (2014) we have combined their measurements of the Poisson and clustered components of the DSFGs and compared this to our model's prediction for the DSFG contribution at $\ell = 2000$ and 220 GHz. The radio Poisson contribution at this frequency with a 6.4 mJy mask is negligible. We see that our model has good agreement with the measurements in Komatsu et al. (2009), but differ from Calabrese et al. (2010) by a factor of two. A similar factor of two between the Komatsu et al. (2009) and Calabrese
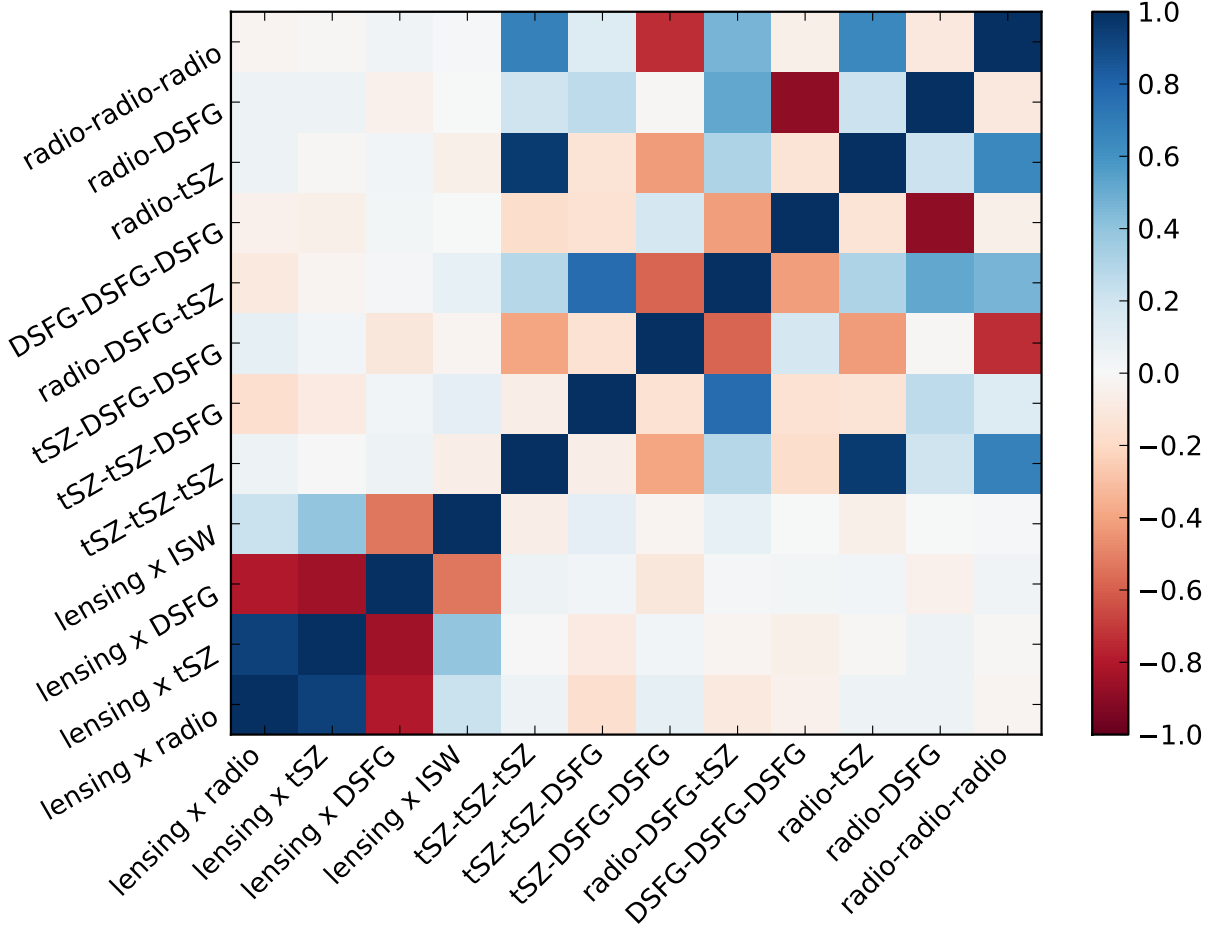
Figure 4. The covariance matrix of template amplitudes only accounting for the Gaussian component to the six point function. The strong covariances between several templates highlights the difficult of disentangling all of the sources of non-Gaussianity.

	Komatsu et al.	Calabrese et al.	Crawford et al.	Our model
WMAP Q band radio galaxies	$6.0 \pm 1.3 \times 10^{-5}$	$13.7 \pm 1.1 \times 10^{-5}$	–	$5.5 \pm 0.7 \times 10^{-5}$
WMAP V band radio galaxies	$0.43 \pm 0.31 \times 10^{-5}$	$2.3 \pm 0.5 \times 10^{-5}$	–	$0.49 \pm 0.06 \times 10^{-5}$
WMAP W band radio galaxies	$0.12 \pm 0.24 \times 10^{-5}$	$1.0 \pm 0.5 \times 10^{-5}$	–	$0.053 \pm 0.07 \times 10^{-5}$
220 GHz DSFG (Poisson and Clustered)	–	–	$0.917 \pm 0.14 \times 10^{-9}$	$1.41 \pm 0.38 \times 10^{-9}$

Table 2. A comparison between various different bispectrum measurements with all results given in $\mu K^3 sr^2$. The Komatsu et al. column contains the results from Komatsu et al. (2009), Calabrese et al. column contains the results from Calabrese et al. (2010) and Crawford et al. column is from Crawford et al. (2014). The final column shows what our model, with amplitudes fitted from the ACTPol and Planck data as presented here, would predict for these measurements. To compare our DSFG result to the Crawford et al. result we evaluate our model at $\ell = 3000$, 220 GHz and at their flux cut of 6.4 mJy. The factor of two difference between our results and those from Calabrese et al. (2010) is discussed in the text.

et al. (2010) results was discussed in Calabrese et al. (2010). The origin of this disparity is not understood. We see good agreement with the DSFG measurements, though the errors are large; a more interesting comparison will be possible with future data sets. Our measurements of the tSZ bispectrum and lensing-ISW bispectrum are consistent with previous work (Crawford et al. 2014; Planck Collaboration 2016j; Wilson et al. 2012; Calabrese et al. 2010), but with large errors.

From our tSZ bispectrum template amplitude we can compute a constraint on σ_8, the amplitude of matter fluctuations smoothed on $8h^{-1}$ Mpc scales. Our analysis follows that of Crawford et al. (2014). Using priors from Planck Collaboration (2016c), we
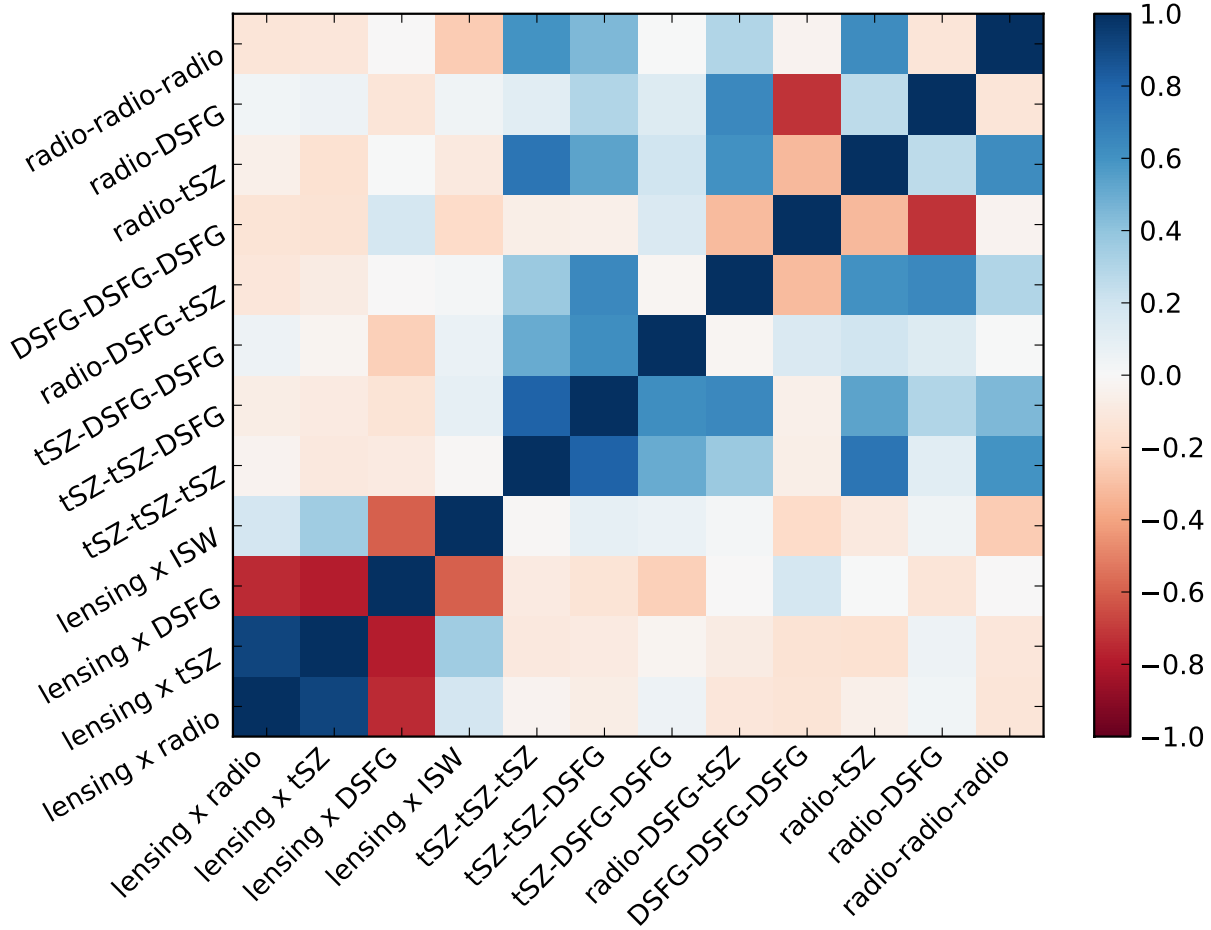
Figure 5. The full one-halo covariance matrix for our measured estimators. The strong covariances present in this figure, but absent in Fig 4 arise due to the physical covariances of these templates.

marginalize over the following cosmological parameters: the baryon density, Ω_b; the matter density, Ω_m; the Hubble constant, H_0; and the scalar spectral index with n_s. Finally we assume an astrophysical modeling uncertainty of 35% (Bhattacharya et al. 2012; Hill & Sherwin 2013). The resulting constraint on σ_8 is

$$\sigma_8 = 0.79 \pm 0.07. \tag{38}$$

We jointly fit the tSZ amplitude with the other secondary templates which provides some robustness to possible contamination. However, large degeneracies mean that our results are sensitive to the theoretical modeling; this is discussed further in the paragraph below. We do not include any uncertainty in our σ_8 constraint from the joint modeling.

We now consider the ability of the data set used in this work to discriminate among the dozen bispectrum templates in our model. The model is not a good fit to the data, with $\chi^2 = 43.6$ for 12 degrees of freedom. This poor fit is likely due to the strong degeneracies seen in the Gaussian covariance matrix (figure 4) and deficiencies in the halo model templates. These degeneracies mean that small changes to the covariance matrix lead to large shifts in our results. Physical fluctuations, for example those caused by fluctuations in the number of massive dark matter halos in the patch, or deficiencies in our modeling could mean that our model covariance differs from the true Gaussian covariance of our data and could cause leakage of one template into another. The interpretation of the chi-squared is further complicated as the distribution of the template amplitudes is non-Gaussian. The full non-Gaussian probability distribution for the template amplitudes is difficult to calculate and will be the subject of future work. This suggests that the theoretical uncertainty of our results could be significant and is likely the limiting factor in our analysis; as such, we stress that our results are a proof of principle for this approach, rather than accurate measurements of these secondary anisotropies.
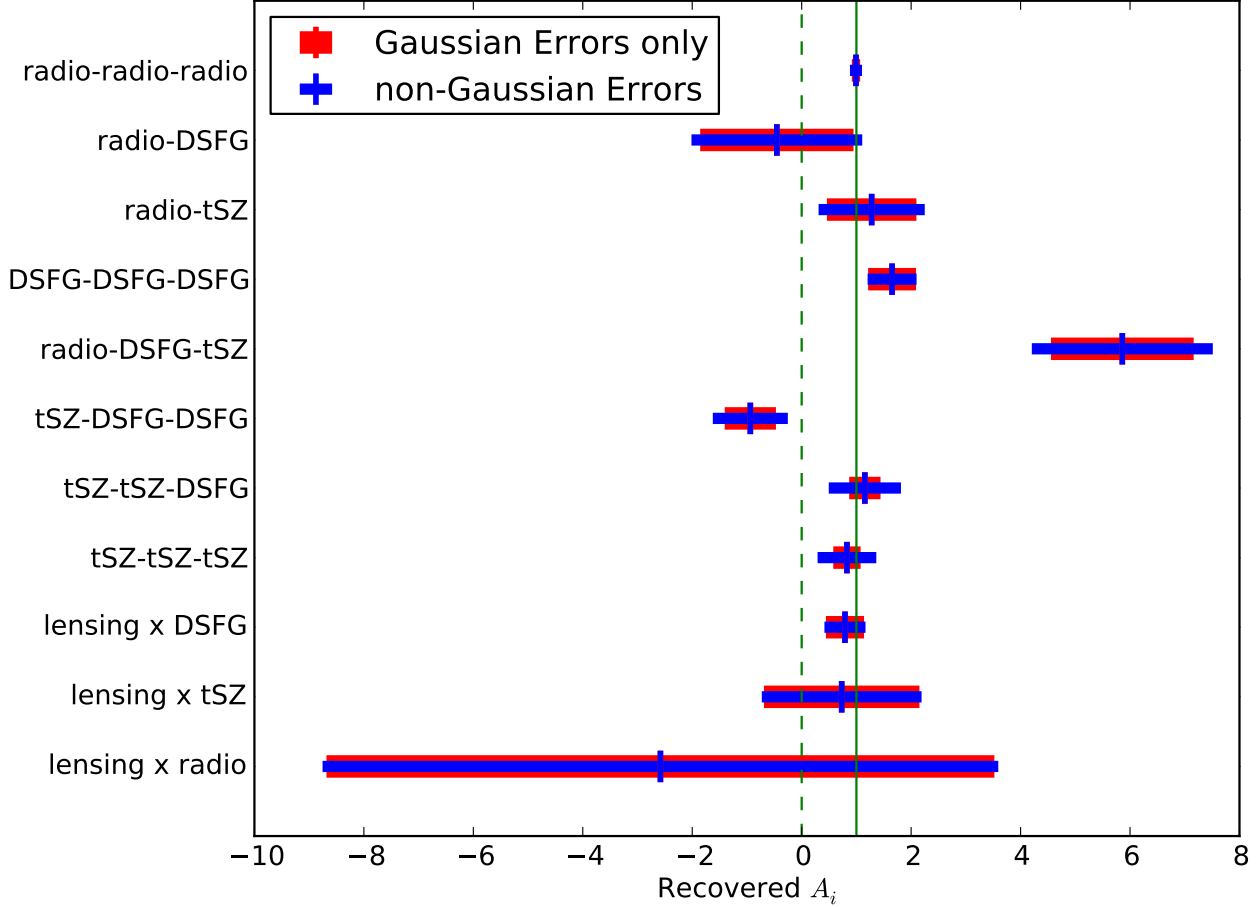
Figure 6. A graphical representation of our joint fit results from table 1 excluding the lensing-ISW constraint in order to restrict the scales. The solid green line is the values predicted by our model and the dashed green line is the null value.

The issues raised above will be addressed with future measurements. Multi-frequency small-scale information, particularly at low frequencies, which will be available from Advanced ACT, SPT3G, Simons Observatory and CMB-S4 (Abazajian et al. 2016; Henderson et al. 2016; Benson et al. 2014), should break the strong degeneracies. This will significantly reduce the dependence on the modeling and allow the use of advanced statistics, such as Skew-C_ℓs (Munshi & Heavens 2010), to assess the accuracy of our templates, and to study the astrophysics of these sources. In particular, it will be interesting to see if the high amplitudes of the radio-tSZ-DSFG terms are caused by template degeneracies or are due to radio galaxies occupying lower mass clusters at redshift ~ 1 than is assumed in our model. If it is the later then it suggests that radio galaxies, DSFGs and the tSZ effect can be found in the same halos and understanding the correlations among these terms will be important for searches for the kSZ signal.

The calculation of the six point function indicated that non-Gaussian contributions for such measurements are already important and cannot be neglected in future more sensitive measurements. The contribution of the non-Gaussian component to the variance has the potential to limit the power of the bispectrum to measure astrophysics and cosmology through its reduction on the information. However there are many measures that can be employed to reduce this error. For example, masking the largest clusters will significantly reduce the importance of these terms as the six point function receives the largest contribution from low redshift and massive clusters (even more so than the bispectrum itself). An even better approach, though computationally very costly, would be to jointly analyze the bispectrum and cluster counts; such an approach would allow a reduction of the errors without loss of the information from the largest clusters.

APPENDIX

A. THE FLAT-SKY -FULL-SKY BISPECTRUM CORRESPONDENCE

In this section we follow the power spectrum derivation of Loverde & Afshordi (2008) and derive a relation between the full-sky and flat-sky bispectrum.

A.1. *Full-sky*

The 2-D three-point function of three fields is

$$\langle A(\boldsymbol{n}_1)B(\boldsymbol{n}_2)C(\boldsymbol{n}_3)\rangle = \left\langle \int \mathrm{d}r_1 \mathrm{d}r_2 \mathrm{d}r_3 F_A(r_1) A^{\mathrm{3D}}(r_1\boldsymbol{n}_1) F_B(r_2) B^{\mathrm{3D}}(r_2\boldsymbol{n}_2) F_C(r_3) C^{\mathrm{3D}}(r_3\boldsymbol{n}_3) \right\rangle, \tag{A1}$$

where $F_X(\boldsymbol{x})$ is the projection kernel and in harmonic space we have

$$\langle A_{\ell_1,m_1}B_{\ell_2,m_2}C_{\ell_3,m_3}\rangle = \int \mathrm{d}\Omega_1 \mathrm{d}\Omega_2 \mathrm{d}\Omega_3 \left\langle A(\boldsymbol{n}_1)B(\boldsymbol{n}_2)C(\boldsymbol{n}_3) Y^*_{\ell,m}(\boldsymbol{n}_1) Y^*_{\ell',m'}(\boldsymbol{n}_2) Y^*_{\ell'',m''}(\boldsymbol{n}_3) \right\rangle. \tag{A2}$$

To proceed we decompose each 3D field into its Fourier components and then use the plane wave expansion:

$$e^{i\boldsymbol{k}\cdot\boldsymbol{x}} = 4\pi \sum_{\ell,m} i^\ell j_\ell(kr) Y^*_{\ell,m}(\hat{\boldsymbol{k}}) Y^*_{\ell,m}(\hat{\boldsymbol{n}}). \tag{A3}$$

Performing these steps and evaluating the angular integrals leaves

$$\langle A_{\ell_1,m_1}B_{\ell_2,m_2}C_{\ell_3,m_3}\rangle = \int \mathrm{d}r_1 \mathrm{d}r_2 \mathrm{d}r_3 F_A(r_1) F_B(r_2) F_C(r_3) \int \prod_i \frac{\mathrm{d}\boldsymbol{k}_i^3}{(2\pi)^3} 4\pi i^\ell_i j_\ell(k_i r_i) Y^*_{\ell_i,m_i}(\hat{\boldsymbol{k}}) \langle \tilde{A}^{\mathrm{3D}}(\boldsymbol{k}_1) \tilde{B}^{\mathrm{3D}}(\boldsymbol{k}_2) \tilde{A}^{\mathrm{3D}}(\boldsymbol{k}_3) \rangle. \tag{A4}$$

We then assume translational and rotational invariance, i.e.

$$\langle \tilde{A}^{\mathrm{3D}}(\boldsymbol{k}_1) \tilde{B}^{\mathrm{3D}}(\boldsymbol{k}_2) \tilde{A}^{\mathrm{3D}}(\boldsymbol{k}_3) \rangle = (2\pi)^3 \delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) B(k_1, k_2, k_3) \tag{A5}$$

so

$$\langle A_{\ell_1,m_1}B_{\ell_2,m_2}C_{\ell_3,m_3}\rangle = \int \mathrm{d}r_1 \mathrm{d}r_2 \mathrm{d}r_3 F_A(r_1) F_B(r_2) F_C(r_3) \times$$
$$\int \prod_i \frac{\mathrm{d}\boldsymbol{k}_i^3}{(2\pi)^3} 4\pi i^\ell_i j_\ell(k_i r_i) Y^*_{\ell_i,m_i}(\hat{\boldsymbol{k}}) B(k_1, k_2, k_3) \int \mathrm{d}\boldsymbol{x} \, e^{i(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)\cdot\boldsymbol{x}}. \tag{A6}$$

Again using the plane wave expansion and evaluating the angular integrals gives

$$\langle A_{\ell_1,m_1}B_{\ell_2,m_2}C_{\ell_3,m_3}\rangle = \mathcal{G}^{\ell_1 \ell_2 \ell_3}_{m_1 m_2 m_3} \int \mathrm{d}r_1 \mathrm{d}r_2 \mathrm{d}r_3 F_A(r_1) F_B(r_2) F_C(r_3) \times$$
$$\int \mathrm{d}x \, x^2 \prod_i \frac{\mathrm{d}k_i (4\pi k_i)^2}{(2\pi)^3} j_{\ell_i}(k_i r_i) j_{\ell_i}(k_i x) B(k_1, k_2, k_3). \tag{A7}$$

where:

$$\mathcal{G}^{\ell_1 \ell_2 \ell_3}_{m_1 m_2 m_3} = \int \mathrm{d}\Omega^2 Y_{\ell_1,m_1}(\boldsymbol{n}) Y_{\ell_2,m_2}(\boldsymbol{n}) Y_{\ell_3,m_3}(\boldsymbol{n}) \tag{A8}$$

is the Gaunt integral. Note that the bispectrum can now be expressed in terms of the reduced bispectrum of the full-sky, $b^{\text{FULL}}_{\ell_1,\ell_2,\ell_3}$,

$$\langle A_{\ell_1,m_1} B_{\ell',m'} C_{\ell'',m''} \rangle = \mathcal{G}^{\ell_1\,\ell_2\,\ell_3}_{m_1 m_2 m_3} b^{\text{FULL}}_{\ell_1,\ell_2,\ell_3}. \tag{A9}$$

We will focus on the reduced bispectrum. Next we replace the spherical Bessel functions, $j_\ell(x)$, with Bessel functions of the first kind, $J_{\ell+\frac{1}{2}}(x) = j_\ell(x)\sqrt{\frac{2x}{\pi}}$ to obtain

$$b^{\text{FULL}}_{\ell_1,\ell_2,\ell_3} = \int dr_1 dr_2 dr_3 F_A(r_1) F_B(r_2) F_C(r_3) \int dx x^2 \prod_i dk_i k_i^2 J_{\ell_i+\frac{1}{2}}(k_i r_i) J_{\ell_i+\frac{1}{2}}(k_i x) \frac{1}{k_i \sqrt{x r_i}} B(k_1, k_2, k_3). \tag{A10}$$

For notational simplicity we will define

$$f_i(r) = \frac{F_A(r)}{\sqrt{r}}. \tag{A11}$$

We can now use the relation

$$\int dx J_\nu(x) f(x) = f(\nu) - \frac{1}{2} f''(\nu) - \frac{\nu}{6} f'''(\nu) + ...$$

$$\int dx J_\nu(kx) f(x) = \frac{f(\frac{\nu}{k})}{k} - \frac{1}{2k^3} f''(\frac{\nu}{k}) - \frac{\nu}{6k^4} f'''(\frac{\nu}{k}) + ... \tag{A12}$$

where $\nu = kx$, to get

$$b^{\text{FULL}}_{\ell_1,\ell_2,\ell_3} = \int dx x^2 \prod_i dk_i J_{\ell_i+\frac{1}{2}}(k_i x) \frac{1}{\sqrt{x}} \left(f_A(r_1) - \frac{f''_A(r_1)}{2k_1^2} - \frac{\nu_1}{6k_1^3} f'''_A(r_1) + ... \right)$$

$$\left(f_B(r_2) - \frac{f''_B(r_2)}{2k_2^2} - \frac{\nu_2}{6k_2^3} f'''_B(r_2) + ... \right) \left(f_C(r_3) - \frac{f''_C(r_3)}{2k_3^2} - \frac{\nu_3}{6k_3^3} f'''_C(r_3) + ... \right) B(k_1, k_2, k_3) \tag{A13}$$

with $\nu_i = \ell_i + \frac{1}{2} = k_i r_i$. Applying that relation again and retaining only the lowest terms in ν_i leads to

$$b^{\text{FULL}}_{\ell_1,\ell_2,\ell_3} = \int \frac{dx}{x^4} F_A(x) F_B(x) F_C(x) B\left(\frac{\nu_1}{x}, \frac{\nu_2}{x}, \frac{\nu_3}{x}\right) \times$$

$$\left(1 - \sum_i \frac{1}{6\nu_i^2} \left[\left(\frac{d \ln(f_i B)}{d \ln k_i}\right)^3 + \left(\frac{d \ln f_i}{d \ln x}\right)^3 + \frac{d^3 \ln(f_i B)}{d \ln^3 k_i} + \frac{d^3 \ln f_i}{d \ln^3 x} + 3\frac{d^2 \ln(f_i B)}{d \ln^2 k_i} \frac{d \ln(f_i B)}{d \ln k_i} + 3\frac{d^2 \ln f_i}{d \ln^2 x} \frac{d \ln f_i}{d \ln x} - \frac{d \ln B}{d \ln k_i} \right] + ...O\left(\nu^{-4}\right) \right). \tag{A14}$$

Note that we now have $\nu_i = \ell_i + \frac{1}{2} = k_i r_i = k_i x$.

A.2. Flat-sky

In the flat-sky regime we have

$$\langle A(\mathbf{x}_1^\perp) B(\mathbf{x}_2^\perp) C(\mathbf{x}_3^\perp) \rangle = \left\langle \int dx_1^\| dx_2^\| dx_3^\| F_A(x_1^\|) A^{3D}(x_1^\|, \mathbf{x}_1^\perp) F_B(x_2^\|) B^{3D}(x_2^\|, \mathbf{x}_2^\perp) F_C(x_3^\|) C^{3D}(x_3^\|, \mathbf{x}_3^\perp) \right\rangle \tag{A15}$$

where $x^\|$ is the coordinate along the line of sight and $\mathbf{x}^\perp$ is perpendicular to the line of sight. Expanding in flat-sky Fourier modes gives

$$\langle \bar{A}(\mathbf{k}_1^\perp) \bar{B}(\mathbf{k}_2^\perp) \bar{C}(\mathbf{k}_3^\perp) \rangle = \prod_i \int dx_i^\| \frac{dk_i^\|}{2\pi} F_i(x_i^\|) e^{ik_i^\| x_i^\|} \left\langle \tilde{A}^{3D}(\mathbf{k}_1^\perp, k_1^\|) \tilde{B}^{3D}(\mathbf{k}_2^\perp, k_2^\|) \tilde{C}^{3D}(\mathbf{k}_3^\perp, k_3^\|) \right\rangle. \tag{A16}$$

Assuming translational and rotational invariance we can then write

$$\langle \tilde{A}^{3D}(\mathbf{k}_1^\perp, k_1^\|) \tilde{B}^{3D}(\mathbf{k}_2^\perp, k_2^\|) \tilde{C}^{3D}(\mathbf{k}_3^\perp, k_3^\|) \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(k_1, k_2, k_3) \tag{A17}$$

so

$$\langle \bar{A}(\boldsymbol{k}_1^\perp)\bar{B}(\boldsymbol{k}_2^\perp)\bar{C}(\boldsymbol{k}_3^\perp)\rangle = \prod_i \int \mathrm{d}x_i^\parallel \frac{\mathrm{d}k_i^\parallel}{2\pi} F_i(x_i^\parallel)e^{ik_i^\parallel x_i^\parallel}(2\pi)^3\delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)B(k_1, k_2, k_3). \tag{A18}$$

This can be written in terms of the flat-sky bispectrum:

$$\langle \bar{A}(\boldsymbol{k}_1^\perp)\bar{B}(\boldsymbol{k}_2^\perp)\bar{C}(\boldsymbol{k}_3^\perp)\rangle = (2\pi)^2\delta^{(2)}(\boldsymbol{k}_1^\perp + \boldsymbol{k}_2^\perp + \boldsymbol{k}_3^\perp)b^{\text{FLAT}}(k_1^\perp, k_2^\perp, k_3^\perp) \tag{A19}$$

with

$$b^{\text{FLAT}}(k_1^\perp, k_2^\perp, k_3^\perp) = \prod_i \int \mathrm{d}x_i^\parallel \frac{\mathrm{d}k_i^\parallel}{2\pi} F_i(x_i^\parallel)e^{ik_i^\parallel x_i^\parallel}(2\pi)\delta(k_1^\parallel + k_2^\parallel + k_3^\parallel)B\left(\sqrt{k_1^{\parallel 2} + k_1^{\perp 2}}, \sqrt{k_2^{\parallel 2} + k_2^{\perp 2}}, \sqrt{k_3^{\parallel 2} + k_3^{\perp 2}}\right). \tag{A20}$$

Expanding $B\left(\sqrt{k_1^{\parallel 2} + k_1^{\perp 2}}, \sqrt{k_2^{\parallel 2} + k_2^{\perp 2}}, \sqrt{k_3^{\parallel 2} + k_3^{\perp 2}}\right)$ about $k_i^\parallel$ we obtain

$$b^{\text{FLAT}}(k_1^\perp, k_2^\perp, k_3^\perp) = \prod_i \int \mathrm{d}x F_A(x)F_B(x)F_C(x)B(k_1, k_2, k_3)\left(1 + \frac{1}{2(k_1 x)^2}\frac{\mathrm{d}\ln F_A}{\mathrm{d}\ln x}\frac{\mathrm{d}(\ln F_B \ln F_C)}{\mathrm{d}\ln x}\frac{\mathrm{d}\ln B}{\mathrm{d}\ln k_1}\right.$$
$$\left. + \frac{1}{2(k_2 x)^2}\frac{\mathrm{d}\ln F_B}{\mathrm{d}\ln x}\frac{\mathrm{d}(\ln F_A \ln F_C)}{\mathrm{d}\ln x}\frac{\mathrm{d}\ln B}{\mathrm{d}\ln k_2} + \frac{1}{2(k_3 x)^2}\frac{\mathrm{d}(\ln F_A \ln F_B)}{\mathrm{d}\ln x}\frac{\mathrm{d}\ln F_B}{\mathrm{d}\ln x}\frac{\mathrm{d}\ln B}{\mathrm{d}\ln k_3}\right). \tag{A21}$$

Expanding $1/x^4 B(\frac{\nu_1}{x}, \frac{\nu_2}{x}, \frac{\nu_3}{x})$ in equation A14 around the peak, $\bar{r}$, of $F_A(x)F_B(x)F_C(x)$ we find $k^4 b^{\text{FLAT}}(k_1^\perp, k_2^\perp, k_3^\perp) \approx \ell^4 b^{\text{FULL}}_{\ell_1,\ell_2,\ell_3}$ with $l_i + 1/2 = k_i \bar{r}$. Note that as in the power spectrum case, the convergence of these series depend both on $\ell + 1/2$ and on the projection kernels, if the projection kernels are peaked at different r_i then more terms will be required in this series.

B. DETAILED BISPECTRUM TEMPLATES

In this appendix we briefly summarize the origin and bispectrum structure of the various significant secondary sources of non-Gaussianity. While the kSZ effect is too small to be constrained with the data used in this work we discuss the form of its bispectrum for completeness.

In this section we define a deconvolved reduced bispectrum, $\tilde{b}$, where we have factored out the beam and pixelization function, as

$$b^{X_1,X_2,X_3}_{\ell_1,\ell_2,\ell_3} \equiv \mathrm{w}^{X_1}_{\ell_1}\mathrm{w}^{X_2}_{\ell_3}\mathrm{w}^{X_2}_{\ell_3}\tilde{b}^{X_1,X_2,X_3}_{\ell_1,\ell_2,\ell_3} \tag{B22}$$

and for notational convenience we suppress the map indicies $\{X_1, X_2, X_3\}$ on the reduced bispectrum.

B.1. *Radio galaxies*

Radio galaxies are point sources with strong synchrotron emission and include AGN, flat-spectrum radio quasars (FSRQs) and BL Lacertae type objects (BL Lacs). It has been shown (Toffolatti et al. 1998; González-Nuevo et al. 2005) that at CMB frequencies radio galaxies are unclustered. Thus radio galaxies follow Poisson statistics and their reduced bispectrum is given by:

$$\tilde{b}^{\text{radio}-\text{radio}-\text{radio}}_{\ell_1,\ell_2,\ell_3} = k_\nu^3 \sum_i \int_0^{S_c} \mathrm{d}S_\nu^{(i)}\, S_\nu^{(i)3}\frac{\mathrm{d}n^{(i)}}{\mathrm{d}S\mathrm{d}\Omega}, \tag{B23}$$

where we sum over the different populations of radio galaxies, $S_\nu^{(i)}$ is the flux density, $\mathrm{d}n^{(i)}/\mathrm{d}S\mathrm{d}\Omega$ is the number count per steradian per flux interval, S_c is the flux cut and k_ν is the conversion from flux to map temperature. We compute the theoretical level and frequency dependence using the model described in de Zotti et al. (2010). The model has three components, each described by their own number counts and with flux densities $S \propto \nu^{-\alpha}$. The components are flat-spectrum radio quasars and BL Lacs with $\alpha_{\text{FSRQ}} = \alpha_{\text{BLLac}} = 0.1$, and a population of steep spectrum sources, arising from AGNs, with $\alpha_{\text{steep}} = 0.8$. We extend de Zotti et al. (2010) by using their model for differential source counts at 100 GHz and steepening the spectral indices of the three components in agreement with the results from Tucci et al. (2011). The bispectrum model presented is a simple generalization from the power spectrum for radio point sources which has the form

$$C_\ell^{\text{radio}} = k_\nu^2 \sum_i \int_0^{S_c} \mathrm{d}S_\nu^{(i)}\, S_\nu^{(i)2}\frac{\mathrm{d}n^{(i)}}{\mathrm{d}S\mathrm{d}\Omega}. \tag{B24}$$

The bispectrum is weighted more towards the most luminous galaxies than the power spectrum and, given the relatively slow decrease of the source counts as a function of flux ($dn/dSd\Omega \propto S^{-1.5}$) the bispectrum of radio galaxies is dominated by the brightest sources and thus is sensitive to the level of point source masking.

B.2. *Thermal Sunyaev Zel'dovich*

Light propagating from the surface of last scattering to the present day can be up-scattered by hot gas in a process know as the thermal Sunyaev Zel'dovich effect (tSZ). The magnitude of this effect depends upon the integrated line-of-sight electron pressure, therefore this effect is dominated by galaxy groups and clusters where there is a large amount of hot intercluster gas. The observable signature depends on the frequency of observation; it can be seen as a decrement below frequencies of 220 GHz and an excess at higher frequencies. We briefly review the tSZ bispectrum as developed in Bhattacharya et al. (2012). It should be noted that we only include effects from the one-halo term as previous work (Komatsu & Kitayama 1999) demonstrated that for the angular scales of interest in the power spectrum the two halo term is negligible. We have verified that this extends to the bispectrum. In the non-relativistic approximation the temperature fluctuation at angular position $\boldsymbol{\theta}$ from the center of a cluster of mass M at redshift z is proportional to the integral along the line of site of the electron pressure:

$$
\begin{aligned}
\frac{\Delta T(\boldsymbol{\theta}, M, z)}{T} &= f(x_\nu) y(\boldsymbol{\theta}, M, z) \\
&= f(x_\nu) \frac{\sigma_T}{m_e c^2} \int_{\text{LOS}} dr P_e(\sqrt{r^2 + D_A(z)^2 |\boldsymbol{\theta}|^2}, M, z)
\end{aligned}
\tag{B25}
$$

where $f(x_\nu) = x_\nu(\coth(x_\nu/2) - 4)$ is the spectral function of the SZ effect, $x_\nu = k_b T_{\text{CMB}}/h\nu$ with ν the observing frequency, $D_A(z)$ is the angular diameter distance, and P_e is the electron pressure. The Fourier transform of the projected SZ profile is (Komatsu & Seljak 2002):

$$
\tilde{y}(\ell, M, z) = \frac{4\pi r_s}{\ell_s^2} \frac{\sigma_T}{m_e c^2} \int_0^\infty dx\, x^2 P_e(M, Z, x) \frac{\sin(x\ell/\ell_s)}{x\ell/\ell_s},
\tag{B26}
$$

where $x = r/r_s$ with r_s a characteristic radius of the cluster which we take to be $r_s = r_{\text{vir}}/c_{\text{NFW}}$ following Navarro et al. (1997), $\ell_s = r_s/D_A(z)$. As is consistent with the literature we have included the projection (a) and Limber terms ($1/\chi^2$) in our definition of $\tilde{y}(\ell, M, z)$. The reduced bispectrum can be calculated by summing the cube of the Fourier transform of the projected SZ-profile weighted by the mass function giving:

$$
\tilde{b}_{\ell_1, \ell_2, \ell_3}^{\text{tSZ-tSZ-tSZ}} = f_\nu(x_{\nu_1}) f_\nu(x_{\nu_2}) f_\nu(x_{\nu_3}) \int_0^\infty dz \frac{d^2 V}{dz d\Omega} \int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{y}(\ell_1, M, z) \tilde{y}(\ell_2, M, z) \tilde{y}(\ell_3, M, z),
\tag{B27}
$$

where $\frac{dn}{d\ln M}$ is the halo mass function (for which we used Tinker et al. 2008) and $\frac{dV}{dzd\Omega}$ is the comoving volume per steradian. In this work we use cluster pressure profiles obtain from simulations (Battaglia et al. 2012b) and we verified our results are not significantly changed when using observationally constrained profiles (Arnaud et al. 2010; Planck Collaboration 2013). The tSZ bispectrum is a complementary probe to the tSZ power spectrum, which has the following structure:

$$
\begin{aligned}
C_\ell^{\text{tSZ}} =\ & f_\nu(x_{\nu_1}) f_\nu(x_{\nu_2}) \int_0^\infty dz \frac{d^2 V}{dz d\Omega} \left[\int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{y}(\ell, M, z) \tilde{y}(\ell, M, z) \right. \\
& \left. + \int_0^\infty d\ln M \frac{dn}{d\ln M} b(M)\, \tilde{y}(\ell, M, z) \int_0^\infty d\ln M' \frac{dn}{d\ln M} b(M')\, \tilde{y}(\ell, M', z) P_{\text{lin}}\left(\frac{\ell + 1/2}{\chi(z)}, z\right) \right],
\end{aligned}
\tag{B28}
$$

where $\chi(z)$ is the comoving distance, $b(M)$ is the linear halo bias and $P_{\text{lin}}(k)$ is the linear theory matter power spectrum. We have included both the one and two halo terms. They are complementary as the bispectrum is dominated by lower redshift and more massive clusters than the power spectrum (Bhattacharya et al. 2012).

B.3. *Dusty star-forming Galaxies*

Dusty star-forming galaxies (DSFGs) are high redshift ($z \sim 2 - 3$), dusty, star-forming galaxies; the UV radiation emitted by the newly formed massive stars is absorbed by the dust and re-emitted at longer wavelengths. In CMB maps these objects appear as point sources and whilst these are dominant at high frequencies they are still important for the frequencies considered here (Hauser & Dwek 2001). We model these sources with two components: a Poisson term and a clustered component. We use the halo model (Cooray & Sheth 2002) to calculate a theoretical DSFG template. Modeling the DSFGs in the halo allows us to simply

calculate DSFG cross-correlations. In this we follow the work of Lacasa et al. (2014), Addison et al. (2012) and Xia et al. (2012) and we combine the halo model with halo occupation statistics to populate dark matter halos with DSFGs (Zheng et al. 2005). We use the assumption that the properties of the DSFGs depend only upon redshift and use the parametric model by Béthermin et al. (2012) to model the source properties. This model incorporates two populations of galaxies, main sequence and starburst, with redshift evolution. This parametric model was fit to differential counts of mm objects from *Herschel* and *Spitzer* data (Béthermin et al. 2010; Berta et al. 2011a; Glenn et al. 2010). The two and three halo terms are only important on the largest scales (Addison et al. 2012; Lacasa et al. 2014) and have been measured (Planck Collaboration 2014f), thus in this work we focus only on the one-halo and Poisson terms.

For a detailed description of this bispectrum we refer the reader to Lacasa et al. (2014) and here we summarize their results. We define the nth order differential source intensity as :

$$\frac{dI^{(n)}(\nu_1, \nu_2..., \nu_n)}{dV} = \int\limits_0^{S_{cut}} dS_{\nu_1} S_{\nu_1} S_{\nu_2}...S_{\nu_n} \frac{d^2 n(S_{\nu_1}, z)}{dVdS}, \tag{B29}$$

S_{ν_i} is the observer frame flux at frequency ν_i, S_{cut} is the flux above which all sources are masked and $d^2 n(S_{\nu_1}, z)/dVdS$ is the differential count and is obtained from the Béthermin et al. (2012) model. Note that this can be related to the generalized differential source emissivity $j^{(n)}(\nu_1, \nu_2..., \nu_n)$ (Knox et al. 2001) used in Lacasa et al. (2014) by

$$\frac{dI^{(n)}(\nu_1, \nu_2..., \nu_n)}{dV} = \frac{a^n}{\chi^{2n}} j^{(n)}(\nu_1, \nu_2..., \nu_n); \tag{B30}$$

in other words we absorbed the projection operator (a) and the Limber terms$(1/\chi^2)$ into our differential source count. Through the halo occupation model prescription we have

$$\langle n_{\text{gal}} \rangle = \langle N_{\text{cen}} \rangle + \langle N_{\text{sat}} \rangle \tag{B31}$$

with

$$\langle N_{\text{cen}} \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log(M) - \log(M_{\text{min}})}{\sigma_{\log(M)}} \right) \right] \tag{B32}$$

and

$$\langle N_{\text{sat}} \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log(M) - \log(2M_{\text{min}})}{\sigma_{\log(M)}} \right) \right] \left(\frac{M}{M_{\text{sat}}} \right)^{\alpha_{\text{sat}}}, \tag{B33}$$

where M is the halo mass, and M_{min}, $\sigma_{\log(M)}$, M_{sat} and α_{sat} parameterize the occupation of the halos. M_{min} gives the mass at which a halo has a 50% probability of having a central galaxy and $\sigma_{\log(M)}$ describes the width of the transition from no central galaxy to one central galaxy. M_{sat} and α_{sat} describe the number of satellite galaxies. The parameters for the HOD model were fit to *Planck* power spectra in Pénin et al. (2014) and we use these values in this work. Finally we assume that the distribution of the DSFGs follows that of the dark matter profile in the halo. Thus the Fourier transform of the 3D DSFG spatial distribution, $\tilde{u}_{\text{DSFGs}}$ is

$$\tilde{u}_{\text{DSFG}}(\ell, M, z) = \tilde{u}_{\text{dm}}(\ell, M, z) = \int\limits_0^{r_{vir}} dr \frac{\sin((\ell + 1/2)r/\chi)}{(\ell + 1/2)r/\chi} \frac{\rho(r|M, z)}{M} \tag{B34}$$

where $\tilde{u}_{\text{dm}}(\ell, M, z)$ is the Fourier transformed of the normalised halo profile. Bringing all these pieces together we can write down the DSFG contributions to the bispectrum, for which there are three components. The three galaxy term

$$\tilde{b}_{\ell_1, \ell_2, \ell_3}^{\text{DSFG}-3\text{gal}} = \int_0^\infty dz \frac{d^2 V}{dz d\Omega} \frac{dI^{(1),\text{DSFG}}(\nu_1)}{dV} \frac{dI^{(1),\text{DSFG}}(\nu_2)}{dV} \frac{dI^{(1),\text{DSFG}}(\nu_3)}{dV} \tag{B35}$$

$$\int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{u}_{\text{dm}}(\ell_1, M, z) \tilde{u}_{\text{dm}}(\ell_2, M, z) \tilde{u}_{\text{dm}}(\ell_3, M, z) \frac{\langle n_{\text{gal}}(n_{\text{gal}} - 1)(n_{\text{gal}} - 2) \rangle}{\bar{n}_{\text{gal}}^3}, \tag{B36}$$

the two galaxy term,

$$\tilde{b}^{\text{DSFG}-2\text{gal}}_{\ell_1,\ell_2,\ell_3} = \int_0^\infty dz \frac{d^2 V}{dz d\Omega} \frac{dI^{(2),\text{DSFG}}(\nu_1,\nu_2)}{dV} \frac{dI^{(1),\text{DSFG}}(\nu_3)}{dV} \int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{u}_{\text{dm}}(\ell_1,M,z)\tilde{u}_{\text{dm}}(|\boldsymbol{\ell}_2+\boldsymbol{\ell}_3|,M,z)\frac{\langle n_{\text{gal}}(n_{\text{gal}}-1)\rangle}{\bar{n}_{\text{gal}}^2},$$

(B37)

and the one galaxy term :

$$\tilde{b}^{\text{DSFG}-1\text{gal}}_{\ell_1,\ell_2,\ell_3} = \int_0^\infty dz \frac{d^2 V}{dz d\Omega} \frac{dI^{(3),\text{DSFG}}(\nu_1,\nu_2,\nu_3)}{dV}.$$

(B38)

Typically the two terms are called the clustered component and the last term is called the Poisson component. In this work we fix the relative amplitude of these three contributions and fit for one overall amplitude, hereafter called the DSFG-DSFG-DSFG component. The source count for the Poisson component decrease rapidly as a function of flux and as such are largely insensitive to the cut level used. We also express the n_{gal} averages in terms of N_{cen} and N_{sat}. Assuming that N_{sat} follows a Poisson distribution, we obtain the following (see Appendix E):

$$\langle n_{\text{gal}}(n_{\text{gal}}-1)\rangle = 2\langle N_{\text{sat}}\rangle + \langle N_{\text{sat}}\rangle^2$$

(B39)

and

$$\langle n_{\text{gal}}(n_{\text{gal}}-1)(n_{\text{gal}}-2)\rangle = 3\langle N_{\text{sat}}\rangle^2 + \langle N_{\text{sat}}\rangle^3.$$

(B40)

The halo formalism allows the cross-correlation between the DSFGs and the tSZ effect to be easily calculated:

$$\tilde{b}^{\text{tSZ}-\text{DSFGs}-\text{DSFGs}}_{\ell_1,\ell_2,\ell_3} = \int_0^\infty dz \frac{d^2 V}{dz d\Omega} \left[f(x_{\nu_1}) \frac{dI^{(1),\text{DSFG}}(\nu_2)}{dV} \frac{dI^{(1),\text{DSFG}}(\nu_3)}{dV} \int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{u}_{\text{dm}}(\ell_1,M,z)\tilde{u}_{\text{dm}}(\ell_2,M,z)\tilde{y}(\ell_3,M,z) \right.$$
$$\left. \frac{\langle n_{\text{gal}}(n_{\text{gal}}-1)\rangle}{\bar{n}_{\text{gal}}^2} + f(x_{\nu_1}) \frac{dI^{(2),\text{DSFG}}(\nu_1,\nu_2)}{dV} \int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{u}_{\text{dm}}(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|,M,z)\tilde{y}(\ell_3,M,z)\frac{\langle n_{\text{gal}}\rangle}{\bar{n}_{\text{gal}}^1} \right]$$

(B41)

and

$$\tilde{b}^{\text{tSZ}-\text{tSZ}-\text{DSFGs}}_{\ell_1,\ell_2,\ell_3} = \int_0^\infty dz \frac{d^2 V}{dz d\Omega} f(x_{\nu_1})f(x_{\nu_2}) \frac{dI^{(1),\text{DSFG}}(\nu_3)}{dV} \int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{u}_{\text{dm}}(\ell_1,M,z)\tilde{y}(\ell_2,M,z)\tilde{y}(\ell_3,M,z)\frac{\langle n_{\text{gal}}\rangle}{\bar{n}_{\text{gal}}}.$$

(B42)

B.4. *Cross-bispectra among radio galaxies, the tSZ effect and DSFGs*

Whilst the radio galaxies show little clustering, recent work by Gralla et al. (2014); Gupta et al. (2017) has found evidence for a correlation between radio galaxies and tSZ sources, since radio galaxies occupy the same clusters that source the tSZ effect. Building on this we consider the cross-bispectra that arise from correlations among the tSZ effect, DSFGs and radio galaxies. There are five possible bispectra: radio-tSZ-tSZ, radio-radio-tSZ, radio-DSFG-DSFG, radio-radio-DSFG and radio-DSFG-tSZ. In this work we fix the relative amplitude of the radio-tSZ-tSZ and radio-radio-tSZ terms, hereafter this joint term is called the radio-tSZ template, and the relative amplitude of the radio-DSFG-DSFG, radio-radio-DSFG terms, hereafter the joint term is called the radio-DSFG template. These terms were combined as they cannot be distinguished with our current data sets.

To estimate the radio halo occupation, we use the model from Wake et al. (2008), which is also reexamined in Smolčić et al. (2011). In Wake et al. (2008), the authors constrain HOD parameters with data from a set of combined surveys (see Sadler et al. (2007) for more details). They are unable to constrain the parameters associated with satellite galaxies, implying most halos host only one radio galaxy, and as such we use only a simplified HOD model:

$$n_{\text{radio}} = e^{-\frac{M^*}{M}},$$

(B43)

where M is the halo mass and $M^* = 9.65 \times 10^{13} h^{-1} M_\odot$ was as measured in Wake et al. (2008). We combine this model with the halo model and force the radio galaxies to lie in the center of the cluster as was preferred in Smolčić et al. (2011) and Wake et al. (2008) so that $\tilde{r}(\ell,M,z) = 1$. Using this we can write down the template for the radio-tSZ term

$$\tilde{b}^{\text{radio}-\text{tSZ}}_{\ell_1,\ell_2,\ell_3} = f(x_{\nu_1})f(x_{\nu_2}) \int_0^\infty dz \frac{d^2 V}{dz d\Omega} dz \frac{dI^{(1),\text{radio}}(\nu_3)}{dV} \int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{y}(\ell_1,M,z)\tilde{y}(\ell_2,M,z)\tilde{r}(\ell_3,M,z)\langle\frac{n_{\text{radio}}(M)}{\bar{n}_{\text{radio}}}\rangle$$
$$+ f(x_{\nu_1}) \int_0^\infty dz \frac{d^2 V}{dz d\Omega} \frac{dI^{(2),\text{radio}}(\nu_1,\nu_2)}{dV} \int_0^\infty d\ln M \frac{dn}{d\ln M} \tilde{y}(\ell_1,M,z)\tilde{r}(|\boldsymbol{\ell}_2+\boldsymbol{\ell}_3|,M,z)\langle\frac{n_{\text{radio}}(M)}{\bar{n}_{\text{radio}}}\rangle.$$

(B44)

Similarly, the radio-DSFG template is:

$$
\begin{aligned}
\tilde{b}^{\text{radio}-\text{DSFG}}_{\ell_1,\ell_2,\ell_3} =& \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \mathrm{d}z \frac{\mathrm{d}I^{(1),\text{DSFG}}(\nu_1)}{\mathrm{d}V} \frac{\mathrm{d}I^{(1),\text{DSFG}}(\nu_2)}{\mathrm{d}V} \frac{\mathrm{d}I^{(1),\text{radio}}(\nu_3)}{\mathrm{d}V} \\
& \times \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{u}_{\text{dm}}(\ell_1,M,z)\tilde{u}_{\text{dm}}(\ell_2,M,z)\tilde{r}(\ell_3,M,z) \langle \frac{n_{\text{DSFG}}(M)(n_{\text{DSFG}}(M)-1)}{\bar{n}^2_{\text{DSFG}}}\rangle \langle \frac{n_{\text{radio}}(M)}{\bar{n}_{\text{radio}}}\rangle \\
& + \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(2),\text{DSFG}}(\nu_1,\nu_2)}{\mathrm{d}V} \frac{\mathrm{d}I^{(1),\text{radio}}(\nu_1)}{\mathrm{d}V} \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{u}_{\text{dm}}(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|,M,z)\tilde{r}(\ell_3,M,z)\langle \frac{n_{\text{DSFG}}(M)}{\bar{n}_{\text{DSFG}}} \frac{n_{\text{radio}}(M)}{\bar{n}_{\text{radio}}}\rangle \\
& + \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\text{DSFG}}(\nu_1)}{\mathrm{d}V} \frac{\mathrm{d}I^{(2),\text{radio}}(\nu_2,\nu_3)}{\mathrm{d}V} \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{u}_{\text{dm}}(\ell_1,M,z)\tilde{r}(|\boldsymbol{\ell}_2+\boldsymbol{\ell}_3|,M,z)\langle \frac{n_{\text{DSFG}}(M)}{\bar{n}_{\text{DSFG}}} \frac{n_{\text{radio}}(M)}{\bar{n}_{\text{radio}}}\rangle,
\end{aligned}
$$
$$\tag{B45}$$

and the final template, the radio-DSFG-tSZ term, is

$$
\tilde{b}^{\text{radio}-\text{DSFG}-\text{tSZ}}_{\ell_1,\ell_2,\ell_3} = f(x_{\nu_1}) \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\text{DSFG}}(\nu_2)}{\mathrm{d}V} \frac{\mathrm{d}I^{(1),\text{radio}}(\nu_3)}{\mathrm{d}V} \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{y}(\ell_1)\tilde{u}(\ell_2)\tilde{r}(\ell_2)\langle \frac{n_{\text{DSFG}}(M)}{\bar{n}_{\text{DSFG}}} \frac{n_{\text{radio}}(M)}{\bar{n}_{\text{radio}}}\rangle.
$$
$$\tag{B46}$$

At the power spectrum level it is very challenging to separate the contributions from tSZ-DSFG, radio-tSZ and radio-DSFG terms, which have the following power spectra:

$$
\begin{aligned}
C_\ell^{\text{tSZ}-\text{DSFG}} =& f_\nu(x_{\nu_1}) \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\text{radio}}(\nu_2)}{\mathrm{d}V} \left[\int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{y}(\ell,M,z)\tilde{r}(\ell,M,z)\langle \frac{n_{\text{DSFG}}(M)}{\bar{n}_{\text{DSFG}}}\rangle \right. \\
& \left. + \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} b(M)\, \tilde{y}(\ell,M,z) \int_0^\infty \mathrm{d}\ln M' \frac{\mathrm{d}n}{\mathrm{d}\ln M} b(M')\, \tilde{r}(\ell,M',z)\langle \frac{n_{\text{DSFG}}(M')}{\bar{n}_{\text{DSFG}}}\rangle P_{\text{lin}}\left(\frac{\ell+1/2}{\chi(z)},z\right) \right]
\end{aligned}
$$
$$\tag{B47}$$

$$
\begin{aligned}
C_\ell^{\text{radio}-\text{tSZ}} =& f_\nu(x_{\nu_1}) \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\text{radio}}(\nu_2)}{\mathrm{d}V} \left[\int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{y}(\ell,M,z)\tilde{r}(\ell,M,z)\langle \frac{n_{\text{radio}}(M)}{\bar{n}_{\text{radio}}}\rangle \right. \\
& \left. + \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} b(M)\, \tilde{y}(\ell,M,z) \int_0^\infty \mathrm{d}\ln M' \frac{\mathrm{d}n}{\mathrm{d}\ln M} b(M')\, \tilde{r}(\ell,M',z)\langle \frac{n_{\text{radio}}(M')}{\bar{n}_{\text{radio}}}\rangle P_{\text{lin}}\left(\frac{\ell+1/2}{\chi(z)},z\right) \right],
\end{aligned}
$$
$$\tag{B48}$$

and

$$
\begin{aligned}
C_\ell^{\text{radio}-\text{DSFG}} =& \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\text{DSFG}}(\nu_1)}{\mathrm{d}V} \frac{\mathrm{d}I^{(1),\text{radio}}(\nu_2)}{\mathrm{d}V} \left[\int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{u}_{\text{dm}}(\ell,M,z)\tilde{r}(\ell,M,z)\langle \frac{n_{\text{DSFG}}(M)}{\bar{n}_{\text{DSFG}}} \frac{n_{\text{radio}}(M)}{\bar{n}_{\text{radio}}}\rangle \right. \\
& \left. + \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} b(M)\, \tilde{u}_{\text{dm}}(\ell,M,z)\langle \frac{n_{\text{DSFG}}(M)}{\bar{n}_{\text{DSFG}}}\rangle \int_0^\infty \mathrm{d}\ln M' \frac{\mathrm{d}n}{\mathrm{d}\ln M} b(M')\, \tilde{r}(\ell,M',z)\langle \frac{n_{\text{radio}}(M')}{\bar{n}_{\text{radio}}}\rangle P_{\text{lin}}\left(\frac{\ell+1/2}{\chi(z)},z\right) \right].
\end{aligned}
$$
$$\tag{B49}$$

The strength of the cross-correlations between these sources is still poorly understood. Small-scale power spectrum measurements provide constraints on these terms (George et al. 2015; Dunkley et al. 2013) whilst the tSZ-DSFG cross-correlation at large scales has been explored in Planck Collaboration (2016f). Our model predicts a correlation coefficient between the DSFGs and the tSZ effect of 0.28 at $\ell = 3000$. This is higher than was seen in George et al. (2015) who found $0.113^{+0.057}_{-0.051}$, but consistent with other previous work.

B.5. *Cross-bispectra between lensing and secondary sources*

As was described in section 3.3 we can decompose the temperature anisotropies, having expanded the effect of lensing to first order in ϕ, as

$$
\Delta T(\boldsymbol{n}) = \Delta T^P(\boldsymbol{n}) + \nabla\phi \cdot \nabla\Delta T^P(\boldsymbol{n}) + \Delta T^s(\boldsymbol{n}).
$$
$$\tag{B50}$$

and this leads to a bispectrum with the form

$$
\tilde{b}^{\text{lensing}-\text{sec.}}_{\ell_1,\ell_2,\ell_3} = -\boldsymbol{\ell}_1 \cdot \boldsymbol{\ell}_2 C^{TT}_{\ell_1} C^{\phi S}_{\ell_2} + 5 \text{ permutations}.
$$
$$\tag{B51}$$

Lewis et al. (2011) showed that the effect of higher order terms in ϕ lead to $\sim 10\%$ corrections to the perturbative result in equation B51 and they showed these higher order terms could accurately be approximated by replacing C^{TT} with the lensed temperature power spectrum. Unlike the other bispectra listed here, this source of non-Gaussianity has a significant contribution from polarized maps through the lensing term. In this work we only analyze temperature maps so do not not considered the polarized contribution in this work.

In the following subsections we describe our models of the cross-correlations between the lensing potential and the secondary sources. The two halo term has a larger contribution to the power spectrum than the three-point function so we include its contribution in these terms.

B.5.1. *Thermal Sunyaev Zel'dovich lensing cross-correlation*

The cross-correlation between the lensing potential and thermal SZ effect is calculated in Hill & Spergel (2014) and Battaglia et al. (2015) and has two contributions, the one and two halo terms are

$$
C_\ell^{\phi-t\mathrm{SZ}} = \int \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \int \mathrm{d}M \frac{\mathrm{d}n(M,z)}{\mathrm{d}\ln M} \tilde{y}(\ell, M, z) \tilde{\phi}(\ell, M, z)
$$
$$
+ \int \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \int \mathrm{d}M \frac{\mathrm{d}n(M,z)}{\mathrm{d}\ln M} b(M) \, \tilde{y}(\ell, M, z) \int \mathrm{d}M' \frac{\mathrm{d}n(M',z)}{\mathrm{d}\ln M'} b(M') \, \tilde{\phi}(\ell, M', z) P_{\mathrm{lin}}\left(\frac{\ell + 1/2}{\chi(z)}, z\right) \tag{B52}
$$

where $\tilde{\phi}$ is the lensing potential profile for a cluster

$$
\tilde{\phi}(\ell, M, z) = \frac{2}{\ell(\ell+1)} \frac{4\pi r_{s,\phi}}{l_{s,\phi}^2} \int \mathrm{d}x_\phi x_\phi^2 \frac{\sin((\ell + 1/2)x_\phi/l_{s,\phi})}{(\ell + 1/2)x_\phi/l_{s,\phi}} \frac{\rho(x_\phi r_{s,\phi}, M, z)}{\Sigma_{crit}(z)}, \tag{B53}
$$

$r_{s,\phi}$ is a characteristic scale radius for the cluster, in our case the NFW scale radius, ρ is the cluster density, given by the NFW profile, Σ_{crit} is the critical surface density for lensing at redshift z

$$
\Sigma_{crit}^{-1}(z) = \frac{4\pi G \chi(z)(\chi_* - \chi(z))}{c^2 \chi_*(1+z)}, \tag{B54}
$$

where χ_* is the comoving distance to the surface of last scattering, and the other terms are as in section B.2.

B.5.2. *DSFGs lensing cross-correlation*

The DSFG-lensing cross-correlation has been detected (e.g. Omori et al. 2017; van Engelen et al. 2015; Planck Collaboration 2014d). We use the halo model for the DSFGs as described in section B.3 to model the lensing-DSFGs cross-correlation. Through this formalism we obtain the following relations

$$
C_\ell^{\phi-\mathrm{DSFG}} = \int \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\mathrm{DSFG}}(\nu_1)}{\mathrm{d}V} \int \mathrm{d}M \frac{\mathrm{d}n(M,z)}{\mathrm{d}\ln M} \tilde{u}_{\mathrm{dm}}(\ell, M, z) \tilde{\phi}(\ell, M, z) \langle \frac{n_{\mathrm{gal}}(M)}{\bar{n}_{\mathrm{gal}}} \rangle
$$
$$
+ \int \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\mathrm{DSFG}}(\nu_1)}{\mathrm{d}V} \int \mathrm{d}M \frac{\mathrm{d}n(M,z)}{\mathrm{d}\ln M} b(M) \, \tilde{u}_{\mathrm{dm}}(\ell, M, z) \langle \frac{n_{\mathrm{gal}}(M)}{\bar{n}_{\mathrm{gal}}} \rangle \int \mathrm{d}M' \frac{\mathrm{d}n(M',z)}{\mathrm{d}\ln M'} b(M') \, \tilde{\phi}(\ell, M', z) P_{lin}\left(\frac{\ell + 1/2}{\chi(z)}, z\right). \tag{B55}
$$

B.5.3. *Radio–Lensing cross-correlation*

Given the discussion above in section B.4 we should expect a cross-correlation between the ϕ field and radio sources. This will have the same structure as the DSFG-lensing cross-correlation:

$$
C_\ell^{\phi-\mathrm{radio}} = \int \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\mathrm{radio}}(\nu_1)}{\mathrm{d}V} \int \mathrm{d}M \frac{\mathrm{d}n(M,z)}{\mathrm{d}\ln M} \tilde{u}_{\mathrm{dm}}(\ell, M, z) \tilde{\phi}(\ell, M, z) \langle \frac{n_{\mathrm{radio}}(M)}{\bar{n}_{\mathrm{radio}}} \rangle
$$
$$
+ \int \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(1),\mathrm{radio}}(\nu_1)}{\mathrm{d}V} \int \mathrm{d}M \frac{\mathrm{d}n(M,z)}{\mathrm{d}\ln M} b(M) \, \tilde{u}_{\mathrm{dm}}(\ell, M, z) \langle \frac{n_{\mathrm{radio}}(M)}{\bar{n}_{\mathrm{radio}}} \rangle \int \mathrm{d}M' \frac{\mathrm{d}n(M',z)}{\mathrm{d}\ln M'} b(M') \, \tilde{\phi}(\ell, M', z) P_{lin}\left(\frac{\ell + 1/2}{\chi(z)}, z\right). \tag{B56}
$$

B.5.4. *ISW-Lensing cross-correlation*

As was noted in Verde & Spergel (2002) and seen in Planck Collaboration (2014c, 2016j), secondary CMB anisotropies caused by the late time decay of the gravitational potential, the integrated Sachs Wolfe effect (Sachs & Wolfe 1967), and their growth through non-linear structure (Rees & Sciama 1968) are correlated with CMB lensing leading to bispectrum with the above form. This bispectrum is primarily important on the largest scales and so we expect very limited sensitivity to this bispectrum with ACTPol. We use CAMB to calculate the lensing-ISW cross-correlation, $C^{ISW-\phi}$ and combine this with equation B51 to form a bispectrum template.

B.6. *The kinetic Sunyaev Zel'dovich effect*

The kinetic Sunyaev Zel'dovich effect (kSZ) arises due to CMB photons being scattered from moving gas and this motion imparts a Doppler shift to the CMB photons. This effect can be either positive or negative (depending on the direction of the motion of gas) and so in the bispectrum analysis only kSZ squared terms are non vanishing. In this section we outline an approximate template for the kSZ bispectrum and discuss its measurability with our data set. This approach to measure the kSZ bispectrum is complimentary to other methods such as those discussed in Hill et al. (2016); Ferraro et al. (2016); Doré et al. (2004); DeDeo et al. (2005).

We use the halo model to calculate simplified one-halo templates for cross-correlations of the kSZ squared with other tracers. We stress that these are only approximate templates and that there will be significant contributions from two and three halo terms, which are not considered here. The temperature shift caused by the kSZ effect is given by (Ostriker & Vishniac 1986; Sunyaev & Zeldovich 1972)

$$\frac{\Delta T(\boldsymbol{n})}{T} = -\sigma_T \int \frac{\mathrm{d}\chi}{1+z} e^{-\tau(z)} n_e(\chi\boldsymbol{n}, \chi) \frac{\boldsymbol{v}_e}{c} \cdot \boldsymbol{n} \tag{B57}$$

where $\tau(z)$ is the optical depth, n_e is the electron number density and $\boldsymbol{v}_e$ is the electron velocity. In our one-halo approximation the deconvolved reduced bispectrum arising from the kSZ cross secondary is

$$\tilde{b}^{kSZ-kSZ-X}_{\ell_1,\ell_2,\ell_3} = \sigma_T^2 \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z\mathrm{d}\Omega} \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} \tilde{n}_e(M, z, \ell_1) \tilde{n}_e(M, z, \ell_2) \tilde{X}(M, z, \ell_3) \frac{\sigma_{vc}^2(M)}{3} e^{-2\tau(z)} \tag{B58}$$

where $\tilde{n}_e(M, z, \ell_1)$ is the projected line of sight integral of the electron number density, $\sigma_{vc}(M)$ is the cluster velocity dispersion and $\tilde{X}(M, z, \ell_3)$ is $f(\nu)\tilde{y}(M, z, \ell)$ for the kSZ-kSZ-tSZ bispectrum and $\frac{\mathrm{d}I}{\mathrm{d}z}\tilde{u}(M, z, \ell)$ for the kSZ-kSZ-DSFGs bispectrum. We approximate the cluster velocity dispersion as the velocity of the halo given by peaks theory as (Bardeen et al. 1986)

$$\sigma_{vc}(M) = \sigma_v(M) \sqrt{1 - \frac{\sigma_0^4}{\sigma_1^2 \sigma_{-1}^2}}, \tag{B59}$$

where:

$$\sigma_v = (faH)^2 \sigma_{-1}, \tag{B60}$$

$$\sigma_j = \frac{1}{2\pi} \int \mathrm{d}k P(k) W(Rk) k^{2j+2}, \tag{B61}$$

and $f = \mathrm{d}\ln(g)/\mathrm{d}\ln(a)$ is the derivative of the growth function. The optical depth is given by the integral of the mean ionized electron number density to the halo

$$\tau(z) = \sigma_T \int \frac{\mathrm{d}\chi}{1+z} \bar{n}_e(z). \tag{B62}$$

To close these equations we need the electron number density profile and, motivated by the results in Schaan et al. (2016), we consider a Gaussian distribution for the electrons. The width of the Gaussian is given by the virial radius and we normalize the profile by fixing the baryon mass to be given by $f_b M_{dm}$ and then using the mean molecular weight and the mean molecular weight per free electron to convert to the total number of electrons.

We then performed a fisher forecast to assess their measurability. In table 3 we present the expected errors from fitting just the three kSZ templates separately, fitting a combined kSZ template and jointly fitting these templates with the templates considered in section 3. When we fit the combined template we assume that the relative amplitude of the three different kSZ templates is fixed and that only a possible overall scaling can exist. This is a simplification but the results show that even with such simplifications we have no constraining power current for measuring the kSZ effect; as such we exclude this template from our analysis. It should

Type	One σ A_i	
	Single Fit	Joint Fit
kSZ-kSZ-tSZ	2.2	51.8
kSZ-kSZ-DSFG	3.2	38.1
kSZ-kSZ-radio	6.7	89.0
Combined kSZ Template	3.30	31.2

Table 3. The one sigma errors for the kinetic Sunyaev Zel'dovich templates. We see that whilst the sensitivities for the individual templates is close to having constraining power, when the templates are jointly fit the similarity of the templates to our data sets leaves no constraining power.

be noted that if we could break the degeneracies between these templates and the other secondary source templates, such that the joint fit and single fit errors are similar, then future experiments should be able to start constraining these terms. The degeneracy between these templates will be reduced for experiments with multifrequency small-scale measurements.

C. PIPELINE VALIDATION TESTS

C.1. *Validation Templates*

In this section we describe the suite of tests that we have preformed to verify the accuracy and efficiency of our analysis. To validate our pipeline we used three types of bispectra that can be easily simulated; these are Poisson noise, pseudo-local and primordial local non-Gaussianity. We did not investigate the amplitudes of these templates in the real data and only used them for validation. These types were chosen due to their ease of simulation, which can be challenging for arbitrary non-Gaussianity.

To cross check the analysis pipeline described in section 5 (hereafter pipeline 1) we implemented a second pipeline based on Das et al. (2009) (hereafter pipeline 2). For most of these analyses we limit the ℓ range of our analysis to be $\ell < 4000$, as Pipeline 2 is restricted to these scales.

C.2. *Non-Gaussian Simulations*

In order to validate our pipelines we need to verify that we can accurately measure non-Gaussianity in the maps. In general we use the method described in Smith & Zaldarriaga (2011), hereafter called the indirect method, to generate maps with arbitrary bispectra. We briefly summarize this method here. First we generate a Gaussian map (as in 2.2) and then perform the following operation to its Fourier coefficients (a_ℓ^L):

$$a_{\boldsymbol{\ell}}^{XNL} = a_{\boldsymbol{\ell}}^{XL} + \frac{1}{6} \sum_{X_1, X_2} \int d^2\ell_1 d^2\ell_2 4\pi^2 \delta^{(2)}(\boldsymbol{\ell} + \boldsymbol{\ell}_1 + \boldsymbol{\ell}_2) b_{\boldsymbol{\ell}, \boldsymbol{\ell}_1, \boldsymbol{\ell}_2}^{X, X_1, X_2} C_{\boldsymbol{\ell}_1}^{-1 X_1} a^L C_{\boldsymbol{\ell}_2}^{-1 X_2} a^L \tag{C63}$$

where $b_{\boldsymbol{\ell}, \boldsymbol{\ell}_i, \boldsymbol{\ell}_j}$ is the reduced bispectrum of the non-Gaussianity that you wish to generate. This method is useful for testing our pipelines although higher order correlations are not accurately reproduced. The method of generating this non-Gaussianity is similar to the method used to calculate the normalization of our estimator, and so this method would not be able to identify errors that jointly affect this method and the normalization calculation.

As a cross check of this method and for further robustness of our analysis pipeline we generate several types of non-Gaussianity via direct methods. For this, we consider the three templates Poisson, pseudo-local and local non-Gaussianity.

C.2.1. *Poisson Noise*

The Poisson noise bispectrum has a trivial structure:

$$\tilde{b}_{\ell_1, \ell_2, \ell_3}^{\text{const}} = \text{Const.} \tag{C64}$$

This template is the same as the radio point source term or the DSFG poisson component, except we have removed any frequency dependence to simplify the template. Poisson non-Gaussianity is simulated by adding Poisson noise to the maps.

C.2.2. *Pseudo-Local bispectrum*

This template is inspired by the primordial local bispectrum and it is generated by adding $A_{\text{pseudo-local}}/T_{CMB}^2 (\Delta T(\boldsymbol{n}) - \bar{\Delta T})^2$ to the map. This has a bispectrum of the following form:

$$\tilde{b}_{\ell_1, \ell_2, \ell_3}^{\text{pseudo-local}} = 2C(\ell_1)C(\ell_2) + 3 \text{ permutations.} \tag{C65}$$

This type of non-Gaussianity is not physical. We use it to validate our pipelines due to the simplicity of simulating it and as it is orthogonal to the other test templates.

C.2.3. *Primordial Local non-Gaussianity*

Primordial local non-Gaussianity is a physical primordial type of non-Gaussianity, so-called as it generated by the real space interactions of fields and is thus local. Physically it arises when two short wavelength modes interact with a longer wavelength mode. The local primordial three point structure is given by (Falk et al. 1993; Gangui et al. 1994; Verde et al. 2000; Wang & Kamionkowski 2000; Komatsu & Spergel 2001)

$$B_\Phi^{\text{local}}(k_1, k_2, k_3) = 2f_{local}^{\text{NL}}(P_\Phi(k_1)P_\Phi(k_2) + P_\Phi(k_1)P_\Phi(k_3) + P_\Phi(k_2)P_\Phi(k_3))$$

$$= 2A^2 f_{\text{local}}^{NL}\left(\frac{1}{k_1^{(4-n_s)}k_2^{(4-n_s)}} + \frac{1}{k_1^{(4-n_s)}k_3^{(4-n_s)}} + \frac{1}{k_2^{(4-n_s)}k_3^{(4-n_s)}}\right), \tag{C66}$$

where n_s is the primordial spectral tilt and A is the amplitude of fluctuations. From this we obtain the reduced bipsectrum as

$$\tilde{b}_{\ell_1,\ell_2,\ell_3}^{\text{prim}} = \frac{8}{\pi^3}\int \mathrm{d}r\, r^2 \prod_i \left(\int \mathrm{d}k_i\, k_i^2 g_\ell^{X_i}(k_i) j_\ell(k_i r)\right) B_\Phi(k_1, k_2, k_3), \tag{C67}$$

where $g_\ell^{X_i}(k_i)$ is the transfer function that can be computed with a Boltzmann code such as CAMB (Lewis et al. 2000).

Local non-Gaussianity can be generated by first generating a realization of primordial Gaussian fluctuations,($\Phi(\boldsymbol{x})$ and then modifying the primordial fluctuations in the following manner:

$$\Phi^{NL}(\boldsymbol{x}) = \Phi(\boldsymbol{x}) + A_{\text{local}}\left(\Phi^2(\boldsymbol{x}) - \langle\Phi^2(\boldsymbol{x})\rangle\right). \tag{C68}$$

We then apply the transfer functions to evolve these fluctuations to the present. We use algorithms described in Elsner & Wandelt (2009) and Liguori et al. (2003) to efficiently generate non-Gaussian maps at high resolution. We use their method to generate full-sky maps and then cut out a patch that is appropriate for our fields. This method exploits the fact that the a_{lm} are related to primordial fluctuations in the following manner

$$a_{lm} = \int \mathrm{d}r r^2 \Phi_{lm}(r)\alpha_l(r), \tag{C69}$$

where $\alpha_l(r)$ are the real space transfer function and $\Phi_{lm}(r)$ are harmonic transforms of the real space primordial potential fluctuations. We draw $\Phi_{lm}(r)$ from a Gaussian distribution, with the appropriate radial correlations, perform a spherical harmonic transform of these, square them, subtract the variance and transform back. Then we use equation C69 to evolve these perturbations and attain the late time a_{lm}. This quantity is then added back to the original map scaled to the appropiate f_{NL} to add local non Gaussianity. We then transform these to a HEALpix map (Górski et al. 2005) and cut out a patch to use in our flat-sky analysis. We refer the reader to Liguori et al. (2003) for further details on this method.

C.3. *Pre-whitening Analysis Pipeline*

This pipeline is based on Das et al. (2009) and uses the same real space and Fourier masks as described in section 5. We pre-whiten the maps by using a disc differencing method; this involves convolving two copies of the maps with discs of radius $R = 1.5'$ and $3R$ and then differencing them. We then add back a fraction of the original maps to these maps and then convolve the maps with a Gaussian. This pre-whitening reduces the mode coupling when we apply the point source and the mask (as above). When we move to Fourier space we deconvolve these effects to attain the desired maps. This method is limited to probing a reduced range of ℓ space as deconvolving the disc-differencing involves a sinc function, and the signal near the zero of the sinc function cannot be accurately recovered. For this reason we restrict the use of this method to testing our pipelines at $\ell < 4000$.

C.4. *Fisher Information Comparisons*

It is important to compare how the variance of our estimators compares to the theoretical limits. A Fisher analysis shows that the expected variance of our estimator should be (in the diagonal covariance case)

$$\sigma^2(\mathrm{A_{sec}}^\alpha) = F_{\alpha,\alpha}^{-1}$$

Type	Max ℓ	σ_{Fisher}	σ_{measured} Pipeline 1	σ_{measured} Pipeline 2
Poisson Noise	3000	8.2	8.0	8.8
Poisson Noise	7500	0.19	0.21	–
Pseudo Local	3000	65	68	87
Local	3000	430	420	450
tSZ-tSZ-tSZ	3000	0.23	0.23	0.22
tSZ-tSZ-tSZ	7500	0.085	0.086	–

Table 4. A comparison between the expected and measured standard deviations of our estimator when applied to simulated Gaussian maps. The consistency of our estimator is a check of the optimality of our implementation, especially our approximate inverse covariance filtering.

$$F_{\alpha,\beta} = \frac{f_{\text{sky}}^{(3)} f_{\text{sky}}^{(3)}}{6} \int \prod_i d^2\boldsymbol{\ell}_i (2\pi)^2 \delta(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) \frac{b_{\ell_1,\ell_2,\ell_3} b_{\ell_1,\ell_2,\ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \tag{C70}$$

We generate sets of 100 Gaussian simulations of maps as would be measured by the ACTPol PA2, including beam effects and anisotropic noise as described section 2.2. In table 4 we show the expected standard deviation of our estimators against the measured standard deviation. As mentioned above pipeline 2 only works for $\ell < 4000$ and so when we consider $\ell > 4000$ we limit our analysis to pipeline 1. We see generally good agreement between the different methods, with pipeline 2 having slightly higher variance. There is close agreement between our estimator variances with the Fisher estimations, which implies that our methods are close to satisfying the Cramér-Rao bound.

C.5. *Simulated non-Gaussian Maps*

It is important to verify that our estimators are able to accurately recover known levels of non-Gaussianity. To do this we first consider the three types that we can directly simulate: the constant, pseudo-local and local form and using the methods described in section C.2 we simulated maps with known levels of non-Gaussianity. We compared these direct methods to non-Gaussianity generated by the method described in Smith & Zaldarriaga (2011) and section C.2. By confirming the accuracy of this method, we can then use it to verify the accuracy of estimators on the remaining templates, which are much more challenging to simulate directly.

In figure 7 we compare the accuracy of our simulation methods for several different types of non-Gaussianity with various different cut offs obtained from sets of 100 simulations of PA2 observations. We see that there is generally good agreement between the different methods and pipelines. However we see that the residuals are predominately slightly negative, this indicates a slight bias in our analysis at the level of $\sim 1\%$. We believe this arises from ignoring the linear term in our normalization. For the measurements reported here this effect is not significant but for future precision measurements this should be accounted for. It should be noted that the variances of the two non-Gaussian simulation methods for large non-Gaussianity is different. This is because the method in Smith & Zaldarriaga (2011) accurately reproduces the power spectrum and bispectrum but does not reproduce the higher order statistics, which results in an inaccurate variance for the bispectrum (as the variance of the bispectrum is a six point function).

Another useful diagnostic is to ensure that our pipelines measure similar levels of non-Gaussianity on individual maps, as well as for the statistical ensemble. In figure 8 we view the recovered amplitudes, A_i, for our local and constant estimator on a map-by-map basis for our two analysis pipelines. We see that in general there is reasonable agreement between the two methods. Together, these tests give us confidence that our estimators are unbiased and efficient.

D. TEMPLATE ERRORS

When the non-Gaussianity is weak the variance on the template is given by the inverse of the normalization term. However as we have detections of these templates, and further as we wish to estimate the template amplitude rather than just the significance of the deviation from null, it is important to either verify that these errors are still dominant or, in the case that they are subdominant, calculate the correct error bars. For simplicity we will consider the case when we only have a single map, but the generalization
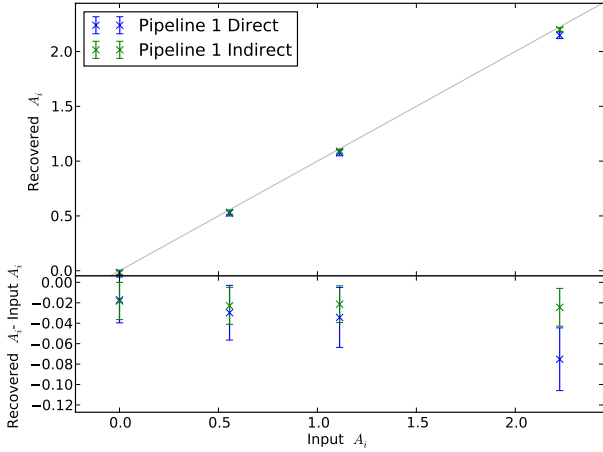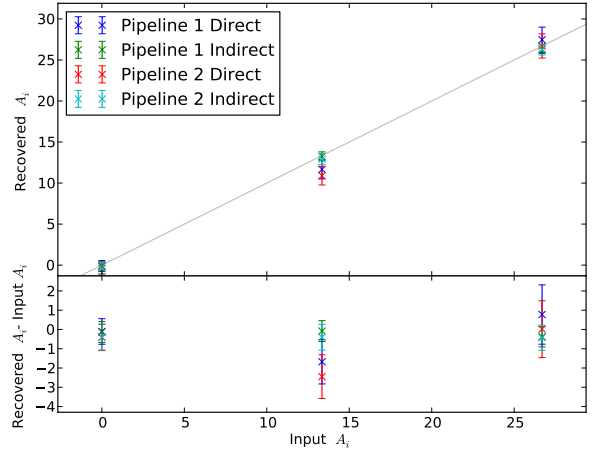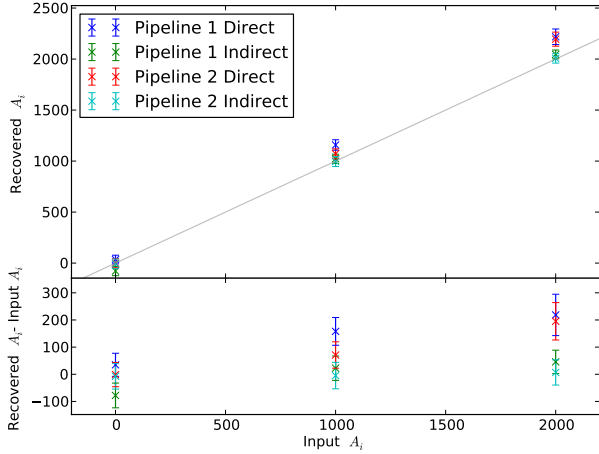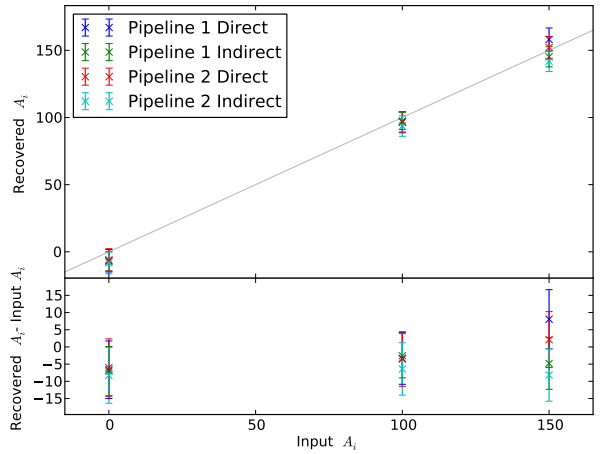
(a) Constant non-Gaussianity with $\ell_{\mathrm{max}} = 7500$.

(b) Constant non-Gaussianity with $\ell_{\mathrm{max}} = 3000$.

(c) Local non-Gaussianity with $\ell_{\mathrm{max}} = 3000$.

(d) Pseudo local non-Gaussianity with $\ell_{\mathrm{max}} = 3000$.

Figure 7. A set of graphs demonstrating how our pipelines can recover an input level of non-Gaussianity, input A_i, for a set of different templates with different max ℓ. The amplitudes, A_i, are overall scalings of our templates and so are dimensionless. We compare our two methods of simulating non-Gaussian templates, the "direct" method and the "indirect" method, as described in section C.2. Having verified that our estimators are unbiased on this set of templates we then use the "indirect" method to validate our full range of templates, which cannot easily be verified by direct simulation methods.

is simple. From our estimator the full covariance is given as:

$$\langle (\hat{A}_i - \hat{A}_i^{\mathrm{mean}})(\hat{A}_j - \hat{A}_j^{\mathrm{mean}}) \rangle = \sum_{i',j'} \mathcal{N}_{i,i'}^{-1} \mathcal{N}_{j,j'}^{-1} \sum_{\boldsymbol{\ell}_k, t_i} \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) \delta^{(2)}(\boldsymbol{\ell}_4 + \boldsymbol{\ell}_5 + \boldsymbol{\ell}_6) b_{\ell_1,\ell_2,\ell_3}^{i'} b_{\ell_4,\ell_5,\ell_6}^{j'} C_{\boldsymbol{\ell}_1}^{-1} C_{\boldsymbol{\ell}_2}^{-1} C_{\boldsymbol{\ell}_3}^{-1} C_{\boldsymbol{\ell}_4}^{-1} C_{\boldsymbol{\ell}_5}^{-1} C_{\boldsymbol{\ell}_6}^{-1}$$

$$\left(C_{\boldsymbol{\ell}_1} C_{\boldsymbol{\ell}_2} C_{\boldsymbol{\ell}_3} \delta(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_4) \delta(\boldsymbol{\ell}_2 + \boldsymbol{\ell}_5) \delta(\boldsymbol{\ell}_3 + \boldsymbol{\ell}_6) + \delta(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_4) b_{\ell_1,\ell_2,\ell_3}^{t_i} b_{\ell_4,\ell_5,\ell_6}^{t_j} + \delta(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_4) C_{\boldsymbol{\ell}_1} T_{\ell_2,\ell_3,\ell_5,\ell_6} \right.$$

$$\left. + S_{\ell_1,\ell_2,\ell_3,\ell_4,\ell_5,\ell_6} + V_{\ell_1,\ell_2,\ell_3,\ell_4,\ell_5,\ell_6}^{SSV} + \mathrm{permutations} \right) \tag{D71}$$

where S is the fully connected six point function, T is the trispectrum and V^{SSV} is the halo sample variance or the super sample variance (Hamilton et al. 2006). The first term is the 'Gaussian' term described in equation 17. To calculate these contributions
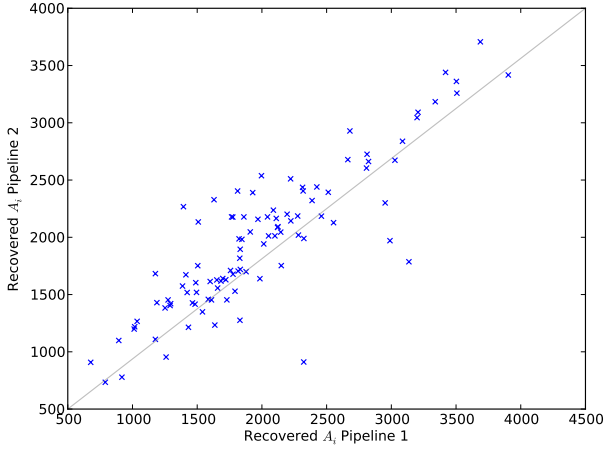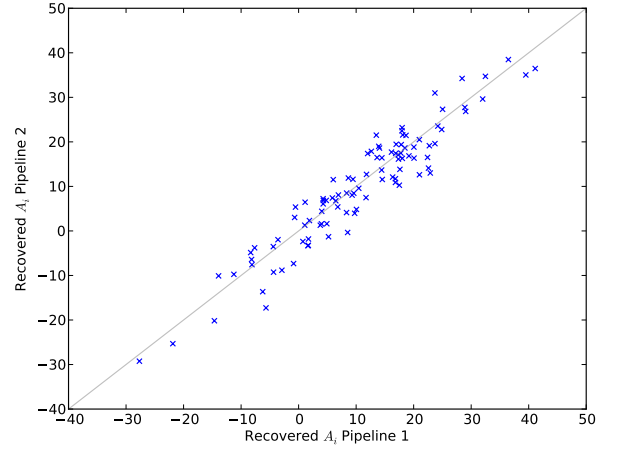
(a) Local non-Gaussianity with $\ell_{\max} = 3000$.

(b) Poisson non-Gaussianity with $\ell_{\max} = 3000$,

Figure 8. A map-by-map comparison of our estimators applied to maps with local type and Poisson noise non-Gaussianity simulated with the direct method. The input level of non-Gaussianity for local type this is $A_{\text{local}} = 2000$ and for Poisson noise this is $A_{\text{Poisson}} = 10.6$. It can be seen the two pipelines agree well on a map by map basis, increasing the confidence in the accuracy of our pipelines.

we use the halo model and keep only the one-halo term. This results in the following from for the trispectrum:

$$T_{\ell_2,\ell_3,\ell_5,\ell_6} = \sum_{i,j,k,l} \int_0^\infty dz \frac{d^2V}{dzd\Omega} \int_0^\infty d\ln M \frac{dn}{d\ln M} X^i_{\ell_2} X^j_{\ell_3} X^k_{\ell_5} X^l_{\ell_6} \tag{D72}$$

where X^i_ℓ is the Fourier transform of the integrated line of sight halo property, *i.e.* X^1 could be the Compton Y parameter and X^2 could be the DSFG profile etc. We sum over all the contributions to the bispectrum from the tSZ effect, DSFGs and radio galaxies (including both the Poisson and clustered terms). Similarly for the six point function:

$$S_{\ell_1,\ell_2,\ell_3,\ell_4,\ell_5,\ell_6} = \sum_{i,j,k,l,m,n} \int_0^\infty dz \frac{d^2V}{dzd\Omega} \int_0^\infty d\ln M \frac{dn}{d\ln M} X^i_{\ell_1} X^j_{\ell_2} X^k_{\ell_3} X^l_{\ell_4} X^m_{\ell_5} X^n_{\ell_6} \tag{D73}$$

where we again sum over all of the contributions listed above. The super sample variance term can be though of additional variance induced by the presence of modes longer than the survey volume. For example, if the region is located near a peak of such a long mode then more regions will have density fluctuations greater than the level required for collapse, resulting in more halos. We refer the reader to Takada & Hu (2013) and Schaan et al. (2014) for a more detailed description of the super sample variance term. As is shown in Kayo et al. (2013) and Schaan et al. (2014) the one-halo term is most important for the bispectrum and has the following form:

$$V^{SSV}_{\ell_1,\ell_2,\ell_3,\ell_4,\ell_5,\ell_6} = \sum_{i,j,k,l,m,n} \int_0^\infty dz \frac{d\sigma_s^2(V_s)}{\Delta z} \frac{d^2V}{dzd\Omega} \int_0^\infty d\ln M \frac{dn}{d\ln M} b(m) X^i_{\ell_1} X^j_{\ell_2} X^k_{\ell_3} \int_0^\infty d\ln M \frac{dn}{d\ln M} b(m) X^l_{\ell_4} X^m_{\ell_5} X^n_{\ell_6} \tag{D74}$$

where $b(m)$ is the linear bias and $\sigma_s(V_s)$ encapsulates the finite survey volume by:

$$\frac{d\sigma_s^2(V_s)}{\Delta z} = \frac{1}{\Delta z 4\pi^2} \int d^2k |W(k)|^2 P_{lin}(k,z) \tag{D75}$$

where $W(k)$ is the 3D Fourier transform of the 3D survey window function.

Analytical calculation of the projection of bispectra onto the six point terms is challenging and so in this work we use an ensemble average method to calculate these terms. In particular we need to exploit some form of factorization as the six point matrix $S_{\ell_1,\ell_2,\ell_3,\ell_4,\ell_5,\ell_6}$ for modern surveys is extremely large ($O(10^{40})$ elements for our data) and so cannot be calculated fully. We use methods inspired by the non-Gaussian simulation methods described in Smith & Zaldarriaga (2011) and section C.2 . As is done in Smith & Zaldarriaga (2011) and above in equation C63 we first calculate :

$$a_\ell^{NL} = \frac{1}{2} \sum_{X_1,X_2} \int d^2\ell_1 d^2\ell_2 4\pi^2 \delta^{(2)}(\ell + \ell_1 + \ell_2) b_{\ell,\ell_1,\ell_2} C^{-1}_{\ell_1} a C^{-1}_{\ell_2} a \tag{D76}$$

This has the property that $\langle a_{\boldsymbol{\ell}_1} a_{\boldsymbol{\ell}_2} a_{\boldsymbol{\ell}_3}^{NL} \rangle = \delta^{(2)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) b_{\ell_1, \ell_2, \ell_3}$. We calculate this for two different Gaussian realizations, $a^{(1)}, a^{(2)}$ and then calculate the following quantity:

$$\int \prod_{i=1}^{6} \mathrm{d}^2 \ell\, S_{\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6} C_{\boldsymbol{\ell}_1}^{-1} a^{(1)} C_{\boldsymbol{\ell}_2}^{-1} a^{(1)} C_{\boldsymbol{\ell}_3}^{-1} a^{NL(1)} C_{\boldsymbol{\ell}_4}^{-1} a^{(2)} C_{\boldsymbol{\ell}_5}^{-1} a^{(2)} C_{\boldsymbol{\ell}_6}^{-1} a^{NL(2)} \tag{D77}$$

and similar terms for the other components of the six point function. As with the calculation of the estimator, we compute this by utilizing the separability of the integrand of the double integral. This means for each redshift and mass in our numerical calculation of S (or the other six point terms) we factorize the above quantity and efficiently calculate the sums by FFTs. For example we would evaluate the term in equation D73 using the following:

$$\sigma_{6pnt}^2 \propto \sum_{i,j,k,l,m,n} \sum_{\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_4, \boldsymbol{\ell}_5, \boldsymbol{\ell}_6} \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} X_{\ell_1}^i C_{\boldsymbol{\ell}_1}^{-1} a^{(1)} X_{\ell_2}^j C_{\boldsymbol{\ell}_2}^{-1} a^{(1)} X_{\ell_3}^k \times$$
$$C_{\boldsymbol{\ell}_3}^{-1} a^{(1)} X_{\ell_4}^l C_{\boldsymbol{\ell}_4}^{-1} a^{(2)} X_{\ell_5}^m C_{\boldsymbol{\ell}_5}^{-1} a^{(2)} X_{\ell_6}^n C_{\boldsymbol{\ell}_6}^{-1} a^{(2)}. \tag{D78}$$

It is easily seen that the ensemble average of equation D77 generates the corresponding term in equation D71. The calculation of the nth order HOD moments required for the DSFG terms is described in Appendix E. By using two different Gaussian simulations we find that the majority of the contributions converge very rapidly, to within 10 percent for a single pair of simulations. We found that a handful of contributions require large numbers of simulations to converge; these terms arise from Poisson contributions to the trispectrum term such as:

$$C_{\ell_1, \ell_4} T_{\ell_2, \ell_3, \ell_5, \ell_6} = \sum_l \int_0^\infty \mathrm{d}z \frac{\mathrm{d}^2 V}{\mathrm{d}z \mathrm{d}\Omega} \frac{\mathrm{d}I^{(3)}(\nu_1, \nu_2 \nu_5)}{\mathrm{d}z} \int_0^\infty \mathrm{d}\ln M \frac{\mathrm{d}n}{\mathrm{d}\ln M} u(\boldsymbol{\ell}_2 + \boldsymbol{\ell}_3 + \boldsymbol{\ell}_5) X_{\ell_6}^l. \tag{D79}$$

As these terms are sub dominant, we neglect their contribution so that we can use orders of magnitude fewer simulations to obtained converged errors.

E. N^{TH} ORDER HOD MOMENTS

When calculating n point functions of DSFG fluctuations we need expectations of drawing n different galaxies from a cluster, *i.e.*:

$$\langle N(N-1)(N-2)..(N-(n-1)) \rangle \tag{E80}$$

Simulations show that the number of galaxies in a cluster has first a step, then a plateau and then a power law (Kravtsov et al. 2004; Berlind et al. 2003). This is understood as the galaxies being split into two: a central and satellite galaxies (with the central galaxy not necessarily being in the center). As the mass of the cluster increases the probability of having a central galaxy increases, this is the step and plateau, above a certain mass the cluster can also host satellites, whose number follows a power law. The HOD model parameterizes the probability of central and satellite galaxies. Only clusters with central galaxies can host satellite galaxies, so the probability distribution of satellite is conditional on the central galaxies, ie:

$$P(N_s) = \sum_{N_c} P(N_s|N_c)P(N_c) = P(N_s|N_c = 1)P(N_c = 1), \tag{E81}$$

where N_c is the number of central galaxies and N_s is the number of satellites. We model the distribution of central galaxies as a Bernoulli distribution with mass dependent mean:

$$\langle N_c(M) \rangle = \frac{1}{2} \left(1 + \mathrm{erf} \left(\frac{\log M - \log M_{\min}}{\sigma_M} \right) \right), \tag{E82}$$

where $M_{\min}$ and σ_M parameterize the step function. We model the number of satellite galaxies as Poisson distributed with mass dependent mean:

$$\langle N_s(M) \rangle = \frac{1}{2} \left(1 + \mathrm{erf} \left(\frac{\log M - \log(2M_{\min})}{\sigma_M} \right) \right) (M/M_{\text{sat}})^\alpha \tag{E83}$$

where M_{sat} and α parameterize the power law. With these pieces we can compute the desired expectations values. It can be shown that for $n \geq 2$ the expectation values can be written in the following form:

$$\langle N(N-1)(N-2)..(N-(n-1)) \rangle = \langle (N_c + N_s)(N_c + N_s - 1)(N_c + N_s - 2)..(N_c + N_s - (n-1)) \rangle$$
$$= \left\langle f(N_c, N_s)N_c(N_c - 1) + nN_c \prod_{i=i}^{i=n-2}(N_s - (i-1)) + \prod_{i=i}^{i=n-1}(N_s - (i-1)) \right\rangle. \tag{E84}$$

The first term in the above always vanishes as $N_c \in 0, 1$ and using the Poisson statistics of the satellites we find:

$$\langle N(N-1)(N-2)...(N-(n-1)) \rangle = \langle N_c \rangle n \langle N_s \rangle^{n-1} + \langle N_s \rangle^n \tag{E85}$$

REFERENCES

Abazajian, K. N., et al. 2016, `arXiv:1610.02743`, ArXiv e-prints

Addison, G. E., Dunkley, J., & Spergel, D. N. 2012, `arXiv:1204.5927`, MNRAS, 427, 1741

Arnaud, M., Pratt, G. W., Piffaretti, R., Böhringer, H., Croston, J. H., & Pointecouteau, E. 2010, `arXiv:0910.1234`, A&A, 517, A92

Babich, D., Creminelli, P., & Zaldarriaga, M. 2004, `astro-ph/0405356`, JCAP, 8, 9

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15

Bartolo, N., Matarrese, S., & Riotto, A. 2006, `astro-ph/0604416`, JCAP, 6, 024

—. 2007, `astro-ph/0610110`, JCAP, 1, 019

Battaglia, N., Bond, J. R., Pfrommer, C., & Sievers, J. L. 2012a, `arXiv:1109.3709`, ApJ, 758, 74

—. 2012b, `arXiv:1109.3711`, ApJ, 758, 75

Battaglia, N., Ferraro, S., Schaan, E., & Spergel, D. 2017, `arXiv:1705.05881`, ArXiv e-prints

Battaglia, N., Hill, J. C., & Murray, N. 2015, `arXiv:1412.5593`, ApJ, 812, 154

Benson, B. A., et al. 2014, `arXiv:1407.2973`, in Proc. SPIE, Vol. 9153, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII, 91531P

Berlind, A. A., et al. 2003, `astro-ph/0212357`, ApJ, 593, 1

Bernardeau, F. & Kofman, L. 1995, `astro-ph/9403028`, ApJ, 443, 479

Berta, S., Magnelli, B., Nordon, R., Lutz, D., & Pep Team. 2011a, in Astronomical Society of the Pacific Conference Series, Vol. 446, Galaxy Evolution: Infrared to Millimeter Wavelength Perspective, ed. W. Wang, J. Lu, Z. Luo, Z. Yang, H. Hua, & Z. Chen, 309

Berta, S., et al. 2011b, `arXiv:1106.3070`, A&A, 532, A49

Béthermin, M., Dole, H., Beelen, A., & Aussel, H. 2010, `arXiv:1001.0896`, A&A, 512, A78

Béthermin, M., et al. 2012, `arXiv:1208.6512`, ApJL, 757, L23

Bhattacharya, S., Nagai, D., Shaw, L., Crawford, T., & Holder, G. P. 2012, The Astrophysical Journal, 760, 5

Blanchard, A. & Schneider, J. 1987, A&A, 184, 1

Calabrese, E., Smidt, J., Amblard, A., Cooray, A., Melchiorri, A., Serra, P., Heavens, A., & Munshi, D. 2010, `arXiv:0909.1837`, PhRvD, 81, 043529

Chen, X. 2010, `arXiv:1002.1416`, Advances in Astronomy, 2010, 638979

Cooray, A., Hu, W., & Tegmark, M. 2000, `astro-ph/0002238`, ApJ, 540, 1

Cooray, A. & Sheth, R. 2002, `astro-ph/0206508`, PhR, 372, 1

Crawford, T. M., et al. 2014, `arXiv:1303.3535`, ApJ, 784, 143

Creminelli, P., Nicolis, A., Senatore, L., Tegmark, M., & Zaldarriaga, M. 2006, `astro-ph/0509029`, JCAP, 5, 4

Das, S., Hajian, A., & Spergel, D. N. 2009, Phys. Rev. D, 79, 083008

Das, S., et al. 2011, The Astrophysical Journal, 729, 62

De Bernardis, F., et al. 2017, `arXiv:1607.02139`, JCAP, 3, 008

de Haan, T., et al. 2016, `arXiv:1603.06522`, ApJ, 832, 95

de Zotti, G., Massardi, M., Negrello, M., & Wall, J. 2010, `arXiv:0908.1896`, A&A Rv, 18, 1

DeDeo, S., Spergel, D. N., & Trac, H. 2005, `astro-ph/0511060`, ArXiv Astrophysics e-prints

Doré, O., Hennawi, J. F., & Spergel, D. N. 2004, `astro-ph/0309337`, ApJ, 606, 46

Dunkley, J., et al. 2013, Journal of Cosmology and Astroparticle Physics, 2013, 025

Elbaz, D., et al. 2007, `astro-ph/0703653`, A&A, 468, 33

Elsner, F. & Wandelt, B. D. 2009, The Astrophysical Journal Supplement Series, 184, 264

Elsner, F. & Wandelt, B. D. 2013, `arXiv:1210.4931`, A&A, 549, A111

Falk, T., Rangarajan, R., & Srednicki, M. 1993, `astro-ph/9208001`, ApJL, 403, L1

Ferraro, S., Hill, J. C., Battaglia, N., Liu, J., & Spergel, D. N. 2016, `arXiv:1605.02722`, PhRvD, 94, 123526

Frigo, M., Steven, & Johnson, G. 2005, in Proceedings of the IEEE, 216–231

Gangui, A., Lucchin, F., Matarrese, S., & Mollerach, S. 1994, `astro-ph/9312033`, ApJ, 430, 447

George, E. M., et al. 2015, `arXiv:1408.3161`, ApJ, 799, 177

Glenn, J., et al. 2010, `arXiv:1009.5675`, MNRAS, 409, 109

Gold, B., et al. 2011, `arXiv:1001.4555`, ApJS, 192, 15

Goldberg, D. M. & Spergel, D. N. 1999, `astro-ph/9811251`, PhRvD, 59, 103002

González-Nuevo, J., Toffolatti, L., & Argüeso, F. 2005, The Astrophysical Journal, 621, 1

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, `astro-ph/0409513`, ApJ, 622, 759

Gralla, M. B., et al. 2014, `arXiv:1310.8281`, MNRAS, 445, 460

Gupta, N., et al. 2017, `arXiv:1605.05329`, MNRAS, 467, 3737

Hamilton, A. J. S., Rimes, C. D., & Scoccimarro, R. 2006, `astro-ph/0511416`, MNRAS, 371, 1188

Hand, N., et al. 2012, `arXiv:1203.4219`, Physical Review Letters, 109, 041101

Hasselfield, M., et al. 2013, The Astrophysical Journal Supplement Series, 209, 17

Hasselfield, M., et al. 2013, `arXiv:1301.0816`, JCAP, 7, 008

Hauser, M. G. & Dwek, E. 2001, `astro-ph/0105539`, ARA&A, 39, 249

Henderson, S. W., et al. 2016, `arXiv:1510.02809`, Journal of Low Temperature Physics, 184, 772

Hill, J. C., Ferraro, S., Battaglia, N., Liu, J., & Spergel, D. N. 2016, `arXiv:1603.01608`, Physical Review Letters, 117, 051301

Hill, J. C. & Pajer, E. 2013, `arXiv:1303.4726`, PhRvD, 88, 063526

Hill, J. C. & Sherwin, B. D. 2013, `arXiv:1205.5794`, PhRvD, 87, 023527

Hill, J. C. & Spergel, D. N. 2014, `arXiv:1312.4525`, JCAP, 2, 030

Hill, J. C., et al. 2014, `arXiv:1411.8004`, ArXiv e-prints

Hinshaw, G., et al. 2007, `astro-ph/0603451`, ApJS, 170, 288

Hojjati, A., et al. 2017, `arXiv:1608.07581`, MNRAS, 471, 1565

Hu, W. 2000, Phys. Rev. D, 62, 043007

Juszkiewicz, R., Weinberg, D. H., Amsterdamski, P., Chodorowski, M., & Bouchet, F. 1995, `astro-ph/9308012`, ApJ, 442, 39

Kayo, I., Takada, M., & Jain, B. 2013, `arXiv:1207.6322`, MNRAS, 429, 344

Knox, L., Cooray, A., Eisenstein, D., & Haiman, Z. 2001, `astro-ph/0009151`, ApJ, 550, 7

Komatsu, E. & Kitayama, T. 1999, The Astrophysical Journal Letters, 526, L1

Komatsu, E. & Seljak, U. 2002, `astro-ph/0205468`, MNRAS, 336, 1256

Komatsu, E. & Spergel, D. N. 2001, `astro-ph/0005036`, PhRvD, 63, 063002

Komatsu, E., Spergel, D. N., & Wandelt, B. D. 2005, `astro-ph/0305189`, ApJ, 634, 14

Komatsu, E., et al. 2009, `arXiv:0803.0547`, ApJS, 180, 330

—. 2011, `arXiv:1001.4538`, ApJS, 192, 18

Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottlöber, S., Allgood, B., & Primack, J. R. 2004, `astro-ph/0308519`, ApJ, 609, 35

Lacasa, F., Pénin, A., & Aghanim, N. 2014, `arXiv:1312.1251`, MNRAS, 439, 123

Lewis, A., Challinor, A., & Hanson, D. 2011, `arXiv:1101.2234`, JCAP, 3, 018

Lewis, A., Challinor, A., & Lasenby, A. 2000, `astro-ph/9911177`, Astrophys. J., 538, 473

Liguori, M., Matarrese, S., & Moscardini, L. 2003, The Astrophysical Journal, 597, 57

Liguori, M., Sefusatti, E., Fergusson, J. R., & Shellard, E. P. S. 2010, `arXiv:1001.4707`, Advances in Astronomy, 2010, 980523

Louis, T., et al. 2017, `arXiv:1610.02360`, JCAP, 6, 031

Loverde, M. & Afshordi, N. 2008, `arXiv:0809.5112`, PhRvD, 78, 123506

Munshi, D. & Heavens, A. 2010, `arXiv:0904.4478`, MNRAS, 401, 2406

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, `astro-ph/9611107`, ApJ, 490, 493

Nelson, K., Lau, E. T., Nagai, D., Rudd, D. H., & Yu, L. 2014, `arXiv:1308.6589`, ApJ, 782, 107

Omori, Y., et al. 2017, `arXiv:1705.00743`, ApJ, 849, 124

Osborne, S. J., Hanson, D., & Doré, O. 2014, `arXiv:1310.7547`, JCAP, 3, 024

Ostriker, J. P. & Vishniac, E. T. 1986, ApJL, 306, L51

Pénin, A., Lacasa, F., & Aghanim, N. 2014, `arXiv:1312.1252`, MNRAS, 439, 143

Pettinari, G. W., Fidler, C., Crittenden, R., Koyama, K., Lewis, A., & Wands, D. 2014, `arXiv:1406.2981`, PhRvD, 90, 103010

Pettinari, G. W., Fidler, C., Crittenden, R., Koyama, K., & Wands, D. 2013, `arXiv:1302.0832`, JCAP, 4, 003

Planck Collaboration. 2013, `arXiv:1207.4061`, A&A, 550, A131

—. 2014a, A&A, 571, A6

—. 2014b, `arXiv:1303.5068`, A&A, 571, A7

—. 2014c, `arXiv:1303.5079`, A&A, 571, A19

—. 2014d, A&A, 571, A18

—. 2014e, A&A, 571, A24

—. 2014f, A&A, 571, A30

—. 2016a, A&A, 594, A8

—. 2016b, A&A, 594, A11

—. 2016c, A&A, 594, A13

—. 2016d, A&A, 594, A17

—. 2016e, A&A, 594, A22

—. 2016f, A&A, 594, A23

—. 2016g, A&A, 594, A24

—. 2016h, A&A, 594, A26

—. 2016i, `arXiv:1509.06348`, A&A, 594, A12

—. 2016j, `arXiv:1502.01595`, A&A, 594, A21

—. 2016k, A&A, 586, A140

—. 2017, `arXiv:1707.00132`, ArXiv e-prints

Rasia, E., et al. 2006, `astro-ph/0602434`, MNRAS, 369, 2013

Rees, M. J. & Sciama, D. W. 1968, Nature, 217, 511

Reichardt, C. L., et al. 2012, `arXiv:1111.0932`, ApJ, 755, 70

Reinecke, M. & Seljebotn, D. S. 2013, `arXiv:1303.4945`, A&A, 554, A112

Sachs, R. K. & Wolfe, A. M. 1967, ApJ, 147, 73

Sadler, E. M., et al. 2007, `astro-ph/0612019`, MNRAS, 381, 211

Schaan, E., Takada, M., & Spergel, D. N. 2014, `arXiv:1406.3330`, PhRvD, 90, 123523

Schaan, E., et al. 2016, arXiv:1510.06442, PhRvD, 93, 082002

Senatore, L., Smith, K. M., & Zaldarriaga, M. 2010, arXiv:0905.3746, JCAP, 1, 28

Smith, K. M., Senatore, L., & Zaldarriaga, M. 2009, arXiv:0901.2572, JCAP, 9, 6

Smith, K. M. & Zaldarriaga, M. 2011, Monthly Notices of the Royal Astronomical Society, 417, 2

Smolčić, V., Finoguenov, A., Zamorani, G., Schinnerer, E., Tanaka, M., Giodini, S., & Scoville, N. 2011, arXiv:1108.2952, MNRAS, 416, L31

Soergel, B., et al. 2016, arXiv:1603.03904, MNRAS, 461, 3172

Spergel, D. N. & Goldberg, D. M. 1999, astro-ph/9811252, PhRvD, 59, 103001

Staniszewski, Z., et al. 2009, arXiv:0810.1578, ApJ, 701, 32

Sunyaev, R. A. & Zeldovich, Y. B. 1970, Ap&SS, 7, 3

—. 1972, Comments on Astrophysics and Space Physics, 4, 173

Takada, M. & Hu, W. 2013, arXiv:1302.6994, PhRvD, 87, 123504

Taylor, A. & Watts, P. 2001, Monthly Notices of the Royal Astronomical Society, 328, 1027

Thornton, R. J., et al. 2016, arXiv:1605.06569, ApJS, 227, 21

Timmons, N., Cooray, A., Feng, C., & Keating, B. 2017, arXiv:1707.02298, ApJL, 849, L6

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S., & Holz, D. E. 2008, The Astrophysical Journal, 688, 709

Toffolatti, L., Argüeso Gómez, F., De Zotti, G., Mazzei, P., Franceschini, A., Danese, L., & Burigana, C. 1998, Monthly Notices of the Royal Astronomical Society, 297, 117

Tucci, M., Toffolatti, L., de Zotti, G., & Martínez-González, E. 2011, arXiv:1103.5707, A&A, 533, A57

van Engelen, A., Bhattacharya, S., Sehgal, N., Holder, G. P., Zahn, O., & Nagai, D. 2014, arXiv:1310.7023, ApJ, 786, 13

van Engelen, A., et al. 2015, arXiv:1412.0626, ApJ, 808, 7

Van Waerbeke, L., Hinshaw, G., & Murray, N. 2014, arXiv:1310.5721, PhRvD, 89, 023508

Verde, L., Jimenez, R., Alvarez-Gaume, L., Heavens, A. F., & Matarrese, S. 2013, arXiv:1301.6017, JCAP, 6, 23

Verde, L. & Spergel, D. N. 2002, astro-ph/0108179, PhRvD, 65, 043007

Verde, L., Wang, L., Heavens, A. F., & Kamionkowski, M. 2000, Monthly Notices of the Royal Astronomical Society, 313, 141

Wake, D. A., Croom, S. M., Sadler, E. M., & Johnston, H. M. 2008, arXiv:0810.1050, MNRAS, 391, 1674

Wang, L. & Kamionkowski, M. 2000, astro-ph/9907431, PhRvD, 61, 063504

Wilson, M. J., et al. 2012, arXiv:1203.6633, PhRvD, 86, 122005

Wright, E. L., et al. 2009, arXiv:0803.0577, ApJS, 180, 283

Xia, J.-Q., Negrello, M., Lapi, A., De Zotti, G., Danese, L., & Viel, M. 2012, Monthly Notices of the Royal Astronomical Society, 422, 1324

Yadav, A. P. S. & Wandelt, B. D. 2010, arXiv:1006.0275, Advances in Astronomy, 2010, 565248

Zheng, Z., et al. 2005, astro-ph/0408564, ApJ, 633, 791