Detection of gravitational lensing in the cosmic microwave background

Kendrick M. Smith

Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA

Oliver Zahn

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

Olivier Doré

Canadian Institute for Theoretical Astrophysics, 60 St. George Street, University of Toronto, Toronto, Ontario, Canada M5S 3H8

Gravitational lensing of the cosmic microwave background (CMB), a long-standing prediction of the standard cosmological model, is ultimately expected to be an important source of cosmological information, but first detection has not been achieved to date. We report a 3.4σ detection, by applying quadratic estimator techniques to all sky maps from the Wilkinson microwave anisotropy probe (WMAP) satellite, and correlating the result with radio galaxy counts from the NRAO VLA sky survey (NVSS). We present our methodology including a detailed discussion of potential contaminants. Our error estimates include systematic uncertainties from density gradients in NVSS, beam effects in WMAP, galactic microwave foregrounds, resolved and unresolved CMB point sources, and the thermal Sunyaev-Zel'dovich effect.

I. INTRODUCTION

Within just two decades, cosmology has progressed from a rather speculative science to one of the most successful fields of physics, driven by an exemplary interplay between experiment and theory. Much of this progress has been owing to the well understood physics underlying the cosmic microwave background (CMB) anisotropy, seeded by oscillations in the baryon-photon plasma of the early universe.

Measurements of these fluctuations by a number of experiments have given rise to a basic cosmological paradigm, with the tightest current constraints on the cosmological parameter budget coming from combinations of data from the Wilkinson microwave anisotropy probe (WMAP) satellite [1,2] in conjunction with small-scale CMB experiments (e.g. [3–5]), and other rich probes of cosmological clustering and dynamics such as supernovae, galaxy surveys, the Lyman-alpha forest, weak lensing, and others (e.g. [6–14]).

The CMB promises to remain a gold mine for precision cosmology, and two new frontiers lie ahead. First, a polarized component has recently been detected by a number of groups [15–20], offering e.g. the prospects of detecting primordial gravitational waves and constraining recombination physics.

Second, large-scale structure between the last scattering surface and us alters the primary CMB anisotropy, through gravitational lensing (for a recent review of the theory see [21]), through scattering off hot electrons in large-scale structure (the Sunyaev-Zel'dovich effects) [22,23], and through redshifting during the traverse of time-dependent potential fluctuations (the integrated Sachs-Wolfe, here-after ISW, effect) [24]. A number of specialized instruments will soon begin to study details of these secondary anisotropies [25,26].

As important as constraining cosmological and astrophysical parameters, detecting any of these effects is a crucial milestone for cosmological physics. The Sunyaev-Zel'dovich effect has been found by targeting clusters detected in x-ray [27–30], also at high significance level using WMAP [31], and it has been observed in cross correlation of galaxy surveys with WMAP [32–34]. The ISW effect has been detected by cross correlating WMAP with galaxy surveys and with the hard x-ray background [32,35–43].

A detection of gravitational lensing in the CMB has so far been outstanding. The main difficulty at millimeter wavelengths is the high angular resolution needed, as typical deflection angles over a cosmological volume are only a few arcminutes. Non-Gaussianity imprinted by lensing into the primordial CMB may allow statistical detection with surveys at lower angular resolution, but the signal-to-noise is currently too low for internal detections. Cross correlation with other tracers of large-scale structure offers a way to limit systematics and increase the signal-to-noise.

A first attempt [44] was made by cross correlating the WMAP first year release [1] with data from the Sloan digital sky survey (SDSS) [8]. These authors used a sample of 503 944 SDSS luminous red galaxies (LRG's) overlapping with $\simeq 10\%$ of the sky observed by WMAP. They were not able to find evidence for gravitational lensing within statistical error bounds. While SDSS LRG's have a well understood redshift distribution, their number density drops rapidly beyond $z = 0.5$, and has only marginal over-

lap with the higher redshift range that is geometrically optimal for CMB lensing. Photometric quasars found in SDSS may offer an additional handle.

Here we go a different route, using the 1.9×10^6 radio sources found in the NRAO VLA sky survey (NVSS) [45]. The large sky coverage and estimated depth of NVSS make it an excellent candidate for a search for CMB lensing in cross correlation with WMAP. The survey covers 77% of the sky, 58% of which is found to overlap with WMAP, once masks to limit systematics have been applied.

The structure of this paper is as follows.

First, we describe the data sets (Sec. II), theory (Sec. III), and pipeline (Sec. IV) that will be used for detecting CMB lensing by reconstructing the lensing potential from WMAP, and cross correlating the result to NVSS. The detection is shown, with statistical errors only, in Fig. 5. The rest of the paper is devoted to null tests and assigning systematic errors: NVSS systematics (Sec. V), WMAP beam effects and galactic foregrounds (Sec. VI), resolved and unresolved point sources (Sec. VII), and Sunyaev-Zel'dovich fluctuations (Sec. VIII). We quote our final result including systematic errors (Fig. 19) in Sec. IX, where we also mention future directions.

In our calculations we will assume throughout the cosmological model favored by a combination of WMAP, smaller-scale CMB experiments, and other data (the WMAP + ALL analysis [2]): a local expansion rate $H_0 = 70.4$ km/s/Mpc, primordial power spectrum slope $n_s = 0.947$, matter and dark energy fractions of $\Omega_0 = 0.267$ and $\Omega_\Lambda = 0.733$, respectively, and amplitude $\sigma_8 = 0.773$.

II. DATA SETS

A. WMAP

With the goal of producing full-sky maps of the CMB with unprecedented accuracy, the WMAP satellite was launched in June 2001. Since then, it has been mapping the sky using 10 differential assemblies (DA's) covering 5 frequency bands centered at 23 (K), 33 (Ka), 41 (Q), 61 (V), and 94 GHz (W). In our analysis we use the 2 Q-band, 2 V-band, and 4 W-band temperature maps produced using 3 years of observations [46] and made publicly available [47]. We will use as a default mask the Kp0 mask, which cuts out the galactic plane and point sources bright enough to be resolved by WMAP, leaving about 78.46% of the sky [1].

The intrinsic quality of this data set leaves us with few instrumental systematic effects to worry about [48–50]. Nonetheless, noise inhomogeneities and beam effects could be of particular concern for our lensing statistic. The former will be optimally handled by our estimator. Although the latter are well controlled for the power spectrum estimation [46,48], they could potentially affect our lensing estimator as will be discussed below. We will show how the formalism presented in [46] allows us to control them in our particular context too. Another source of

systematic error might come from other astrophysical sources, namely, residual galactic foregrounds (synchrotron, free-free, and dust), residual point sources and the signature of galaxy clusters via the Sunyaev-Zel'dovich effect. These potential contaminants will be discussed in later sections.

B. NVSS

As a tracer of the large-scale density field, we use observations resulting from the NRAO VLA sky survey (NVSS) 1.4 GHz continuum survey. This survey covers 82% of the sky [45] with a source catalog containing over 1.8×10^6 sources that is 50% complete at 2.5 mJy. It is appropriate for our purpose since most of the bright sources are active radio galaxies and quasars whereas the less bright ones correspond to nearby star-forming galaxies. As a consequence, almost all the sources away from the galactic plane ($|b| > 2°$) are extragalactic.

We pixelize the NVSS catalog using HEALPIX [51] maps with $N_{side} = 256$ corresponding to around $14'$ square pixels [52]. As an extra precaution, we removed sources with a flux greater than 1 Jy as well as a 1 degree disk around them. We also mask out pixels at low galactic latitude ($|b| < 10°$) and those unobserved by the survey ($\delta < -36.87°$). We ended up with 1.29×10^6 sources with an average density $\bar{G} = 159\,000$ gal/sr.

III. CMB LENSING

Weak lensing by large-scale structure remaps the CMB temperature field on the sky; the lensed temperature $\tilde{T}(\hat{\mathbf{n}})$ and unlensed temperature $T(\hat{\mathbf{n}})$ are related by [53]

$$\tilde{T}(\hat{\mathbf{n}}) = T(\hat{\mathbf{n}} + \mathbf{d}(\hat{\mathbf{n}})), \qquad (1)$$

where $\mathbf{d}(\hat{\mathbf{n}})$ is a vector field representing the deflection angles. To first order in perturbation theory, $\mathbf{d}(\hat{\mathbf{n}})$ is expected to be a pure gradient:

$$d_a(\hat{\mathbf{n}}) = \nabla_a \phi(\hat{\mathbf{n}}), \qquad (2)$$

where the scalar potential ϕ is given by the line of sight integral:

$$\phi(\hat{\mathbf{n}}) = -2 \int_0^{\chi_*} d\chi \left(\frac{\chi_* - \chi}{\chi \chi_*} \right) \Psi(\chi \hat{\mathbf{n}}, \eta_0 - \chi), \quad (3)$$

where χ denotes conformal distance along the line of sight in the assumed flat cosmology, χ_* is the conformal distance to recombination, and η_0 is conformal time today. The integral in Eq. (3) receives contributions from a broad redshift range with median around $z \sim 2$.

How can CMB lensing be detected in data? At the power spectrum level, lensing slightly smoothes the acoustic peaks in the temperature power spectrum C_ℓ^{TT} and adds power in the damping tail [54]. However, these effects are too small to be detectable in existing data sets. Going beyond the power spectrum, the effect of CMB lensing

on higher-point statistics of the CMB is stronger and requires less instrumental sensitivity to detect [55].

The theory of CMB lens reconstruction [56–59] provides a framework for extracting this higher-point signal which we will use throughout this paper. One first defines a quadratic (in the CMB temperature T) estimator for the CMB lensing potential ϕ. The simplest higher-point estimator for detecting CMB lensing would be the power spectrum $C_\ell^{\phi\phi}$: a quadratic estimator in the reconstruction ϕ or a four-point estimator in T.

However, the three-year WMAP data do not have sufficient sensitivity to detect CMB lensing via the auto power spectrum $C_\ell^{\phi\phi}$. This can be seen by considering the statistical "noise" in the reconstruction; in [56] it is shown that the reconstruction noise power spectrum $N_\ell^{\phi\phi}$ is given by

$$\frac{1}{N_\ell^{\phi\phi}} = \frac{1}{2\ell+1} \sum_{\ell_1\ell_2} \frac{(C_{\ell_2}^{TT}F_{\ell_1\ell\ell_2} + C_{\ell_1}^{TT}F_{\ell_2\ell\ell_1})^2}{2(C_{\ell_1}^{TT} + N_{\ell_1}^{TT})(C_{\ell_2}^{TT} + N_{\ell_2}^{TT})}, \quad (4)$$

where $F_{\ell_1\ell_2\ell_3}$ is defined by

$$F_{\ell_1\ell_2\ell_2} = \mathcal{G}_{\ell_1\ell_2\ell_3}f_{\ell_1\ell_2\ell_3} \quad (5)$$

$$\mathcal{G}_{\ell_1\ell_2\ell_3} = \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}}\begin{pmatrix}\ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0\end{pmatrix} \quad (6)$$

$$f_{\ell_1\ell_2\ell_3} = \frac{\ell_2(\ell_2+1) + \ell_3(\ell_3+1) - \ell_1(\ell_1+1)}{2}. \quad (7)$$

In Fig. 2 (left panel) we have shown the noise power spectrum $N_\ell^{\phi\phi}$ for three-year WMAP sensitivity, with the fiducial signal power spectrum $C_\ell^{\phi\phi}$ shown for comparison. Although the CMB temperature anisotropies are signal dominated across a wide range of angular scales (Fig. 1), the lens reconstruction is highly noise dominated. At this level of signal-to-noise, an "internal" (to WMAP) detec-

tion of CMB lensing, by measuring the auto power spectrum $C_\ell^{\phi\phi}$, is not possible.

It is frequently the case that a signal which is too noisy for internal detection can nonetheless be detected via cross correlation to a second, less noisy signal. (For example, the first-year WMAP data had poor sensitivity to the EE polarization signal, but contained a many-sigma detection of CMB polarization via the TE cross correlation [60].) In this paper, we will detect the lensing signal in WMAP by cross correlating to radio galaxy counts in NVSS, thus detecting a nonzero cross power spectrum $C_\ell^{\phi g}$. The galaxy field g is much less noisy than ϕ (Fig. 1), but the two fields have a significant redshift range in common and so are highly correlated; the correlation in the fiducial model is ~ 0.65 on angular scales $\ell \lesssim 100$. Therefore, the effective signal-to-noise is higher for the cross correlation (Fig. 2, right panel). A forecast based on this signal-to-noise ratio, and the assumption of simple $f_{\rm sky}$ scaling, predicts that a $\sim 3.5\sigma$ detection can be made. If the same forecast is repeated using the parameters from [44] (i.e. first-year WMAP sensitivity and Sloan LRG's over 4000 deg^2), we find a $\sim 1\sigma$ result, in agreement with previous results.

In addition to the improved statistical errors from higher signal-to-noise, obtaining the detection as a cross correlation is more robust to systematics, as we will see in detail in Secs. V, VI, VII, and VIII. Any source of systematic contamination which appears in either WMAP or NVSS, but not both, will not bias our estimates for the cross power spectrum $C_\ell^{\phi g}$, since it does not correlate the two surveys. At worst, such a contaminant can affect the statistical significance of the detection, by increasing the error bars on each bandpower.

Our estimator for $C_\ell^{\phi g}$ will be defined by cross correlating the quadratic reconstruction of the lensing potential ϕ to the NVSS overdensity field g. Thus the estimator is three-point: two-point in the CMB temperature and one-point in the galaxy field. The same three-point estimator
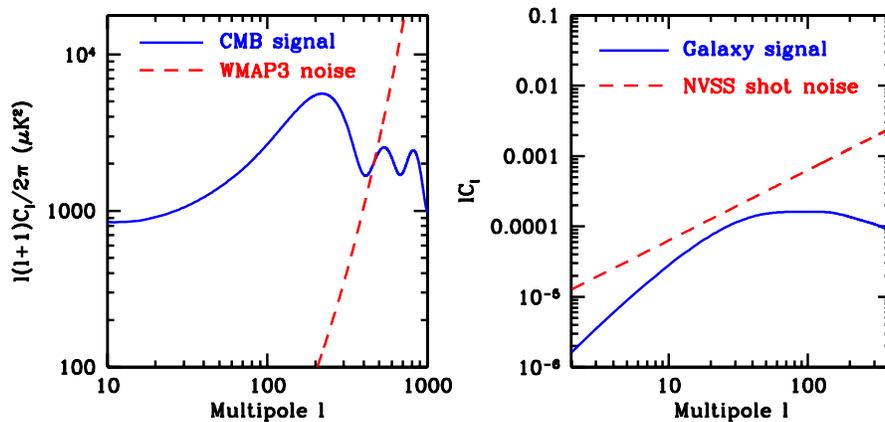
FIG. 1 (color online). Left panel: CMB signal power spectrum, and three-year WMAP noise power spectrum. Right panel: Fiducial NVSS signal power spectrum, and NVSS shot noise ($\bar{G} = 159\,000$ gal/sr).
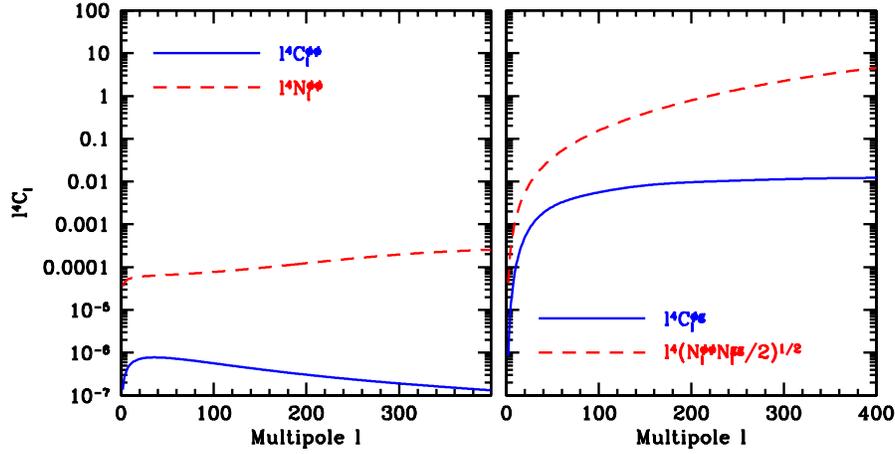
FIG. 2 (color online).   Left panel: Auto power spectrum $C_\ell^{\phi\phi}$ of the CMB lensing potential, and reconstruction noise power spectrum $N_\ell^{\phi\phi}$ [Eq. (4)] at three-year WMAP noise levels. Right panel: Cross power spectrum $C_\ell^{\phi g}$ between the CMB lensing potential and NVSS galaxy counts, and the effective noise power spectrum $[N_\ell^{\phi\phi} N_\ell^{gg}/2]^{1/2}$ for detecting the cross correlation. The "boost" in signal-to-noise between the two cases is sufficient to obtain a several-sigma detection of CMB lensing.

can also be derived from the general theory of bispectrum estimation [61–63].

The most general three-point correlation between two CMB multipoles and one galaxy multipole which is allowed by rotational and parity invariance is of the form

$$\langle a_{\ell_1 m_1}^T a_{\ell_2 m_2}^T a_{\ell_3 m_3}^g \rangle = b_{\ell_1 \ell_2 \ell_3} \mathcal{G}_{\ell_1 \ell_2 \ell_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \quad (8)$$

This equation defines the bispectrum $b_{\ell_1 \ell_2 \ell_3}$. (More properly, with the $\mathcal{G}_{\ell_1 \ell_2 \ell_3}$ prefactor included, we have defined the "reduced bispectrum" in Eq. (8); with this prefactor $b_{\ell_1 \ell_2 \ell_3}$ reduces to the flat sky bispectrum in the limit of large ℓ [64].) Whenever we write bispectra in this paper, ℓ_1, ℓ_2 are understood to denote CMB multipoles and ℓ_3 denotes a galaxy multipole.

From this perspective, the CMB lensing signal simply gives a contribution to the bispectrum which we want to measure. The lensing bispectrum is proportional to $C_\ell^{\phi g}$:

$$b_{\ell_1 \ell_2 \ell_3} = (f_{\ell_1 \ell_2 \ell_3} C_{\ell_2}^{TT} + f_{\ell_2 \ell_1 \ell_3} C_{\ell_1}^{TT}) C_{\ell_3}^{\phi g}. \quad (9)$$

One can think of this as a single bispectrum which is estimated to give an overall detection, or a linear combination of independent bispectra corresponding to bandpowers in $C_\ell^{\phi g}$.

In Appendix B, we show that the lens reconstruction and bispectrum formalisms are equivalent, so that it is a matter of convenience which to use. In this paper, we have generally used the lens reconstruction formalism, but will occasionally refer to the bispectrum formalism when it provides additional perspective.

One issue which is clearer from the bispectrum perspective is the distribution of statistical weight. Suppose we consider the total squared detection significance σ^2, rather than splitting the signal into bandpowers. Starting from the bispectrum in Eq. (9), one can write σ^2 as a sum over multipoles (ℓ_1, ℓ_2, ℓ_3). In Fig. 3, we have split up this sum
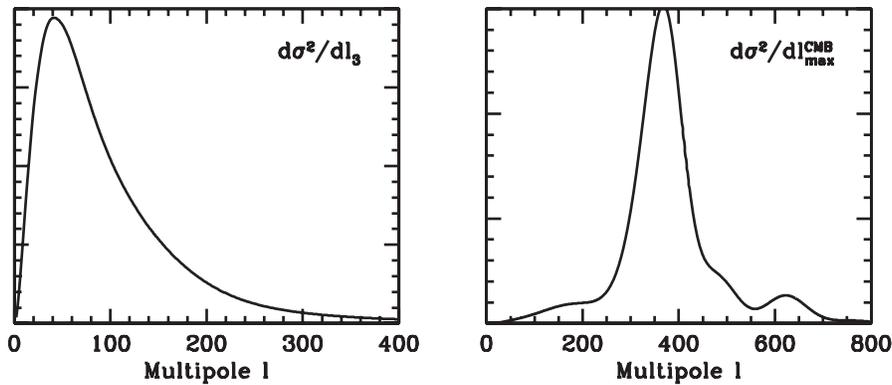
FIG. 3.   Mean contribution to the squared total detection significance σ^2 per NVSS galaxy multipole ℓ_3 (left panel), and per unit increase in maximum CMB multipole $\ell_{\max}^{CMB} = \max(\ell_1, \ell_2)$ (right panel). Most of the statistical weight comes from galaxy multipoles near $\ell \sim 50$, and CMB multipoles near $\ell \sim 400$.

to show the contribution per multipole. (Since there are two CMB multipoles, we show the contribution per unit increase in the maximum multipole $\ell_{\max}^{\mathrm{CMB}} = \max(\ell_1, \ell_2)$.) It is seen that the greatest statistical weight comes from galaxy multipoles near $\ell_3 \sim 50$, and CMB multipoles near $\ell \sim 400$ corresponding to an acoustic trough in the primary CMB. In bispectrum language, most of the signal is in "squeezed" triangles where the galaxy wave number is much smaller than the two CMB wave numbers. This corresponds to the intuitive statement that lens reconstruction estimates degree-scale lenses indirectly through their effect on smaller-scale hot and cold spots in the CMB.

IV. PIPELINE

In this section, we describe our simulation and analysis pipeline for estimating the cross power spectrum $C_\ell^{\phi g}$ from the WMAP and NVSS data sets, and present results with statistical errors. (Systematics will be treated in Secs. V, VI, VII, and VIII.)

A. Pipeline description

Our pipeline is shown in Fig. 4. Steps (1)–(4) represent the simulation direction and produce simulated WMAP
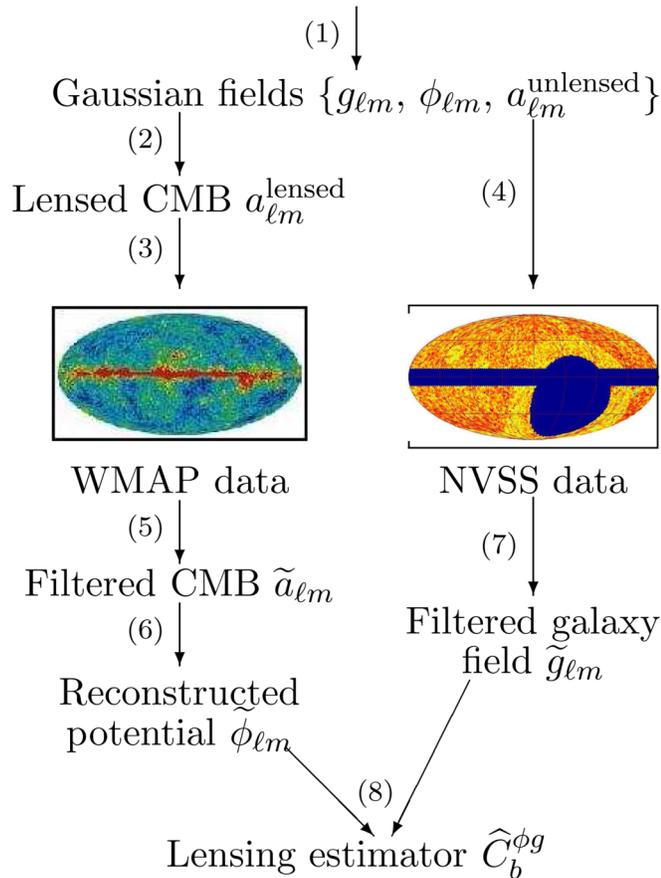
FIG. 4 (color online). Simulation + analysis pipeline used in this paper; the stages (1)–(7) are described in detail in Sec. IV.

and NVSS data sets with CMB lensing. Steps (5)–(8) are the analysis direction and produce power spectrum estimates $\hat{C}_b^{\phi g}$ in bands b, starting from the WMAP and NVSS data sets. We now describe each step in detail.

The first step (1) is simulating Gaussian fields: the unlensed CMB temperature, lensing potential, and (shot noise free) radio galaxy field g. We use the power spectra $C_\ell^{TT}, C_\ell^{\phi\phi}, C_\ell^{gg}, C_\ell^{\phi g}$ in the fiducial model. The last two are computed using the Limber approximation (e.g. [65]) and a simple constant galaxy bias model: we take the galaxy overdensity to be given by the line of sight integral

$$\delta g(\hat{\mathbf{n}}) = b_g \frac{\int d\chi \frac{dN}{d\chi} \delta(\chi\hat{\mathbf{n}}, \eta_0 - \chi)}{\int d\chi \frac{dN}{d\chi}} \qquad (10)$$

using a fiducial redshift distribution $dN/d\chi$ and galaxy bias b_g that will be discussed in the next section.

In step (2), we compute the lensed CMB from the lensing potential and unlensed CMB. The lensing operation

$$\tilde{T}(\hat{\mathbf{n}}) = T(\hat{\mathbf{n}} + \mathbf{d}(\hat{\mathbf{n}})) \qquad (11)$$

is performed directly in position space [rather than relying on an approximation to Eq. (11) such as the gradient approximation]. The right-hand side of Eq. (11) is evaluated using cubic interpolation on a high resolution ($\approx$ 0.5 arc min) map.

In step (3), we simulate the eight Q-, V-, and W-band channels of WMAP. The maps are simulated at HEALPIX resolution $N_{\mathrm{side}} = 1024$ and downgraded to $N_{\mathrm{side}} = 512$ to minimize pixelization artifacts. To simulate each map, we first convolve with the beam and pixel window in harmonic space:

$$a_{\ell m} \rightarrow B_\ell W_\ell a_{\ell m}, \qquad (12)$$

where B_ℓ is the beam transfer function (distinct for each channel) and W_ℓ is the pixel window function. We then take the spherical transform and add Gaussian noise to each pixel. The noise rms is pixel dependent but the noise is assumed uncorrelated between pixels.

As the last step in the simulation direction, in step (4) we simulate NVSS, including clustering which is controlled by the Gaussian field g, by generating a Poisson galaxy count in each pixel p whose mean is given by

$$\lambda(p) = \bar{n}(1 + g(p)), \qquad (13)$$

where $\bar{n}$ is the mean number of galaxies per pixel over the survey. We simulate NVSS at $N_{\mathrm{side}} = 1024$ and downgrade to $N_{\mathrm{side}} = 256$.

Step (5) is the first step in the analysis direction: we start with the pixel-space maps corresponding to the eight Q-, V-, and W-band WMAP channels, and compute a single harmonic-space map $\tilde{a}_{\ell m}$ representing the inverse signal + noise filtered temperature $\tilde{a} = (S + N)^{-1}a$. This reduction step is a common ingredient in many types of optimal

estimators [62,63,66–69]. The general principle is that the filtering operation completely incorporates the sky cut and noise model, so that optimal estimators can be constructed by simple subsequent operations directly in harmonic space. For example, the optimal TT power spectrum estimator is obtained by straightforwardly computing the power spectrum $C_\ell^{\tilde{a}\tilde{a}}$.

Here and throughout the body of the paper, we will defer technical details of the estimators to Appendices A and B and concentrate on conveying intuition. In this case, the idea is that the $(S + N)^{-1}$ filter simply weights each mode of the data by the inverse of its total variance, so that poorly measured modes are filtered out. For example, the sky cut is incorporated into the noise covariance N by assigning infinite noise variance to pixels which are masked (in implementation, we use N^{-1} rather than N and set the relevant matrix entries to zero). Data outside the sky cut is then completely filtered out: the map $\tilde{a}$ is independent of the map values in masked pixels, and everything "downstream" in the analysis pipeline will be blind to the masked data. As a similar example, we marginalize the CMB monopole and dipole modes by assigning them infinite variance. Finally, the beam transfer functions [Eq. (12)] are kept distinct in the filtering operation, so that optimal frequency weighting is performed: the filtered map $\tilde{a}_{\ell m}$ will receive contributions from all frequencies at low ℓ, but will depend mainly on the highest-frequency channels (i.e., the channels with narrow beams) at high ℓ. The filtered map $\tilde{a} = (S + N)^{-1}a$ can also be thought of as the least-squares estimate of the signal, given data from all channels.

In step (6), we perform lens reconstruction. Given the filtered CMB temperature $\tilde{a}_{\ell m}$ from step (5), we compute the reconstructed potential $\tilde{\phi}_{\ell m}$, defined by the equation:

$$\sum_{\ell m} \tilde{\phi}_{\ell m} Y_{\ell m}(x) = \nabla^a(\alpha(x)\nabla_a\beta(x)), \qquad (14)$$

where α and β are defined by

$$\alpha(x) = \sum_{\ell m} \tilde{a}_{\ell m} Y_{\ell m}(x) \qquad (15)$$

$$\beta(x) = \sum_{\ell m} C_\ell^{TT} \tilde{a}_{\ell m} Y_{\ell m}(x). \qquad (16)$$

As explained in [56], $\tilde{\phi}_{\ell m}$ is a noisy reconstruction of the CMB lensing potential (or more precisely, the inverse noise weighted potential $N_\phi^{-1}\phi$, where N_ϕ is the noise covariance of the reconstruction) which is quadratic in the CMB temperature. Note that both $\tilde{a}$ and $\tilde{\phi}$ are defined in harmonic space, but Eq. (14) involves multiplication and derivative operations in real space; in Appendix B, we explain in detail how $\tilde{\phi}_{\ell m}$ is computed.

In step (7), we perform inverse signal + noise filtering on the NVSS data: given pixel-space galaxy counts, we compute the harmonic-space map $\tilde{g}_{\ell m} = (S + N)^{-1}g$ where the noise covariance N represents shot noise. This

is analogous to the WMAP filtering operation in step (5), but there is one new ingredient. In addition to marginalizing data outside the sky cut, and the monopole and dipole, we marginalize any mode which is independent of the angular coordinate φ in equatorial coordinates. (In harmonic space, this is equivalent to marginalizing modes with $m = 0$.) This is needed to remove a systematic effect in NVSS which we will discuss in detail in Sec. V; for now we remark in advance that all results in this paper include this marginalization.

Finally, in step (8), we compute the bandpower estimator $\hat{C}_b^{\phi g}$ by cross correlating the fields $\tilde{\phi}_{\ell m}$ and $\tilde{g}_{\ell m}$ from steps (6) and (7). There is one wrinkle here: as we show in Appendix B, to obtain the optimal estimator, we must include an extra term which subtracts the Monte Carlo average $\langle\tilde{\phi}\rangle$ taken over unlensed simulations of WMAP:

$$\hat{C}_b^{\phi g} \stackrel{\text{def}}{=} \frac{1}{\mathcal{N}_b} \sum_{\substack{\ell \in b \\ -\ell \le m \le \ell}} \frac{1}{\ell^2}(\tilde{\phi}_{\ell m} - \langle\tilde{\phi}_{\ell m}\rangle)^*(\tilde{g}_{\ell m}), \qquad (17)$$

where $\mathcal{N}_b$ is a normalization constant to be discussed shortly. (We have included the factor $1/\ell^2$ since we estimate bandpowers assuming that $\ell^2 C_\ell^{\phi g}$ is flat in each band.) Note that the Monte Carlo average $\langle\tilde{\phi}_{\ell m}\rangle$ vanishes for symmetry reasons in the case of full-sky coverage and isotropic noise, but sky cuts or noise inhomogeneities will give rise to a nonzero average. The extra term in Eq. (17) simply improves the variance of the estimator by subtracting the spurious cross correlation between this average and the galaxy field $\tilde{g}$.

We determine the estimator normalization $\mathcal{N}_b$ by end-to-end Monte Carlo simulations of the pipeline, including a nonzero $C_\ell^{\phi g}$ in the simulations for calibration. (Strictly speaking, the normalization should be a matrix which couples bands $b \neq b'$, but we have neglected the off-diagonal terms, which are small for our case of large sky coverage and wide bands.) As we will see in Appendix B, the normalization $\mathcal{N}_b$ is proportional to a cut-sky Fisher matrix element, which must be computed by Monte Carlo (unless an approximation is made such as simple f_{sky} scaling). In addition, Monte Carlo simulations are also needed to compute the one-point term in Eq. (17).

This concludes our description of the pipeline. We have not motivated the details in the construction of our lensing estimator $\hat{C}_b^{\phi g}$, but in Appendix B we show that the estimator is optimal, by proving that it achieves statistical lower limits on the estimator variance, so that the best possible power spectrum uncertainties are obtained. This justifies the combination of ingredients presented here: inverse signal + noise filtering (steps 5 and 7), keeping the lensing potential in harmonic space (step 6), and including the one-point term in the cross correlation (step 8); and shows that no further improvements are possible.

B. Results

The result of applying this analysis pipeline to the WMAP and NVSS data sets is shown in Fig. 5. We emphasize that the uncertainties are purely statistical. Systematic errors will be studied in Secs. V, VI, VII, and VIII, and an updated version of the result shown in Sec. IX, where we also show that the detection significance with systematic errors included is 3.4σ.

Our error bars were obtained by Monte Carlo, cross-correlating simulations of WMAP and NVSS. As a consistency check, Fig. 5 shows that nearly identical error bars are obtained if WMAP simulations are cross correlated to the real NVSS data, or vice versa. This is an important check; if it failed, then we would know that our simulations were failing to capture a feature of the data sets which contributes significant uncertainty to the lensing estimator. In addition, it shows that the uncertainties only depend on correctness of one of the simulation pipelines. Suppose, for example, that the NVSS data set contains unknown catastrophic systematics which invalidate our simulations. Because the same result is obtained by treating NVSS as a black box to be cross correlated to WMAP simulations, it is still valid (provided that WMAP contains no ''catastrophic'' systematics)

As another consistency check, in Fig. 6 we show the detection that is obtained if each frequency in WMAP is analyzed separately. No signs of inconsistency are seen, although we have not attempted to quantify this precisely: the results obtained from different frequencies are corre-

FIG. 6 (color online). CMB lensing detection obtained by analyzing Q-band (left/black error bar in each triple), V-band (middle/blue), and W-band (right/red) data from WMAP separately, showing consistency of the result between CMB frequencies.

lated even though the CMB noise realizations are independent, because NVSS is identical and so is the underlying CMB realization. For the same reason, we caution the reader that the three sets of error bars in Fig. 6 cannot be combined in a straightforward way to obtain an overall result. The best possible way of combining the data is already shown in Fig. 5: the maps from the three frequencies are combined into a single CMB map which is cross correlated to NVSS.

C. Curl null test

Our lensing estimator $\hat{C}_b^{\phi g}$ detects a gradient component in the deflection field d_a via cross correlation to radio galaxy counts. If we instead decompose the deflection field into gradient and curl,

$$d_a(\hat{\mathbf{n}}) = \nabla_a \phi(\hat{\mathbf{n}}) + \epsilon_{ab} \nabla^b \psi(\hat{\mathbf{n}}), \qquad (18)$$

then one can similarly devise an estimator $\hat{C}_b^{\psi g}$ to detect the curl component. Since the curl component is expected to be absent cosmologically, this is a null test [70]. Note that we have parametrized the curl component by a pseudoscalar potential ψ, for notational uniformity with the gradient component which is parametrized by its scalar potential ϕ.

In Appendix B, we show that the optimal estimator is constructed as follows. First, we define a reconstructed potential $\tilde{\psi}$ which is quadratic in the CMB temperature:

$$\sum_{\ell m} \tilde{\psi}_{\ell m} Y_{\ell m}(x) = \epsilon^{ab} \nabla_a(\alpha(x)\nabla_b \beta(x)) \qquad (19)$$

with α, β as in Eqs. (15) and (16). Second, we define a power spectrum estimator by cross correlating to galaxy

FIG. 5 (color online). Detection of CMB lensing via the cross power spectrum $C_\ell^{\phi g}$ between the reconstructed potential and galaxy counts. The three 1σ error bars on each bandpower represent different Monte Carlo methods: WMAP simulations vs NVSS simulations (left/black); WMAP data vs NVSS simulations (middle/blue); and WMAP simulations vs NVSS data (right/red). These error bars represent statistical errors only; the result with systematic errors included will be shown in Fig. 19.

counts, subtracting the one-point term:

$$\hat{C}_{b}^{\psi g} \stackrel{\text{def}}{=} \frac{1}{\mathcal{N}_b} \sum_{\substack{\ell \in b \\ -\ell \le m \le \ell}} \frac{1}{\ell^2} (\tilde{\psi}_{\ell m} - \langle \tilde{\psi}_{\ell m} \rangle)^* (\tilde{g}_{\ell m}). \qquad (20)$$

This construction is identical to our construction [Eqs. (14) and (17)] of the lensing estimator $\hat{C}_b^{\phi g}$, except that a 90° rotation has been included (via the antisymmetric tensor ϵ_{ab}) in Eq. (19).

The result of the curl null test is shown in Fig. 7. The χ^2 for the null test is 12.1 with 8 degrees of freedom (i.e. high at 1σ), so the null test passes.

How strong is the null test obtained by demanding that $\hat{C}_{\psi g}$ be consistent with zero? One might hope that astrophysical contaminants, such as point sources or the Sunyaev-Zel'dovich effect, would contribute both gradient and curl components to the reconstructed deflections, and thus be monitored by the null test. However, parity invariance requires $C_\ell^{\psi g} = 0$ even when $\psi \neq 0$. Since astrophysical contaminants are expected to obey parity-invariant statistics, they will not bias $C_\ell^{\psi g}$ on average. Our null test therefore only monitors contaminants which can violate parity invariance, such as galactic foregrounds or instrumental systematics. This is analogous to the $C_\ell^{EB} = 0$ null test in CMB polarization experiments: it is not sensitive to all sources of contamination, but is nevertheless an important sanity check.

We remark that for a detection of CMB lensing which is internal to the CMB (detecting lensing via the auto power spectrum $C_\ell^{\phi\phi}$, rather than the cross spectrum $C_\ell^{\phi g}$ consid-

ered here), one would have one null test ($C_\ell^{\psi\psi} = 0$) which can monitor parity-invariant contaminants, and one null test ($C_\ell^{\phi\psi} = 0$) which cannot.

V. NVSS SYSTEMATICS

In the previous section, we obtained a statistical detection of CMB lensing (Fig. 5) by cross correlating WMAP and NVSS, and showed that two consistency checks were satisfied: frequency independence (Fig. 6) and a curl null test (Fig. 7). The rest of the paper is devoted to studying potential instrumental and astrophysical contaminants of the lensing signal, to show that the observed lensing cross correlation is not due to systematic contamination. In this section, we will consider NVSS systematics.

If a maximum likelihood galaxy power spectrum is calculated from NVSS using the sky cut described in Sec. II, the power spectrum C_ℓ^{gg} shown in the top panel of Fig. 8 is obtained. The very high bandpower in the lowest ℓ band is a clear sign of systematic contamination. If the low ℓ modes are isolated by low-pass filtering the
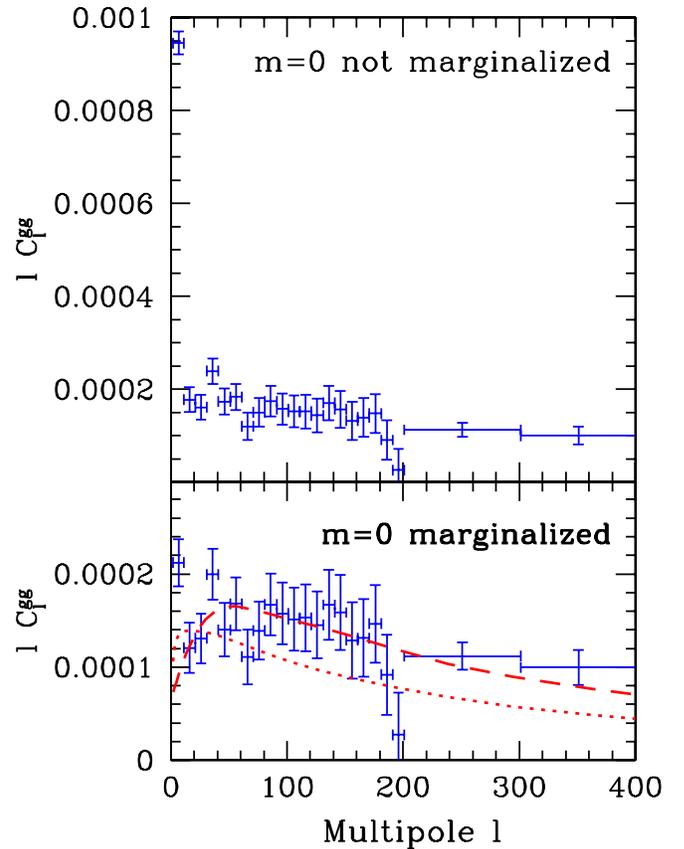
FIG. 8 (color online). Maximum likelihood NVSS galaxy power spectrum, calculated without (top panel) and with (bottom panel) marginalization of $m = 0$ modes in equatorial coordinates. In the bottom panel, fiducial spectra are shown (both for $b_g = 1.7$) from the model for dN/dz by [73] (dotted line) and our fit in Eq. (21) (dashed line).

FIG. 7 (color online). Result of the curl null test ($C_\ell^{\psi g} = 0$). As in Fig. 5, the three error bars on each bandpower represent different Monte Carlo methods: WMAP simulations vs NVSS simulations (left/black); WMAP data vs NVSS simulations (middle/blue); and WMAP simulations vs NVSS data (right/red).

NVSS galaxy counts to $\ell \leq 10$, the resulting map shows azimuthal "striping" when plotted in equatorial coordinates (Fig. 9). This is a known systematic effect in NVSS [71]: due to calibration problems at low flux densities, the galaxy density has a systematic dependence on declination, which can mimic long-wavelength modes in the galaxy field.

To remove this contaminant, we analyze NVSS in equatorial coordinates, and marginalize any modes in the data which are constant in the azimuthal coordinate φ. The marginalization is performed by modifying the NVSS noise model so that all such modes are assigned infinite variance, as described in Appendix A. Thus any signal which is constant in φ is completely filtered out in the inverse signal + noise weighted map $\tilde{g}$ which appears in our estimators. Note that treating the marginalization as part of the noise model means that the loss in sensitivity due to marginalizing $m = 0$ modes is already included in the statistical errors; it is not necessary to assign systematic errors separately.

After including this marginalization in the analysis, the NVSS galaxy power spectrum shown in the bottom panel of Fig. 8 is obtained, showing reasonable agreement with our fiducial C_ℓ^{gg}. Marginalizing $m = 0$ modes produces a large shift in the lowest bandpower and a much smaller shift in higher bands. In Fig. 10 (top panel), we show the shift in each bandpower when $m = 0$ modes are marginalized; the error bars show the rms shift obtained when the same marginalization is performed in simulations. It is seen that the shift is statistically significant not only in the lowest ℓ band, but all the way to $\ell \sim 100$. We conclude that declination gradients in NVSS are an important systematic on a range of scales and should always be marginalized in cosmological studies.

Has marginalizing $m = 0$ completely removed the systematic? To answer this, we tried marginalizing the $m = 1$ Fourier mode in the azimuthal coordinate φ, in addition to the $m = 0$ mode. In this case, we find (Fig. 10, bottom panel) that the shift in C_ℓ^{gg} bandpowers is consistent with simulations. (There is a possible glitch at $\ell \sim 200$, but this
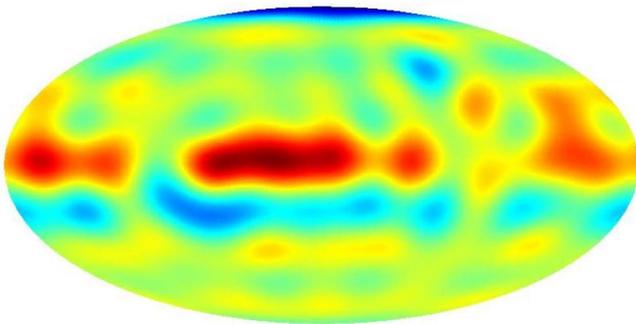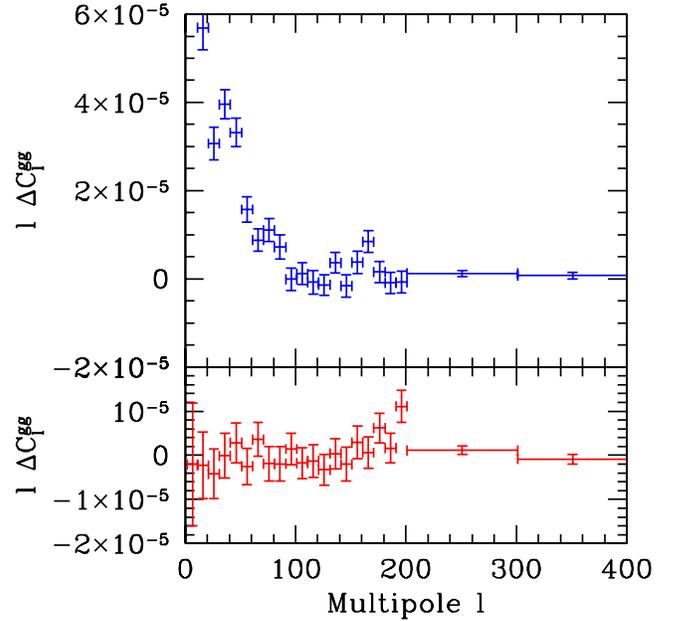
FIG. 10 (color online). Change ΔC_ℓ^{gg} in maximum likelihood galaxy power spectrum, when NVSS is analyzed with $m = 0$ marginalization vs no marginalization (top panel) or $m = 0, 1$ marginalization vs $m = 0$ marginalization (bottom panel) in equatorial coordinates. The error bars represent the rms shift obtained when Monte Carlo simulations are analyzed in the same way.

is outside the range of angular scales which contribute to the lensing detection.) Therefore, we believe that marginalizing all modes with $m = 0$ in equatorial coordinates completely removes the systematic; there is no evidence that the contamination extends to higher m.

In addition to declination gradients, there is another NVSS systematic which has been relevant for cosmological studies: multicomponent sources [71,72]. Radio galaxies whose angular size is sufficiently large to be resolved by the 45-arcsec NVSS beam will appear as multiple objects in the NVSS catalog. This can contribute extra power to the auto spectrum C_ℓ^{gg}, at a level which is a few percent of the shot noise. At worst, this could increase the variance of our cross-correlation estimator $\hat{C}_b^{\phi g}$ by a few percent without biasing the estimator. Furthermore, as can be seen in Fig. 8 (bottom panel), we see no evidence for galaxy power in excess of fiducial in the highest ℓ band, which is most sensitive to this systematic. We conclude that multicomponent sources are a negligible source of systematic error for CMB lensing.

Next we consider uncertainties in the NVSS redshift distribution dN/dz and galaxy bias b_g. These uncertainties affect our fiducial power spectra $C_\ell^{\phi g}$, C_ℓ^{gg} in a given cosmology, and would need to be understood in detail if we wanted to constrain cosmological parameters from our lensing detection. However, since we are merely measuring the cross spectrum $C_\ell^{\phi g}$, there is only one effect to con-

FIG. 9 (color online). NVSS galaxy overdensity field in equatorial coordinates, low-pass filtered to multipoles $\ell \leq 10$, showing visible azimuthal striping.

sider: the Monte Carlo error bars we assign depend on the fiducial galaxy spectrum C_ℓ^{gg} used in the simulations. (We verified in simulations that the fiducial *cross* spectrum $C_\ell^{\phi g}$ does not significantly affect the error bars.) If we use a fiducial C_ℓ^{gg} with too little power, we will underestimate our errors. Therefore, it is important to check that our fiducial C_ℓ^{gg} agrees with the galaxy power spectrum obtained from the data.

Estimates for the radio luminosity function inspired by optical and infrared observations were given in [73]. Using their mean-z, model 1 for average sources, the authors of [74] were able to reproduce the NVSS autocorrelation function rather well. However, the dotted curve in Fig. 8 shows the galaxy power spectrum C_ℓ^{gg}, calculated using a mean bias of $b_g = 1.7$ (in agreement with the values in [72,74]) and the same model for dN/dz. For our fiducial value of σ_8, the model power spectrum is deficient relative to the observed power spectrum.

Therefore, we search for a NVSS redshift distribution that better reproduces our angular power spectrum measurement. We find that for $b_g = 1.7$, a near Gaussian which is lopsided toward low redshift and centered at $z_0 = 1.1$:

$$\frac{dN}{dz} \propto \begin{cases} \exp(-\frac{(z-z_0)^2}{2(0.8)^2}) & (z < z_0) \\ \exp(-\frac{(z-z_0)^2}{2(0.3)^2}) & (z > z_0) \end{cases} \quad (21)$$

results in a good fit. This match to the NVSS angular power spectrum is shown in the dashed curve of Fig. 8. We have used this fiducial C_ℓ^{gg} in all simulations in this paper.

We make no claim that our fiducial (dN/dz) is a more accurate model for the real NVSS redshift distribution than the previously considered model. It is just a device for generating simulations with the same power spectrum as the data, so that we do not underestimate our error bars. As a check, in Sec. IV we compared Monte Carlo based error estimates for WMAP data versus NVSS data on one hand, and WMAP data versus NVSS simulations on the other, and obtained agreement (Fig. 5). Using the dotted line in Fig. 8 would underestimate the power spectrum errors by $\sim 20\%$ due to the disagreement with the power spectrum seen in the data. We have not investigated the reason for the disagreement in detail since it is somewhat peripheral to the primary purpose of this paper. However, the redshift distribution and galaxy bias assumed in the modeling would be critical if we were to infer constraints on cosmological parameters (such as the normalization of matter fluctuations σ_8 or the total matter density Ω_0) from our measurements of the NVSS angular power spectrum and the cross correlation $C_\ell^{\phi g}$. We return to this issue in Sec. IX.

VI. WMAP SYSTEMATICS

Because our lensing estimator receives contributions from CMB anisotropies on small angular scales (Fig. 3), the WMAP systematics most likely to affect the detection are point sources and beam effects. In our pipeline, beam effects are incorporated by convolving the CMB with an isotropic beam [Eq. (12)] which is different for each DA. This is approximate in two ways: first, the real WMAP beams are not perfectly isotropic, but contain asymmetries which also convolve small-scale modes of the CMB by a sky varying kernel defined by the details of the scanning strategy. Second, the isotropic part of each beam is not known perfectly; uncertainty in the beam transfer function acts as a source of systematic error in our lensing detection. We study these two effects in Secs. VI A and VI B.

In Sec. VI C, we consider galactic microwave foregrounds and show that their effect on the lensing detection is small. Point sources and thermal Sunyaev-Zel'dovich (SZ) will be treated separately in Secs. VII and VIII. The ISW effect [24] does not affect our lensing estimator, since the signal is negligible on CMB angular scales ($\ell \sim 400$) which contribute. The Rees-Sciama effect [75] would give a small contribution on these scales, but we will ignore it since it is negligible compared to the SZ signal.

A. Beam asymmetry

The WMAP beams are asymmetric due to: (1) the feeds not being at the primary focus, and (2) substructure caused by 0.02 cm rms deformations in the primary mirror [50]. The Q-band beams are elliptical with minor/major axis ratio of ≈ 0.8. The V- and W-band beams show significant substructure at the -10 to -20 dB level, leading to $\approx 0.7\%$ distortions in the inferred power spectrum [46].

Although deviations from azimuthal symmetry of the beams have a small effect when estimating the WMAP temperature power spectrum, it is unclear whether the same is true when estimating lensing. At an intuitive level, CMB lens reconstruction recovers degree-scale modes of the lensing potential indirectly, through their distorting effect on smaller-scale hot and cold spots in the CMB. Beam asymmetries which convolve the small-scale CMB modes have a qualitatively similar effect and may be degenerate with lensing. For example, a beam quadrupole imparts an overall ellipticity or shear to the hot and cold spots.

To incorporate beam asymmetry into our pipeline, we expand the beam profile in spherical harmonics $Y_{\ell s}$. The $s = 0$ multipoles of the beam represent the azimuthally averaged beam and are already incorporated in both the analysis and simulation directions of our pipeline. The higher-s multipoles have been estimated by the WMAP team and represent corrections to the azimuthally symmetric approximation. In Appendix D, we show how to incorporate the higher multipoles into the simulation direction of the pipeline, generalizing the convolution in Eq. (12). In contrast to the $s = 0$ multipoles, convolving with the higher multipoles depends on the scan strategy; our method incorporates the details of the WMAP scan based on full timestream pointing. In Fig. 11, we illustrate our simula-
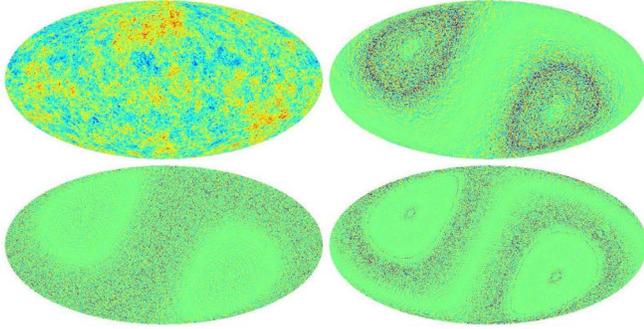
FIG. 11 (color online). Result of convolving a single noiseless CMB realization with the WMAP V1 beam, including beam asymmetry. We have shown the output map separated into contributions from different beam multipoles: $s = 0$ (isotropic component, top left); $s = 1$ (top right); $s = 2$ (bottom left); and $s = 3$ (bottom right). Each map has been scaled independently for visibility; the rms temperature in the $s = 0, \ldots, 3$ maps is 88, 0.4, 1.0, 0.04 μK. The convolution with the $s > 0$ multipoles is scan dependent and shows alignments with the ecliptic poles reflecting the WMAP scan strategy.

tion procedure for a single noiseless realization in V-band, showing the contribution of the $s = 0, \ldots 3$ multipoles to the beam-convolved map.

It would be very difficult to incorporate asymmetric beams into the analysis direction of the pipeline, so our approach is to treat beam asymmetry as a source of systematic error. We assign each lensing bandpower $C_b^{\phi g}$ a systematic error given by the Monte Carlo rms *change* in the bandpower when the same WMAP + NVSS simulation is analyzed with and without including beam asymmetry in the simulation pipeline. We find that the systematic error in each band is small compared to the statistical error. The result is shown, as part of a larger systematic error budget, in the "Beam asymmetry" column of Table I in Sec. IX.

B. Beam uncertainty

We have shown that systematic errors from beam asymmetry are small, so that the beam may be treated as the simple convolution in Eq. (12) to a good approximation. This leaves only one remaining beam-related source of systematic error: measurement uncertainty in the beam transfer function B_ℓ.

We model the beam transfer function uncertainty following [46], Sec. A2. The beam covariance matrix is dominated by a small number of modes. We diagonalize the matrix for each DA and keep only the 10 most significant modes. Then we construct realizations of the beam transfer function using

$$B_\ell = B_\ell^{(0)}\Big(1 + \sum_i u_i m_\ell^i\Big), \qquad (22)$$

where $B^{(0)}$ is the standard beam transfer function, u_i are

unit-variance normal random deviates, and m_ℓ^i are the beam covariance modes.

Armed with this simulation procedure, we assign systematic errors by computing the rms change in each band-power when the same simulation is analyzed with and without simulated beam uncertainty. We find that the systematic errors are extremely small.

C. Galactic foregrounds

In addition to the CMB, the sky at microwave frequencies contains other foreground signals which must be considered as a source of systematic error in lensing. We will find that the most important of these are point sources and the thermal Sunyaev-Zel'dovich effect, which will be discussed in Secs. VII and VIII, respectively. The other relevant microwave foregrounds are galactic in origin: dust, free-free emission, and synchrotron radiation. For descriptions of the foreground components, we refer the reader to [76].

Following [46], we will model dust contamination by adding a template derived from "Model 8" from Finkbeiner *et al.* [77], evaluated at 94 GHz and scaling to frequency ν by

$$T_A(\nu) = \Big(\frac{\nu}{94 \text{ GHz}}\Big)^{2.0} T_A(94 \text{ GHz}), \qquad (23)$$

where T_A denotes antenna temperature. The dust template is shown in Fig. 12, left panel.

When we cross correlate simulations of WMAP and NVSS, we find that including the dust template in the WMAP simulation results in a very small change in the estimated lensing signal. We take the Monte Carlo rms average of the *change* in each bandpower when the same pair of simulations is analyzed with and without the template as a systematic error estimate, shown in the "Dust" column of Table I in Sec. IX.

One might worry that this way of assigning systematic errors, based entirely on simulations, is too optimistic because it fails to account for unknown correlations between the templates and the real data sets. As a check, we obtain consistent results if we cross correlate an ensemble of WMAP simulations against the real NVSS data, or the
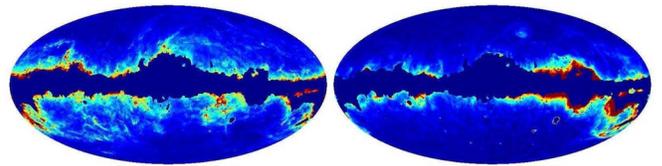
FIG. 12 (color online). Foreground templates used in this paper, shown with Kp0 mask (Sec. II) applied. Left panel: Dust template, based on [77] with frequency dependence given by Eq. (23). Right panel: Free-free template, based on [76,78] with frequency dependence given by Eq. (24). The masked rms of the templates in V-band is 6.4 μK and 4.8 μK, respectively.

real WMAP data (with and without template *subtraction*) against an ensemble of NVSS simulations. Finally, when the real WMAP and NVSS data sets are cross correlated with and without template subtraction, the change in each bandpower is consistent with our systematic error estimates, and no evidence for an overall bias is seen.

We treat free-free emission similarly; in this case we use the full-sky Hα map from [78], with the correction for dust extinction from [76], and frequency dependence:

$$T_A(\nu) = b_2 \left(\frac{\nu}{22.8 \text{ GHz}}\right)^{-2.14} I_{H\alpha}, \qquad (24)$$

where $b_2 = 6.7$ μK/Rayleigh and $I_{H\alpha}$ denotes the $H\alpha$ intensity. Again we find consistent systematic errors in the simulation-simulation, simulation-data, and data-data cases described in the previous paragraph. The results are shown in the "Free-free" column of Table I in Sec. IX; the systematic errors from free-free are slightly higher than dust, but still small.

Finally, we turn to galactic synchrotron emission. The WMAP team has derived synchrotron templates both from the Haslam 408 MHz survey [76,79], and internally by differencing the K and Ka band WMAP channels [46]. However, both of these templates are intended for use at degree scales, and do not have sufficient resolution to measure the synchrotron signal on the angular scales ($\ell \sim$ 400) which contribute to our lensing estimator. Therefore, it would not be meaningful to assign systematic errors from synchrotron emission by using either of these templates.

In the absence of a template for synchrotron, the best we can do is to make the assumption that the synchrotron contamination at $\ell \sim 400$ is comparable to the other galactic foregrounds. In V-band, synchrotron, free-free, and dust emission all contaminate the CMB at roughly similar levels [76]. In addition, synchrotron and dust appear to have similar spatial distributions [46], Fig. 5, so the dust template should give us a reasonable estimate of possible synchrotron contamination. However, a direct test of this assumption will have to await future higher-resolution measurements of synchrotron emission.

These results and the consistency of our measurement between frequencies (Fig. 6) lead us to conclude that our lensing detection is not contaminated by significant residual foregrounds. However, we quantify it by assigning each lensing bandpower a total systematic error from foregrounds by adding the systematic errors from the dust and free-free templates (treating the two as correlated) and then doubling the rms error to account for a synchrotron contribution with the same order of magnitude. The result is shown in the "Total galactic" column of Table I in Sec. IX.

VII. POINT SOURCE CONTAMINATION

Point sources which are bright enough to be resolved by WMAP are excluded by the Kp0 mask (Sec. II), but unresolved point sources act as a contaminating signal in the CMB. If the unresolved CMB point source signal were uncorrelated to NVSS, we would not expect point sources to affect our lensing estimator $\hat{C}_b^{\phi g}$ significantly. However, NVSS radio galaxies will contribute some nonzero flux at microwave frequencies and so appear directly as part of the point source contribution to the CMB. In addition, CMB point sources which do not actually appear as objects in NVSS may be correlated to NVSS objects in some way, e.g. if both are tracers of the same large-scale potential. Therefore, point sources are a possible contaminant of our lensing detection.

In this section, we will place limits on the level of point source contamination and assign systematic errors. Point sources will turn out to be our dominant source of systematic error, and so we will devote considerable effort to constructing robust error estimates.

A. Point source estimator

It is difficult if not impossible to construct a realistic model which would allow the level of point source contamination to be reliably estimated from general principles. At radio and microwave frequencies, several populations of point sources have been identified [80–83] with significant uncertainties in spectral index and clustering properties.

Therefore, our approach will be to estimate the level of point source contamination directly from the data. In this subsection, we will motivate and construct an estimator which is optimized for detecting point sources instead of CMB lensing, to use as a monitor for point source contamination. The first candidate for the point source estimator is simply the cross power spectrum C_ℓ^{Tg}.

However, consider the following toy model for point sources: suppose that there are N distinct populations of unclustered Poisson point sources which appear as objects in the NVSS catalog, and the ith population has number density n_i and constant flux per source S_i at CMB frequencies. In this model, the cross power spectrum is

$$C_\ell^{Tg} \propto \sum_{i=1}^{N} S_i n_i \qquad (25)$$

whereas the bias to the lensing estimator is proportional to

$$\Delta \hat{C}_\ell^{\phi g} \propto \sum_{i=1}^{N} S_i^2 n_i. \qquad (26)$$

Because the right-hand sides of Eqs. (25) and (26) are not related in any model-independent way, one cannot translate a value of the cross spectrum C_ℓ^{Tg} to an estimate of the point source contamination in the lensing estimator, without making implicit assumptions about the point source model.

For this reason, we next consider a different candidate for the point source estimator: the three-point estimator

optimized to detect the "Poisson" bispectrum

$$b_{\ell_1\ell_2\ell_3} = \text{constant}, \tag{27}$$

where, following Eq. (8), ℓ_1, ℓ_2 denote CMB multipoles and ℓ_3 denotes a galaxy multipole. (We will construct the estimator shortly; for now we "define" the point source estimator by writing down the bispectrum which we want to detect.)

To motivate this form, we note that the bispectrum in our toy model is

$$b_{\ell_1\ell_2\ell_3} \propto \sum_{i=1}^{N} S_i^2 n_i. \tag{28}$$

Comparing to Eq. (26), we see that each point source population makes contributions to the Poisson bispectrum and lensing estimator $\hat{C}_b^{\phi g}$ which are proportional. Therefore, an estimate of the Poisson bispectrum will directly translate to a systematic error estimate for the lensing estimator.

This aspect of our toy model illustrates a general point: a statistical contaminant, such as unresolved point sources, affects the lensing detection by making a contribution to the bispectrum $b_{\ell_1\ell_2\ell_3}$ which may be coupled to the lensing bispectrum [Eq. (9)] which is measured by our estimator. Therefore, when trying to understand point source contamination, one should first ask: what bispectrum do point sources contribute?

We will actually consider a more general point source bispectrum than the Poisson form in Eq. (27), which relaxes two assumptions of the toy model. First, we have assumed that point sources do not cluster (i.e., are purely Poisson distributed). Furthermore, we have assumed that each CMB point source appears as an object in NVSS; there is a second case to consider in which the point sources do not actually appear as objects, but are merely clustered in a way which is correlated to NVSS.

Consider a population of clustered point sources which are tracers of a Gaussian field ρ. [We assume that the bias is absorbed into the definition of ρ, so that the probability of a point source at position x is $\propto (1 + \rho(x))$.] For our second case, where the point sources do not appear as NVSS objects, a short calculation shows that the point source bispectrum is

$$b_{\ell_1\ell_2\ell_3} = \langle S^2 \rangle n C_{\ell_3}^{\rho g}. \tag{29}$$

In the first case, where the sources do appear as NVSS objects, the bispectrum is given by

$$b_{\ell_1\ell_2\ell_3} = \frac{\langle S^2 \rangle n}{N} + \frac{\langle S^2 \rangle n^2}{N} C_{\ell_3}^{\rho\rho} + \frac{\langle S \rangle^2 n^2}{N}(C_{\ell_1}^{\rho\rho} + C_{\ell_2}^{\rho\rho}), \tag{30}$$

where $\langle S \rangle$ is the average temperature at CMB frequencies, n is the number density of the point source population, and N is the number density of NVSS.

In Eq. (30), the first term represents contributions from Poisson statistics, the second represents point source clustering on the galaxy angular scales ($\ell \sim 50$) which contribute to the lensing detection, and the third represents clustering on CMB angular scales ($\ell \sim 400$). We will assume that the third term is small compared with the first two and can be neglected. This is a critical assumption for our methodology and so we justify it carefully, giving two arguments.

The first argument is that a realistic point source clustering power spectrum $C_\ell^{\rho\rho}$ will be rapidly decreasing with ℓ and so the C_ℓ factors in the third term (with $\ell \sim 400$) will be small compared with the C_ℓ factor in the second term (with $\ell \sim 50$).

The second argument is more formal and shows that the third term in Eq. (30) is small compared to the first term. The ratio r of the third and first terms is given by

$$\begin{aligned} r &= \frac{\langle S \rangle^2 n}{\langle S^2 \rangle}(C_{\ell_1}^{\rho\rho} + C_{\ell_2}^{\rho\rho}) \leq n(C_{\ell_1}^{\rho\rho} + C_{\ell_2}) \\ &\leq N(C_{\ell_1}^{gg} + C_{\ell_2}^{gg}) \\ &\lesssim (1.59 \times 10^5)(2)(2.5 \times 10^{-7}) \\ &= 0.04. \end{aligned} \tag{31}$$

In the second line, we have used the fact that the contribution to the NVSS galaxy power spectrum C_ℓ^{gg} from the point source population alone is given by $\Delta C_\ell^{gg} = (n/N)C_\ell^{\rho\rho}$. In the third line, we have used our measurement of C_ℓ^{gg} (Fig. 8), which shows that $\ell C_\ell^{gg} \lesssim 10^{-4}$ for $\ell \gtrsim 400$. The intuition behind this formal argument is that if point source clustering were important on small angular scales, we would see this signal in the NVSS power spectrum.

We have now shown that the most general point source bispectrum is a combination of Eqs. (29) and (30) with the third term neglected. This motivates our final choice of point source estimator: we will use the three-point estimator optimized to detect any bispectrum of the form

$$b_{\ell_1\ell_2\ell_3} = F_{\ell_3}, \tag{32}$$

where F_{ℓ_3} is arbitrary (our estimator will estimate F_ℓ in bands). This generalizes the Poisson bispectrum considered previously [Eq. (27)].

We have shown that Eq. (32) is a sufficiently general form of the point source bispectrum to allow an arbitrary clustering power spectrum between point sources, an arbitrary cross correlation to the NVSS overdensity field g, and applies whether the CMB point sources actually appear as objects in NVSS, or are merely correlated to NVSS. Indeed, by putting an arbitrary ℓ_3 dependence in Eq. (32), we have been conservative by allowing a very general point source contribution. However, there is one caveat: we have assumed that point sources are biased tracers of Gaussian fields. Non-Gaussian contributions

from nonlinear evolution have not been included. In halo model language [84], we have incorporated one-halo and two-halo terms in the bispectrum but not the three-halo term.

Now that we have determined the most general bispectrum contributed by point source contamination [Eq. (32)], how do we construct the point source estimator? In Appendix B, we show that the optimal estimator for this bispectrum is constructed in a way which is analogous to the lensing estimator $\hat{C}_b^{\phi g}$ (or the curl null test $\hat{C}_b^{\psi g}$). First, we define a field $\tilde{s}$ which is quadratic in the CMB:

$$\sum_{\ell m} \tilde{s}_{\ell m} Y_{\ell m}(x) = \alpha(x)^2, \qquad (33)$$

where $\alpha(x)$ was defined previously in Eq. (15). Then we cross correlate $\tilde{s}$ to galaxy counts, subtracting the one-point term as usual:

$$\hat{C}_b^{sg} = \frac{1}{\mathcal{N}_b} \sum_{\substack{\ell \in b \\ -\ell \leq m \leq \ell}} (\tilde{s}_{\ell m} - \langle \tilde{s}_{\ell m} \rangle)^* (\tilde{g}_{\ell m}). \qquad (34)$$

This defines the optimal estimator $\hat{C}_b^{sg}$ for the point source bispectrum [Eq. (32)], with the galaxy multipole ℓ_3 binned into a bandpower b.

Intuitively, the field $\tilde{s}$ can be thought of as a "quadratic reconstruction" of CMB point source power, in the same sense that $\tilde{\phi}$ is a quadratic reconstruction of the CMB lensing potential. Our estimator $\hat{C}_b^{sg}$ is obtained by cross correlating $\tilde{s}$ to the filtered galaxy field $\tilde{g}$: we are only interested in point source power which is correlated to NVSS. By using $\hat{C}_b^{sg}$ to directly estimate the bispectrum due to point sources from data, we can assign systematic errors to the lensing bandpower $\hat{C}_b^{\phi g}$ which do not depend on the details of the point source model, as we will now see.

B. Results

In Fig. 13, we show the result of applying the point source estimator $\hat{C}_b^{sg}$, constructed in the previous section, to the WMAP and NVSS data sets. The χ^2 to zero is 11.7 with 12 degrees of freedom. Therefore, no evidence for point source contamination is seen. This lets us put strong constraints on the systematic error in lensing due to point sources: the point source contribution must be small enough to be hidden in Fig. 13, even though the estimator $\hat{C}_b^{sg}$ is optimized for point sources. The rest of this subsection is devoted to assigning systematic errors based on this observation.

We find that for distinct bands $b \neq b'$, the point source and lensing estimators in band b are uncorrelated to the estimators in band b'. This is unsurprising; it follows from the definitions that the bands are independent for all-sky coverage and homogeneous noise, so that the only correlation is due to inhomogeneities. Since we have large sky
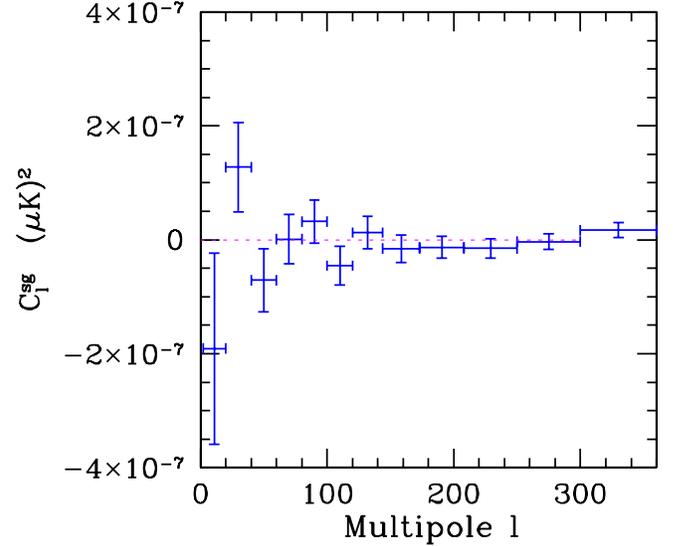
FIG. 13 (color online). Point source estimator $\hat{C}_b^{sg}$ applied to the WMAP and NVSS data sets, showing no evidence for CMB point source power which is correlated to NVSS. The error bars were obtained from Monte Carlo WMAP + NVSS simulations without point sources.

coverage and wide bands, the correlations should be small. We will treat each band independently, for consistency with our point source model, which allows an arbitrary ℓ dependence in the point source amplitude [Eq. (32)]. We will illustrate our method in detail for the band $b = (\ell_{\min}, \ell_{\max}) = (20, 40)$.

First, we use simulations to study the effect of point sources on the estimators $\hat{C}_b^{\phi g}$, $\hat{C}_b^{sg}$, using the following fiducial point source model. (We will show shortly that the final result does not depend on the details of the point source model.) Each simulated NVSS galaxy is assigned a randomly generated flux $S_{1.4 \text{ GHz}}$ between 2 mJy and 1 Jy, drawn from the distribution

$$\frac{dN}{dS} \propto \frac{S^{-1.8}}{1 + (S/200 \text{ mJy})^{1.1}}. \qquad (35)$$

This distribution was obtained empirically from the flux distribution seen in the real NVSS data (Fig. 14). We then assign the flux

$$S_\nu = \Lambda \left(\frac{\nu}{1.4 \text{ GHz}} \right)^\alpha S_{1.4 \text{ GHz}} \qquad (36)$$

at each WMAP frequency ν, where Λ is a constant which will be varied to simulate different overall levels of point source contamination. Following [76], we take spectral index $\alpha = 0$ in our fiducial point source model.

In Fig. 15, we show the values of the point source estimator $\hat{C}_b^{sg}$ obtained in an ensemble of simulations with varying point source amplitude Λ, and the *change* $\Delta\hat{C}_b^{\phi g}$ in the lensing bandpower which is due to the point source contribution. (Note that we do not show the true
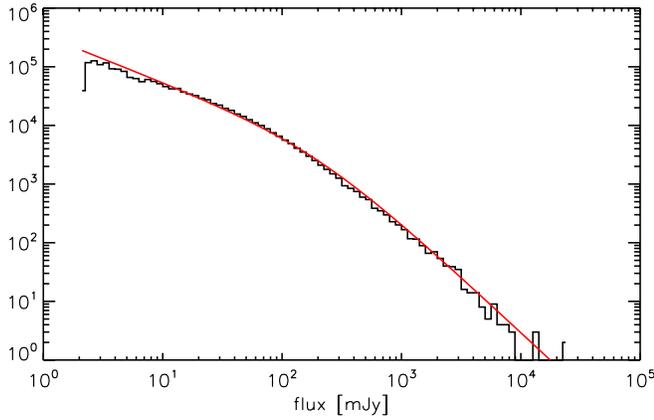
FIG. 14 (color online). Histogrammed 1.4 GHz flux distribution in NVSS, with the fitting function in Eq. (35) shown for comparison.

point source amplitude Λ for each simulation; we show the observed point source level $\hat{C}_b^{sg}$, estimated the same way as in the data.)

We find that the results can be fit by treating $\Delta \hat{C}_b^{\phi g}$ as a Gaussian variable whose mean and variance depend on $\hat{C}_b^{sg}$:

$$\langle \Delta \hat{C}_b^{\phi g} \rangle = -\alpha \hat{C}_b^{sg} \qquad \mathrm{Var}(\Delta \hat{C}_b^{\phi g}) = \beta^2 + \gamma^2 (\hat{C}_b^{sg})^2, \tag{37}$$
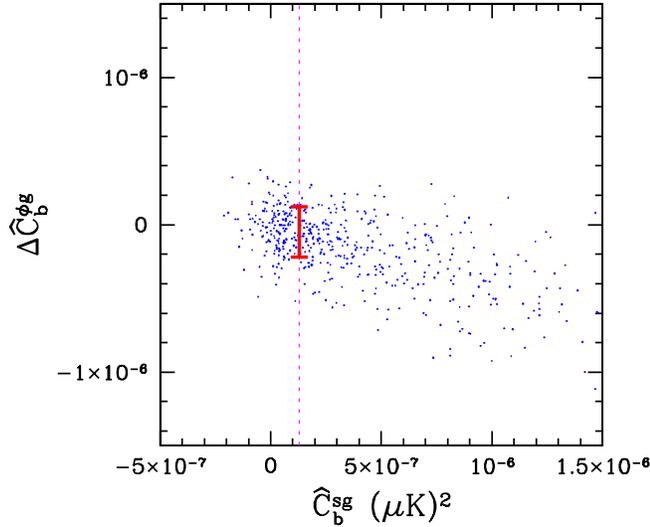
FIG. 15 (color online). Ensemble of simulations in the fiducial point source model [Eqs. (35) and (36)] with varying point source amplitude Λ. For each realization, we show the observed point source level $\hat{C}_b^{sg}$ in the band $b = (\ell_{\min}, \ell_{\max}) = (20, 40)$ and the change in the lensing estimator $\Delta \hat{C}_b^{\phi g}$ due to the point source contribution. The dotted vertical line shows the point source level in this band estimated from the real WMAP + NVSS data; the smaller vertical error bar shows the mean and rms $\Delta \hat{C}_b^{\phi g}$ among simulations whose observed point source level matches the measured value.

where $\alpha = 0.38 \ \mu\mathrm{K}^{-2}$, $\beta = 1.64 \times 10^{-7}$, and $\gamma = 0.21 \ \mu\mathrm{K}^{-2}$.

Based on this picture, how can we assign systematic errors due to point sources? Consider the distribution of $\Delta \hat{C}_b^{\phi g}$ values obtained by considering only realizations whose *observed* point source level $\hat{C}_b^{sg}$ agrees with the value ($ = 1.3 \times 10^{-7} \ \mu\mathrm{K}^2$) observed in the data (indicated by the dotted vertical line in Fig. 15). Note that this distribution includes realizations with a range of values for the true point source amplitude Λ; we are effectively averaging over point source levels allowed by the observed value of $\hat{C}_b^{sg}$ (i.e. the posterior distribution). By Eq. (37), we get a Gaussian distribution with parameters:

$$\Delta \hat{C}_b^{\phi g} = (-0.5 \pm 1.7) \times 10^{-7} \tag{38}$$

indicated by the vertical error bar in Fig. 15.

We have now arrived at a distribution [Eq. (38)] for the change in $\Delta \hat{C}^{\phi g}$ which is due to the point source contribution. The central value of this distribution is nonzero; point source contamination makes a negative contribution on average, as can be seen in Fig. 15. To be conservative, we will not shift our estimate for $\hat{C}^{\phi g}$ in the positive direction by the central value (this would allow point sources to "help" the lensing detection), but will include the shift as part of the systematic error. Thus, we would quote the systematic error in $C_b^{\phi g}$ as $\pm 2.2 \times 10^{-7}$.

As we have described it, this procedure appears to depend on the fiducial point source model [Eqs. (35) and (36)]. However, we find that the final systematic error estimate in each band is relatively robust even under drastic changes to the model. We tried the following extreme cases: assigning constant flux to each source rather than using Eq. (35), taking spectral index $\alpha = \pm 1$ in Eq. (36) rather than $\alpha = 0$, and finally simulating point sources which are merely correlated to NVSS rather than appearing as NVSS objects. All of these models give similar results to within a factor ~ 2. [Note that our point source estimator in Eq. (34) is actually optimized for point sources with a blackbody spectral distribution, but these results show that we obtain robust systematic error constraints across a reasonable range of spectral indices.]

Repeating this procedure for every ℓ band, we obtain a systematic error estimate for each lensing bandpower $\hat{C}_b^{\phi g}$. Since we have considered several point source models, we assign the systematic error for each band using the model which gives the largest error in that band. The results are shown in the "Resolved point source" column in Table I in Sec. IX. We find a systematic error which is smaller than the statistical error in all bands, but is the largest overall source of systematic error.

The robustness of our error estimate to the point source model is consistent with our discussion in the previous subsection: regardless of the details of the model, the contamination to the lensing estimator is proportional to

the level of the point source bispectrum [Eq. (32)] contributed by point sources. By directly estimating the bispectrum, we can obtain a relatively model-independent constraint on the systematic error due to point sources. This would not be possible if a simpler statistic were used, such as the cross power spectrum C_ℓ^{Tg}.

The procedure we have described is similar to the Fisher matrix based method that is frequently used to marginalize point sources when estimating primordial non-Gaussianity from the CMB bispectrum [61], but differs in several details. First, we use a general form of the point source bispectrum [Eq. (32)] which allows point source clustering, and also allows CMB point sources to appear or not appear as NVSS objects. Second, we do not shift the lensing estimator by the central value of the posterior distribution in Eq. (38), but treat the shift as part of the systematic error. Finally, the Fisher matrix formalism would not predict the increased *variance* in $\Delta \hat{C}_\ell^{sg}$ in the presence of point sources [Eq. (37)]. This is included in the Monte Carlo based procedure presented here. The Fisher matrix does predict the overall negative slope in Fig. 15, which is a property of the point source and CMB lensing bispectra. As a check, if we directly compute the Fisher matrix [see Eq. (43) below], we find a weak negative correlation (≈ -0.1 in each band) between the lensing and point source shapes.

C. Resolved point sources

Now that we have analyzed systematic errors in lensing from unresolved CMB sources, we consider resolved sources. Resolved CMB point sources have been treated in the pipeline by simply masking each source (Sec. II). If the sources are correlated to radio galaxies, so that the WMAP *mask* is correlated to NVSS, one may wonder whether the masking procedure can bias the lensing detection.

We can prove the following general result (Appendix C): in the absence of CMB lensing, correlations between the mask and galaxy field cannot fake the lensing signal, i.e. the expectation value $\langle \hat{C}_b^{\phi g} \rangle$ is zero even if the mask is correlated. Interestingly, our proof depends on the presence of the one-point term in the estimator [Eq. (17)] and does not rule out the possibility of bias if this term is omitted.

Given this general result, the lowest-order effect that might be expected from mask-galaxy correlations is a bias proportional to the lensing signal, i.e. a calibration error. We looked for a calibration error in simulations, by randomly generating a point source mask by assigning a point source to pixel x with probability

$$\rho(x) \propto \begin{cases} \delta g(x) & \text{if } \delta g(x) > 0 \\ 0 & \text{otherwise.} \end{cases} \qquad (39)$$

[This is an extreme case, corresponding to a linear bias model $\rho(x) \propto 1 + b(\delta g(x))$ in the maximally biased limit $b \to 0$.] An example of this simulation procedure is shown in Fig. 16.
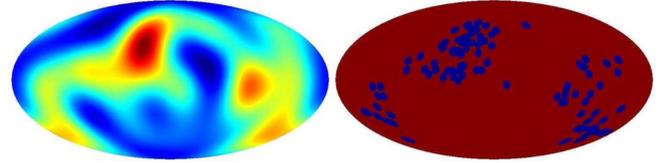
FIG. 16 (color online). Illustration of our procedure [Eq. (39)] for simulating correlations between the NVSS galaxy field (left) and source mask (right). For visibility, we have bandlimited the galaxy field δg to $\ell \leq 6$, and used 100 sources with masking radius 4° rather than the mask parameters of the data sets (Sec. II).

With the source mask density of the real data sets (Sec. II), we see no evidence for a calibration error after 1024 Monte Carlo simulations of the full pipeline. The same result was obtained replacing the NVSS overdensity δg by the lensing potential ϕ on the right-hand side of Eq. (39), or bandlimiting the right-hand side for several choices of ℓ band.

Since we do not have a general proof that the calibration error is small, we can only conclude that it is smaller than the ~3% statistical limit from our Monte Carlo sample. In Table I, we have assigned each bandpower a 3% systematic calibration error in the "Resolved point sources" column, but we see no evidence for the effect and it may be much smaller.

VIII. SUNYAEV-ZEL'DOVICH FLUCTUATIONS

A further source of possible contamination of the WMAP-NVSS correlation comes from rescattering of the primordial microwave background off hot electrons inside the large-scale structure field that also underlies the distribution of NVSS sources. The largest effect is the thermal Sunyaev-Zel'dovich (SZ) effect [22,23], due to inverse Compton scattering which shifts photons away from their originally blackbody spectrum. The kinetic Sunyaev-Zel'dovich effect, due to Doppler scattering of CMB photons by large-scale structure moving along the line of sight, is expected to be a concern for lensing reconstruction with future CMB experiments that are able to frequency clean the thermal effect [85,86]. On the angular scales relevant for WMAP, the kinetic effect is much smaller and more Gaussian than the thermal effect, and we neglect it in the following analysis.

The induced temperature change of the thermal SZ compared to the CMB, $\Delta T(\hat{n})/T_{\text{CMB}} = g(\nu)y$, is proportional to the line of sight integral over the cluster gas pressure, $y = \int dl n_e \sigma_T (k_B T / m_e c^2)$ (the Compton-y parameter), where n_e is the free electron density, k_B the Boltzmann constant, T_e the electron temperature, m_e the electron mass, and σ_T the Thomson scattering cross section. It also has a characteristic frequency dependence, given in terms of the dimensionless frequency $x = (h\nu/k_B T)$ by

$$g(x) = x \frac{e^x + 1}{e^x - 1} - 4. \tag{40}$$

This frequency dependence causes $\simeq 13(18)\%$ changes in the expected amplitude of the SZ between the WMAP V (Q) and W channels. These differences are smaller than the statistical error of our WMAP-NVSS cross-correlation measurement, making it impossible to distinguish the SZ effect from lensing on frequency basis alone.

We therefore rely on angular separation. Our preferred way to describe the SZ effect and assign systematic errors would be to use full hydrodynamical simulations of the effect (e.g. [87–89]). Unfortunately these have to date only been performed on scales of $\simeq 100$ comoving megaparsec, allowing modeling of secondary anisotropies on scales of only a few square degrees. Our lensing estimator on the other hand receives contributions from $\ell \gtrsim 20$, requiring simulation on scales substantially larger than 10×10 square degrees. A somewhat less computationally expensive route would be to establish halo catalogues based on perturbation theory schemes (e.g. [90,91]) that are then decorated with semianalytic gas-pressure profiles. Even these procedures are however very costly for our purposes of covering 40 000 square degrees on the sky at a depth of about 4 comoving gigaparsec, under the necessary requirement of resolving halos down to 10^{13} solar masses in order to reliably model SZ fluctuations below $l = 1000$ [92].

As we will argue in this section however, on the scales relevant to a lensing detection using WMAP, SZ contamination can be treated as part of the point source contribution which has been studied in the previous section.

To begin with, notice that although at WMAP frequencies SZ clusters contribute a temperature decrement to the CMB, their contribution to the point source estimator $\hat{C}_b^{sg}$ is positive, because the estimator is quadratic in the CMB. Therefore our "point source" estimator will be able to serve as a monitor for the sum of point source and SZ contamination. This is yet another advantage of the three-point estimator over the cross spectrum C_ℓ^{Tg} discussed in Sec. VII A: because point sources make a positive contribution to the cross spectrum but the SZ contribution is negative, the cross spectrum cannot constrain both contaminants.

Next consider the spatial distribution of SZ. The vast majority of the thermal SZ signal stems from collapsed regions with a gas density contrast of hundreds of times the mean density of the universe (see e.g. [88]). If cluster profiles could be approximated as δ-functions, then they could be treated as biased tracers of large-scale structure that is correlated to NVSS galaxies. Since our point source model [Eq. (32)] allows clustering and cross correlation to NVSS, this would allow us to treat the SZ contribution as part of the point source contribution.

To quantify the deviation from pointlike profiles, in Fig. 17, we show galaxy cluster profiles in angular multipole space, calculated with the universal gas-pressure pro-
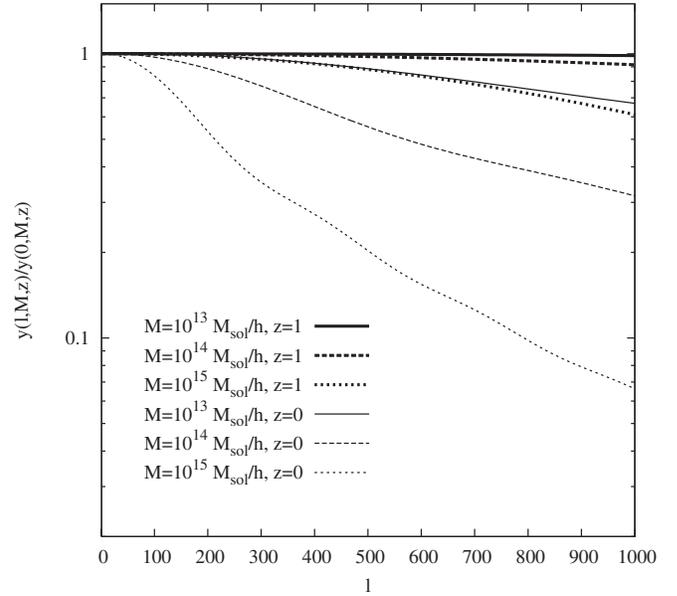
FIG. 17. The Compton-y profile for three different cluster masses at $z = 1$ (thick lines) and $z = 0.1$ (thin lines). The profiles have been normalized to 1 at $l = 0$ to facilitate comparison. According to this panel, at high redshift it may be possible to approximate even rare and massive clusters as point sources on the scale where our lensing estimator gathers most of its information, $l \simeq 400$.

file model of [93], at $z = 0.1$ and $z = 1.0$. This redshift range is chosen to span roughly the range where the SZ might be correlated to NVSS sources. It can be seen that many of the relevant clusters fall below the angular scale ($\ell \sim 400$) where our lensing reconstruction gathers most of its information, but some large nearby SZ clusters have profiles as extended as tens of arcminutes, and show some slope at the relevant angular scales.

To determine whether this slope is important at WMAP resolution, we consider the angular power spectrum C_ℓ^{SZ}, which is an average over redshift and mass of all clusters. In cross correlation with NVSS, this integral would be modulated by the source redshift number density. Since the NVSS redshift distribution is not very well understood, here we apply uniform weight to all objects to obtain an estimate for the scale dependence of the power spectrum. We calculate the power spectrum including both the Poisson (1-halo) and clustering (2-halo) contributions, following the formalism of [92,94]. The results are shown (for the low frequency limit in which $y = -2$) in Fig. 18.

It is seen that the SZ power spectrum is not flat at $\ell \sim 400$, owing to the contribution of the most massive and nearby clusters, but has the rough scaling

$$C_\ell \propto \ell^{-1.2} \tag{41}$$

over the relevant range of angular scales.

We can incorporate this scale dependence into the analysis by considering a bispectrum of the form
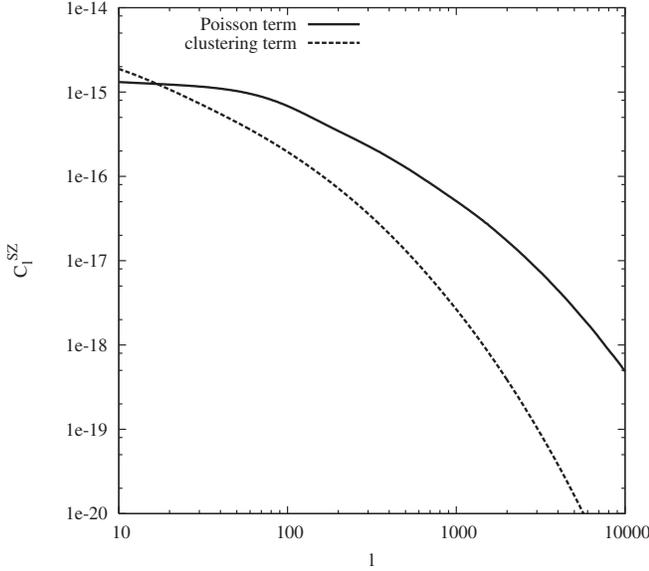
FIG. 18.   The thermal Sunyaev-Zel'dovich angular power spectrum contributions (in the Rayleigh-Jeans limit) from Poisson and clustering terms. On the scale of most interest for our lensing reconstruction, $l \simeq 400$, the SZ Poisson term dominates by an order of magnitude over the clustering part. The angular power spectra were calculated using the gas-pressure profile model by [92,93].

$$b_{\ell_1, \ell_2, \ell_3} \propto \ell_1^{-0.6} \ell_2^{-0.6} F_{\ell_3}. \tag{42}$$

To quantify the effect of scale dependence on the lensing estimator, we compute the correlation between this shape and the point source shape [Eq. (32)], using the Fisher matrix formalism [95]. According to this, the Fisher matrix element between two bispectra $b_{\ell_1 \ell_2 \ell_3}^{(\alpha)}$, $b_{\ell_1 \ell_2 \ell_3}^{(\beta)}$ is defined by

$$F_{\alpha\beta} = \frac{1}{2} \sum_{\ell_1 \ell_2 \ell_3} \frac{(\mathcal{G}_{\ell_1 \ell_2 \ell_3})^2 b_{\ell_1 \ell_2 \ell_3}^{(\alpha)} b_{\ell_1 \ell_2 \ell_3}^{(\beta)}}{(C_{\ell_1}^{TT} + N_{\ell_1}^{TT})(C_{\ell_2}^{TT} + N_{\ell_2}^{TT})(C_{\ell_3}^{gg} + N_{\ell_3}^{gg})}. \tag{43}$$

To a good approximation, when bispectra are estimated from data, the covariance matrix is given by

$$\text{Cov}\,(b^{(\alpha)}, b^{(\beta)}) = f_{\text{sky}}^{-1} F_{\alpha\beta}^{-1}. \tag{44}$$

When we compute the Fisher matrix for the point source [Eq. (32)] and scale-dependent [Eq. (42)] shapes at WMAP and NVSS noise levels, we find a correlation coefficient ~ 0.95. At this level of correlation, the point source shape and SZ shape cannot be distinguished to 1σ, unless a 6σ detection of the point source shape can also be made. Since we do not find any evidence for point source contamination in the data (Fig. 15), we conclude that the difference between the point source and SZ bispectra should be negligible in the context of the WMAP and NVSS data sets.

As an additional check, we tried modifying our point source simulations by giving each point source an $a_{\ell m} \propto \ell^{-0.6}$ profile, and SZ frequency dependence [Eq. (40)], including the negative sign. This crude procedure is of course not an accurate method for simulating SZ in detail, but does incorporate two qualitative features which distinguish SZ from point sources at WMAP resolution: the scale dependence [Eq. (41)] and frequency dependence [Eq. (40)]. We find that the systematic errors in lensing (obtained from Monte Carlo simulations as described in Sec. VII) are within the range of point source models previously considered, showing that neither of these deviations from pure point source behavior significantly affects our method.

Finally, there is one assumption in our point source model which we can check explicitly for the case of SZ: that clustering is unimportant on scales of $l \simeq 400$ [see Eq. (30)]. This can be seen directly from Fig. 18; the clustering term is dominated by the Poisson term by an order of magnitude.

IX. FINAL RESULT AND DISCUSSION

In Table I and Fig. 19, we show our final result: estimated $C_\ell^{\phi g}$ bandpowers, together with statistical and systematic uncertainties. Our procedure for combining errors is as follows. We combine the errors from beam asymmetry (Sec. VI A) and beam uncertainty (Sec. VI B) into a "total beam" error assuming that the two are completely correlated. We obtain a "total galactic" error from galactic CMB foregrounds by combining the dust and free-free systematic errors (Sec. VI C) assuming correlated errors, and double the result to account for synchrotron (where no template is available on the relevant angular scales). We obtain a "total point source" error by combining the errors from unresolved and resolved sources, assuming that the two are correlated. (As we have shown in Sec. VIII, the "point source" errors apply to the total systematic error from CMB point sources and the thermal SZ effect.) We then obtain our final result by combining the statistical, total beam, total galactic, and point source errors, assuming that the four are uncorrelated.

What is the total statistical significance of our detection? To assess this, we combine our bandpower estimates into a single estimator $\hat{\mathcal{C}}$, giving each bandpower a weight proportional to its *fiducial* expectation value $C_{b,\text{fid}}^{\phi g}$ (not the measured value in Table I) and inversely proportional to its total (statistical + systematic) variance:

$$\hat{\mathcal{C}} = \frac{\sum_b (C_{b,\text{fid}}^{\phi g} / \text{Var}(\hat{C}_b^{\phi g})) \hat{C}_b^{\phi g}}{\sum_b (C_{b,\text{fid}}^{\phi g})^2 / \text{Var}(\hat{C}_b^{\phi g})}, \tag{45}$$

where the denominator has been included to normalize $\langle \hat{\mathcal{C}} \rangle = 1$ in the fiducial model. We find $\hat{\mathcal{C}} = 1.15 \pm 0.34$, i.e. a 3.4σ detection.

TABLE I.   Final estimated $C_b^{\phi g}$ bandpowers, together with statistical uncertainties and systematic errors from point sources. All entries in the table are $\ell^2 C_\ell^{\phi g}$ in multiples of 10^{-7}.

(ℓ_{min}, ℓ_{max})	Statistical	Beam Asymmetry	Beam Uncertainty	Beam Total	Galactic Dust	Galactic Free-free	Galactic Total	Point source + SZ Unresolved	Point source + SZ Resolved	Point source + SZ Total	Stat + systematic
(2, 20)	17.4 ± 22.4	± 0.9	± 0.3	± 1.2	± 0.4	± 1.4	± 3.6	± 10.9	± 0.5	± 11.4	17.4 ± 27.4
(20, 40)	33.2 ± 10.5	± 0.2	± 0.1	± 0.3	± 0.2	± 0.5	± 1.4	± 4.9	± 1.0	± 5.9	33.2 ± 13.0
(40, 60)	15.9 ± 7.8	± 0.1	± 0.1	± 0.2	± 0.2	± 0.3	± 1.0	± 2.8	± 1.5	± 4.3	15.9 ± 9.3
(60, 80)	10.1 ± 6.3	± 0.1	± 0.1	± 0.2	± 0.1	± 0.3	± 0.8	± 2.0	± 0.3	± 2.3	10.1 ± 7.0
(80, 100)	5.1 ± 5.8	± 0.1	± 0.1	± 0.2	± 0.1	± 0.3	± 0.8	± 1.1	± 0.2	± 1.3	5.1 ± 6.0
(100, 130)	8.3 ± 4.3	± 0.1	< 0.1	± 0.2	± 0.1	± 0.2	± 0.6	± 0.6	± 0.2	± 0.8	8.3 ± 4.4
(130, 200)	1.6 ± 2.5	< 0.1	< 0.1	± 0.1	± 0.1	± 0.1	± 0.4	± 0.3	± 0.1	± 0.4	1.6 ± 2.6
(200, 300)	-1.9 ± 2.2	< 0.1	< 0.1	± 0.1	± 0.1	± 0.1	± 0.4	± 0.3	± 0.1	± 0.4	-1.9 ± 2.3

Throughout this paper, we have assumed a fiducial cosmology, NVSS redshift distribution, and galaxy bias when computing statistical errors by Monte Carlo simulation, and when constructing the $(S + N)^{-1}$ filters in the analysis pipeline. To what extent do our results depend on the fiducial model? Our $C_\ell^{\phi g}$ bandpowers and error bars depend only on the fiducial power spectra C_ℓ^{TT}, C_ℓ^{gg} used in Monte Carlo simulations, not on the details of the modeling. We have checked these fiducial spectra in two ways: first, by direct comparison with the measured NVSS power spectrum (Fig. 8); we have omitted the comparison for the WMAP power spectrum since our fiducial cosmology is the WMAP + ALL cosmology from [2]. Second, we have shown that consistent statistical errors are obtained by cross-correlating simulations with data (Fig. 5). The fiducial model is also used to construct the $(S + N)^{-1}$ filtering operation, but in this case using incorrect power spectra merely makes our estimator slightly suboptimal and does not significantly affect the detection.
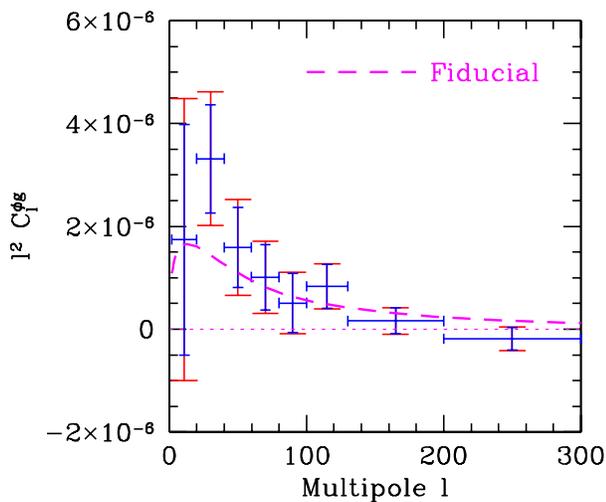
FIG. 19 (color online).   Final result from Table I, showing statistical errors alone (blue/inner error bars) and statistical + systematic errors (red/outer).

The statement that our result only depends on the fiducial spectra C_ℓ^{TT}, C_ℓ^{gg}, not on the details of the model, would not be true if we were attempting to translate our measurement of $C_\ell^{\phi g}$ into a constraint on cosmological parameters. There are several obstacles to doing so which we plan to address in future work. First, $C_\ell^{\phi g}$ depends on cosmology but is also proportional to the NVSS galaxy bias b_g, which must be marginalized. One possible approach is to only consider quantities such as

$$C_\ell^{\phi g} / \sqrt{C_\ell^{gg}} \tag{46}$$

which should be independent of galaxy bias (ignoring subtleties like redshift-dependent bias). Second, the NVSS redshift distribution dN/dz is uncertain and must also be marginalized over some reasonable range. We note that the auto power spectrum C_ℓ^{gg}, which appears in Eq. (46), is more sensitive to changes in dN/dz than the cross spectrum $C_\ell^{\phi g}$. A conservative approach to marginalizing over cosmological parameters as well as redshift and bias uncertainties would be the Markov chain Monte-Carlo method (compare [96]) applied to both $C_\ell^{\phi g}$ and C_ℓ^{gg} constraints.

Finally, we have not considered the impact of magnification bias: the observed NVSS galaxy field is altered by the magnifying and demagnifying effect of gravitational lenses between the source galaxies and observer [97,98]. One can think of this as adding terms to the galaxy field $g(\hat{n})$ which depend on the matter distribution at intermediate redshifts along the line of sight. This introduces additional terms in $C_\ell^{\phi g}$ which are not included in our fiducial spectrum, and have been shown to be significant when deducing cosmological constraints from ISW measurements [99]. In a magnified region, the galaxy surface density $g(\hat{n})$ receives a negative contribution (since magnification spreads a fixed number count over a larger area) and a positive contribution (since magnification brings new galaxies above the flux threshold of the survey), so the effect can have either sign. Note that magnification bias affects the fiducial $C_\ell^{\phi g}$ in a given cosmology, but does not

affect our measured values of $C_\ell^{\phi g}$ or the statistical significance of the detection.

We have constructed an estimator for the lensing cross correlation $C_\ell^{\phi g}$ which is probably optimal (Appendices A and B). The estimator is defined in three steps. First, we filter the WMAP and NVSS data sets by their inverse signal + noise covariance, thus "distilling" the data sets to harmonic-space maps $\tilde{a}_{\ell m}$, $\tilde{g}_{\ell m}$. Second, we perform lens reconstruction on the filtered WMAP data $\tilde{a}_{\ell m}$, producing a noisy reconstruction $\tilde{\phi}_{\ell m}$ of the CMB lensing potential which is quadratic in the data. Third, we cross correlate $\tilde{\phi}$ and $\tilde{g}$, subtracting the one-point term.

Subtracting the one-point term is necessary to make the estimator optimal, and also eliminates systematic bias from resolved point sources (Sec. VII C), although a systematic calibration error may remain. Since the one-point subtraction is trivial to implement in a Monte Carlo pipeline, we recommend that it always be used. The other feature making our estimator optimal is full-blown $(S + N)^{-1}$ filtering (Appendix A). Here, it is unclear whether the optimal filter is practically necessary; it may be possible to construct a simpler filter which approximates $(S + N)^{-1}$ and produces near-optimal estimates in practice. In any case, an optimal implementation is an invaluable tool when studying candidates for such a filter, since the results can be directly compared to optimal.

We have studied potential sources of systematic error from known NVSS systematics (Sec. V), WMAP beam effects (Sec. VI A and VI B) galactic microwave foregrounds (Sec. VI C), point sources (Sec. VII), and the thermal Sunyaev-Zel'dovich effect (Sec. VIII). Error estimates for each of these systematics have been included in our final result.

The most problematic systematic for CMB lensing, at least when measured in cross correlation to large-scale structure, seems to be point source contamination. In general, a statistical contaminant such as point sources affects the lensing detection by contributing some bispectrum $b_{\ell_1 \ell_2 \ell_3}$ which may be correlated to the lensing bispectrum which our estimator measures [Eq. (9)]. We therefore treat point sources by directly estimating the point source bispectrum from the data, to monitor the level of contamination and assign systematic errors. We allow a form of the point source bispectrum [Eq. (32)] which is sufficiently general to include a wide range of point source models, including clustered sources and sources which may or may not appear as objects in NVSS.

We have argued that at WMAP sensitivity levels, thermal Sunyaev-Zel'dovich fluctuations due to hot gas in clusters of galaxies can be treated as part of the point source contribution. We checked that the level of scale dependence in the bispectrum, introduced by large nearby objects, is unimportant at WMAP resolution, but we do not expect this to be the case for smaller-scale experiments such as Planck [100], ACT [25], or SPT [26], which will

begin to observe the sky in the near future. In fact, even the qualitative trends we have found in Table I for systematic error contributions may be different for these future surveys, which will probe new regimes of sensitivity and resolution. The detection from WMAP that has been presented here is a milestone toward detailed measurements of CMB lensing that lie ahead.

ACKNOWLEDGMENTS

APPENDIX A: FAST $(S + N)^{-1}$ FILTERING

In this Appendix, we present the details of our method for computing the inverse signal + noise weighted map $\tilde{a} = (S + N)^{-1}a$, for either the WMAP or NVSS data.

Outside the context of lens reconstruction, this inversion problem also arises for other types of optimal analysis in which the data is weighted by inverse signal + noise, e.g. optimal power spectrum estimation [67], power spectrum analysis by Gibbs sampling [68,69], and bispectrum estimation [63]. We expect that our method will be useful in these contexts as well.

1. Conjugate gradient inversion

First, let us introduce some notation. We assume a data set which is specified by N_{chan} pixel-space maps, with a common underlying harmonic-space signal $s_{\ell m}$. Thus we can write

$$d_i^{pix} = A_i s + (noise), \tag{A1}$$

where A_i is the pointing matrix associated to the ith channel.

This generality is sufficient to describe both the WMAP and NVSS data sets. For WMAP, we have $N_{chan} = 8$

corresponding to the eight Q-, V-, and W-band differencing assemblies used in the analysis, the signal $s_{\ell m}$ is the noiseless CMB, and each pointing matrix A_i includes convolution with the pixel window function and beam of the corresponding DA. Our convention is that the signal s is defined in harmonic space, while the data d_i^{pix} is defined in pixel space. Thus the operator A_i in Eq. (A1) is defined by applying beam and pixel window functions to the harmonic-space signal [see Eq. (12)], then taking the spherical transform to produce a map in pixel space. For NVSS, we have $N_{\mathrm{chan}} = 1$ corresponding to a single galaxy count map, with no beam convolution included in the pointing matrix A, since the 45-arcsec NVSS beam can be neglected on angular scales ($\ell \lesssim 250$) which contribute to the lensing estimator.

In [101], it is shown that the data in Eq. (A1) can be reduced to a single harmonic-space map a, with associated noise covariance matrix N, without losing information. The map a and matrix N are defined by the pair of equations

$$N^{-1} = \sum_i A_i^T (N_i^{\mathrm{pix}})^{-1} A_i \qquad (A2)$$

$$N^{-1}a = \sum_i A_i^T (N_i^{\mathrm{pix}})^{-1} d_i^{\mathrm{pix}}, \qquad (A3)$$

where N_i^{pix} is the noise covariance associated to the ith map.

Let us first assume a noise model (which we will generalize in Appendix A 3) such that the inverse noise covariance $(N_i^{\mathrm{pix}})^{-1}$ in the ith map is diagonal in pixel space. For WMAP, this is the noise model used to analyze the temperature power spectrum [46]; for NVSS, the diagonal noise covariance represents shot noise and is constant between pixels. (In both cases, a sky cut is incorporated by setting N^{-1} to zero inside the mask.) In this noise model, it is trivial to compute $N^{-1}a$ using Eq. (A3), but what we need in our analysis pipeline is $\tilde{a} = (S + N)^{-1}a$. Note that N^{-1} is generally not invertible due to the presence of unconstrained modes (such as pixels excluded by the sky cut), so that a is not determined by Eq. (A3), but the data do determine $N^{-1}a$, and having this is sufficient for $\tilde{a}$. The remainder of this Appendix is devoted to an algorithm for computing $\tilde{a}_{\ell m}$.

Following [69], we will find it convenient to replace the matrix $(S + N)^{-1}$ by the matrix X^{-1}, where

$$X \stackrel{\mathrm{def}}{=} 1 + S^{1/2}N^{-1}S^{1/2} = 1 + \sum_i S^{1/2}A_i^T (N_i^{\mathrm{pix}})^{-1} A_i S^{1/2}. \quad (A4)$$

Using the identity $(S + N)^{-1}a = S^{-1/2}X^{-1}S^{1/2}N^{-1}a$, it suffices to give an algorithm for multiplying a map by X^{-1}. Since the number of degrees of freedom is too large for direct matrix inversion, this multiplication must be performed using conjugate gradient (CG) inversion [102]. Performance of the conjugate gradient method depends on

a good choice of preconditioner, or linear operator which is efficient to compute and approximates X^{-1}.

A common way to construct a preconditioner is to replace X by some simpler approximation X' which can be inverted exactly, and use $(X')^{-1}$ as the preconditioner. The simplest preconditioner of this type would be X_Δ^{-1}, where X_Δ is the matrix defined by keeping only the diagonal of X.

With this diagonal preconditioner, we have found that the conjugate gradient search will eventually converge, but the convergence is extremely slow. To understand why it is slow, note that X_Δ^{-1} will only be a good approximation to X^{-1} when X is diagonally dominated. This will be the case on angular scales which are noise dominated ($S/N \ll 1$), since X will be close to the identity matrix, but on large angular scales where the signal dominates, the preconditioner is not a good approximation to X^{-1}, and the convergence rate becomes limited by these scales.

This picture motivates the following improved preconditioner, which has been used in several previous treatments [44,103]. Define the matrix X_0 by keeping all matrix entries in the dense block corresponding to multipoles (ℓ, m) satisfying $\ell \leq \ell_{\mathrm{split}}$. Then consider the preconditioner

$$\begin{pmatrix} X_0^{-1} & 0 \\ 0 & X_\Delta^{-1} \end{pmatrix}, \qquad (A5)$$

obtained by keeping dense matrix entries below ℓ_{split} and the diagonal above ℓ_{split}. [In practice, the choice of ℓ_{split} is usually dictated by memory limitations, since $\mathcal{O}(\ell_{\mathrm{split}}^4)$ storage is needed to store X_0 in dense form.] In this section, we will refer to (A5) as the "block preconditioner."

We have found that the block preconditioner is very efficient for the NVSS data set, but slow to converge for WMAP. If we terminate the CG search as soon as we find an approximate solution $a' \approx X^{-1}a$ such that the termination criterion $|a - Xa'|/|a| < 10^{-6}$ is satisfied, then block preconditioning requires ~ 3.5 CPU-hours to converge for the three-year WMAP data set with Kp0 mask, and distinct beam transfer functions for each of the eight differencing assemblies in Q-, V-, and W-band.

The slow convergence of this preconditioner is a bottleneck for our lens reconstruction analysis and has also been identified as a limiting factor in other contexts, e.g. Gibbs sampling [68,69]. Therefore, a faster method is desirable.

2. Multigrid preconditioner

So far, we have recalled existing work in the literature: fast $(S + N)^{-1}$ filtering can be performed via conjugate gradient inversion with the block preconditioner [Eq. (A5)]. In this section, we present our improvement. The idea is that, even with the block preconditioner to do the inversion exactly at multipoles below ℓ_{split}, conjugate gradient inversion is still limited by the convergence rate at multipoles just above ℓ_{split} (since the lowest multipoles will have highest signal-to-noise). However, these are precisely

the multipoles which can be represented in a coarser pixelization.

This leads naturally to a multigrid preconditioner: one preconditions the inversion at resolution N_{side} using the result of performing the inversion at coarser resolution $N_{\text{side}}/2$, where the spherical transform is faster by a factor of ~8. This process is recursive; the inversion at resolution $N_{\text{side}}/2$ is preconditioned by an inversion at resolution $N_{\text{side}}/4$, and so on. At the coarsest resolution (typically $N_{\text{side}} = 128$), the inversion is preconditioned using the block preconditioner. For the WMAP example with parameters as described at the end of Appendix A 1, we find a running time of 14 CPU-minutes using the multigrid preconditioner. This represents an improvement by a factor ~15, relative to the block preconditioner alone.

In detail, the multigrid method works as follows. As described in the preceding section, we wish to compute $X^{-1}a$, where $a = a_{\ell m}$ is defined in harmonic space up to some maximum multipole ℓ_{max}, and X is defined by Eq. (A4). Then let $X_{(1)}$ be the matrix defined analogously, with all noise covariance matrices "coarsified" (i.e. with N_{side} decreased by a factor of 2), and with the maximum multipole reduced to some $\ell_{\text{max}}^{(1)} < \ell_{\text{max}}$. Then the multigrid preconditioner is defined by

$$\begin{pmatrix} X_{(1)}^{-1} & 0 \\ 0 & X_\Delta^{-1} \end{pmatrix}, \qquad (A6)$$

i.e. we use the diagonal preconditioner for multipoles above $\ell_{\text{max}}^{(1)}$. Since applying the preconditioner involves a multiplication by $X_{(1)}^{-1}$, and the matrix $X_{(1)}$ is too large for dense inversion, we do the $X_{(1)}^{-1}$ multiplication recursively, using an "inner" instance of conjugate gradient inversion. The preconditioner for the inner CG inversion is obtained analogously by a second round of coarsifying noise covariance matrices and reducing the maximum multipole to some $\ell_{\text{max}}^{(2)} < \ell_{\text{max}}^{(1)}$, and so on. At the coarsest resolution,

we use the block preconditioner described in the preceding subsection.

In Fig. 20, we show the preconditioner chain for WMAP. The parameters $N_{(i)}$, $\epsilon_{(i)}$ control the termination criterion for each CG instance; when evaluating $X_{(i-1)}^{-1}$ with preconditioner $X_{(i)}^{-1}$, we terminate the CG search after $N_{(i)}$ iterations, or when the approximate solution $a' \approx X^{-1}a$ satisfies $|a - Xa'|/|a| < \epsilon_{(i)}$. We have found that it is necessary to include these parameters to avoid spending too much CPU time in the coarse grids. In the WMAP3 example, the first level of preconditioning actually does not reduce the resolution, but instead reduces the number of distinct beams in the problem from eight to two (by making the so-called "equal-beam approximation" in which the average of the beam transfer functions is used). Note that the final output of the inversion does not make the equal-beam approximation, but merely uses inversions with the equal-beam approximation internally, to precondition the top-level CG inversion where no such approximation is made.

It is illuminating to describe the sequence of coarsifying and decoarsifying operations which occur in the multigrid method. Each iteration of the top-level CG loop requires one evaluation of its preconditioner, which in turn is a full-blown CG search (at coarser resolution) which can iterate up to $N_{(1)}$ times. Each of these iterations can iterate at the next coarsest resolution up to $N_{(2)}$ times, and so on. In the parlance of multigrid algorithms, this exponential fanout is referred to as a W-cycle (Fig. 21). Note that, even though the number of iterations spent at each resolution increases exponentially, the total CPU time does not, because the running time of each iteration is exponentially suppressed; in each level, the resolution and value of ℓ_{max} are typically reduced by a factor of 2, which reduces the cost of the spherical harmonic transform by a factor of 8. Indeed, the strength of the multigrid method is that it spends an exponentially large number of CG iterations on the large

| $N_{\text{side}} = 512$ | $\Longrightarrow$ | $N_{\text{side}} = 512$ | $\Longrightarrow$ | $N_{\text{side}} = 256$ | $\Longrightarrow$ | $N_{\text{side}} = 128$ | $\Longrightarrow$ | $\ell_{\text{split}} = 60$ |

$\ell_{\text{max}} = 1000$ $\quad$ $\ell_{\text{max}}^{(1)} = 1000$ $\quad$ $\ell_{\text{max}}^{(2)} = 400$ $\quad$ $\ell_{\text{max}}^{(3)} = 200$

$N_{(1)} = 8$ $\quad$ $N_{(2)} = 10$ $\quad$ $N_{(3)} = 5$

$\epsilon_{(1)} = 10^{-3}$ $\quad$ $\epsilon_{(2)} = 0.1$ $\quad$ $\epsilon_{(3)} = 0.1$
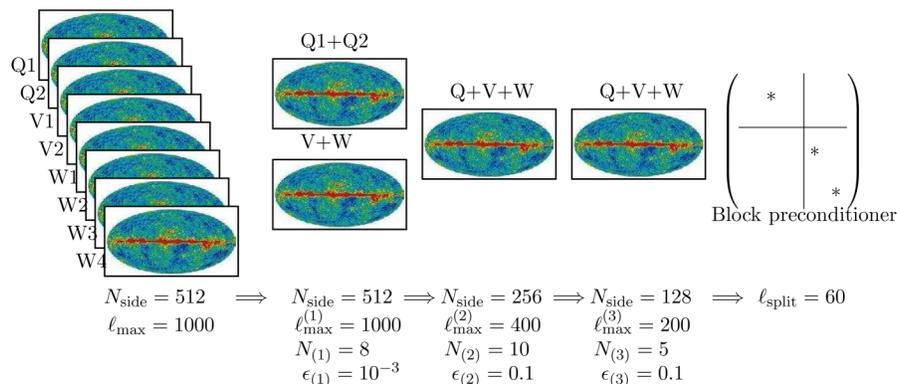
FIG. 20 (color online). Preconditioner chain for multigrid $(S + N)^{-1}$ filtering, using noise maps from the three-year WMAP data set. From left to right, each set of maps represents one conjugate gradient inversion problem, which is preconditioned by the "faster and cruder" approximation which appears next in the chain, obtained by either reducing resolution or the number of distinct beams retained in the problem.
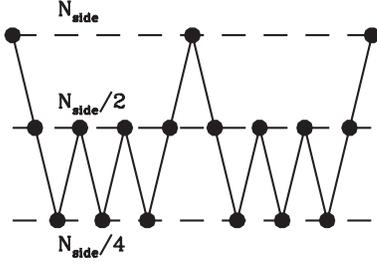
FIG. 21. Sequence of coarsifying and decoarsifying operations in an instance of the multigrid method with $N_{(1)} = 3$, $N_{(2)} = 2$, showing the W-cycle structure. Each solid circle represents one "forward" operation of the operator $X = (1 + S^{1/2}N^{-1}S^{1/2})$ at the appropriate resolution.

angular scales, which are slowest to converge but accurately approximated at coarse resolution, while avoiding a large increase in CPU time.

The performance of the multigrid preconditioner (~ 14 CPU-min per Monte Carlo WMAP simulation) is sufficient for purposes of this paper. However, we have also found that none of the preconditioners described so far give reasonable performance with a realistic sky cut and the noise levels and resolution expected for the Planck satellite mission. Therefore, the multigrid preconditioner is probably not the final word on this subject; additional improvements are still needed for future data sets.

3. Template marginalization

So far, we have assumed a noise covariance N_i^{map} for each map which is diagonal in pixel space. Suppose that, in addition, one wants to marginalize the amplitudes of N_{tmpl} modes in the map. We have seen several examples in the paper:

(1) In both WMAP and NVSS, we marginalize the monopole and dipole ($N_{\mathrm{tmpl}} = 4$).
(2) In NVSS, we remove systematic declination gradients by marginalizing any mode which is constant around each isolatitude ring in equatorial coordinates (Sec. V). This leads to $N_{\mathrm{tmpl}} = N_{\mathrm{ring}}$, where N_{ring} is the number of isolatitude rings in the pixelization.
(3) In WMAP, one could use this formalism to marginalize any signal proportional to external foreground templates, although we have not implemented this because the effect of galactic foregrounds is small (Sec. VI C).

Template marginalization, in this general form, is easy to incorporate in our conjugate gradient framework. Let τ be an N_{tmpl}-by-N_{pix} matrix containing the templates. By the Woodbury formula, template marginalization modifies the map covariance as follows:

$$(N_i^{\mathrm{pix}})^{-1} - (N_i^{\mathrm{pix}})^{-1}\tau^T[\tau(N_i^{\mathrm{pix}})^{-1}\tau^T]^{-1}\tau(N_i^{\mathrm{pix}})^{-1}. \quad \text{(A7)}$$

Since the conjugate gradient method only requires a "black

box" procedure for multiplying a map by the inverse covariance $(N_i^{\mathrm{pix}})^{-1}$, one simply includes the extra term in Eq. (A7).

If N_{tmpl} is small (e.g. in the case of marginalizing the monopole and dipole), one can simply keep the matrix τ in dense form. In cases where N_{tmpl} is large, all that is needed is a procedure for multiplying a map by the matrix τ, i.e. computing each template amplitude given a map. For example, when marginalizing declination gradients in NVSS, we implement "multiplication by τ" by simply averaging pixel values around each isolatitude ring in the input map.

APPENDIX B: THREE-POINT ESTIMATORS

In Appendix A we have described in detail how the filtered CMB map $\tilde{a}_{\ell m}$ and filtered galaxy map $\tilde{g}_{\ell m}$ are computed in our pipeline. In order to completely describe our implementation, there is one remaining loose end: in this Appendix, we will give the details of how our quadratic reconstructions $\tilde{\phi}_{\ell m}$, $\tilde{\psi}_{\ell m}$, $\tilde{s}_{\ell m}$ are computed. We will also prove the statement, made throughout the paper, that our bandpower estimators $\hat{C}_b^{\phi g}$, $\hat{C}_b^{\psi g}$, $\hat{C}_b^{sg}$ for lensing, curl null test, and point sources are optimal. Our proof will depend on the assumption of small deviations from Gaussianity, and we discuss the conditions under which this assumption applies.

1. Quadratic reconstruction

Here, we give the implementational details of how the quadratic reconstuctions $\tilde{\phi}_{\ell m}$, $\tilde{\psi}_{\ell m}$, $\tilde{s}_{\ell m}$ are computed in our pipeline. There is a small subtlety because the reconstructions are defined by position-space equations, e.g. $\tilde{\phi}_{\ell m}$ is defined by

$$\sum_{\ell m} \tilde{\phi}_{\ell m} Y_{\ell m}(x) = \nabla^a(\alpha(x)\nabla_a\beta(x)) \quad \text{(B1)}$$

but the maps $\tilde{a}_{\ell m}$, $\tilde{\phi}_{\ell m}$ are defined in harmonic space. [The quantities $\alpha(x)$, $\beta(x)$ were defined in Eqs. (15) and (16).]

In principle, $\tilde{\phi}$ can be evaluated as a brute force harmonic-space sum:

$$\tilde{\phi}_{\ell m}^* = \sum_{\ell_1 m_1 \ell_2 m_2} f_{\ell_1 \ell \ell_2} C_{\ell_2}^{TT} \mathcal{G}_{mm_1 m_2}^{\ell \ell_1 \ell_2} \tilde{a}_{\ell_1 m_1} \tilde{a}_{\ell_2 m_2}, \quad \text{(B2)}$$

where $f_{\ell_1 \ell_2 \ell_3}$ was defined previously in Eq. (7), and we have introduced the notation

$$\mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} \stackrel{\text{def}}{=} \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \quad \text{(B3)}$$

However, the harmonic-space sum has computational cost $\mathcal{O}(\ell_{\max}^5)$ and so we introduce an optimized position-space method.

Multiplying Eq. (B1) on both sides by $Y_{\ell m}(x)^*$ and integrating over x, one obtains

$$\tilde{\phi}_{\ell m} = \int d^2 x \nabla^a (Y_{\ell m}(x))^* \alpha(x) \nabla_a \beta(x). \quad \text{(B4)}$$

The integral can be done exactly using Gauss-Legendre quadrature in $\cos(\theta)$ and uniform quadrature in φ. We evaluate the quantities $\alpha(x)$, $\nabla_a \beta(x)$ on the isolatitude rings by using a fast spin-0 and spin-1 spherical transform, respectively. The right-hand side of Eq. (B4) can then be evaluated using a fast spin-1 transform. This algorithm provides an exact evaluation of Eq. (B4) with computational cost $\mathcal{O}(\ell_{\max}^3)$. We use an analogous method to evaluate the quadratic quantities $\tilde{\psi}_{\ell m}$, $\tilde{s}_{\ell m}$.

2. Equivalence with the bispectrum

As a preliminary step toward proving optimality, we show how the estimators $\hat{C}_b^{\phi g}$, $\hat{C}_b^{\psi g}$, $\hat{C}_b^{sg}$ can be rewritten purely in terms of the associated bispectra. Throughout this paper, when we write a bispectrum $b_{\ell_1 \ell_2 \ell_3}$, it is understood that ℓ_1, ℓ_2 are CMB multipoles and ℓ_3 is a galaxy multipole.

We write the lensing estimator in the following form:

$$\hat{C}^{\phi g} = \frac{1}{\mathcal{N}} \sum_{\ell m} C_\ell^{\phi g} [\tilde{\phi}_{\ell m} - \langle \tilde{\phi}_{\ell m} \rangle]^* \tilde{g}_{\ell m}. \quad \text{(B5)}$$

In Eq. (B5) and throughout this Appendix, $C_\ell^{\phi g}$ denotes the cross power spectrum we are interested in estimating (typically proportional to $1/\ell^2$ over some band in ℓ), not the fiducial spectrum.

If we replace $\tilde{\phi}_{\ell m}^*$ by the right-hand side of Eq. (B2), we obtain

$$\hat{C}_b^{\phi g} = \frac{1}{\mathcal{N}} \sum_{\ell_i m_i} f_{\ell_1 \ell_2 \ell_3} C_{\ell_2}^{TT} C_{\ell_3}^{\phi g} \mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} [\tilde{a}_{\ell_1 m_1} \tilde{a}_{\ell_2 m_2}$$
$$- \langle \tilde{a}_{\ell_1 m_1} \tilde{a}_{\ell_2 m_2} \rangle] \tilde{g}_{\ell_3 m_3}. \quad \text{(B6)}$$

We can replace $\langle \tilde{a}_{\ell_1 m_1} \tilde{a}_{\ell_2 m_2} \rangle$ by $C_{\ell_1 m_1, \ell_2 m_2}^{T-1}$, where in this Appendix we use the notation $(C^T)^{-1}$, $(C^g)^{-1}$ to distinguish the inverse signal + noise covariances for the CMB and galaxy fields. Now comparing with the form of the bispectrum due to lensing [Eq. (9)], this becomes

$$\hat{C}_b^{\phi g} = \frac{1}{2\mathcal{N}} \sum_{\ell_i m_i} b_{\ell_1 \ell_2 \ell_3} \mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} [\tilde{a}_{\ell_1 m_1} \tilde{a}_{\ell_2 m_2}$$
$$- C_{\ell_1 m_1, \ell_2 m_2}^{T-1}] \tilde{g}_{\ell_3 m_3}. \quad \text{(B7)}$$

We have now written the lensing estimator purely in terms of the lensing bispectrum $b_{\ell_1 \ell_2 \ell_3}$. A similar calculation shows that the same is true for the curl and point source estimators $\hat{C}_b^{\psi g}$, $\hat{C}_b^{sg}$: in both cases the estimator takes the form in Eq. (B7), with $b_{\ell_1 \ell_2 \ell_3}$ replaced by the bispectrum due to lensing by a curl component, or the point source bispectrum in Eq. (32). This allows us to give a uniform proof of optimality which applies to all three cases, as we will now see.

3. Optimality

We will now prove the following general statement: for *any* bispectrum $b_{\ell_1 \ell_2 \ell_3}$, the optimal estimator is given by

$$\hat{\mathcal{C}} = \frac{1}{F} (\hat{\mathcal{C}}_3 - \hat{\mathcal{C}}_1), \quad \text{(B8)}$$

where the three-point and one-point terms are defined by

$$\hat{\mathcal{C}}_3 \overset{\text{def}}{=} \frac{1}{2} \sum_{\ell_i m_i} b_{\ell_1 \ell_2 \ell_3} \mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} \tilde{a}_{\ell_1 m_1} \tilde{a}_{\ell_2 m_2} \tilde{g}_{\ell_3 m_3} \quad \text{(B9)}$$

$$\hat{\mathcal{C}}_1 \overset{\text{def}}{=} \frac{1}{2} \sum_{\ell_i m_i} b_{\ell_1 \ell_2 \ell_3} \mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} C_{\ell_1 m_1, \ell_2 m_2}^{T-1} \tilde{g}_{\ell_3 m_3} \quad \text{(B10)}$$

and F is the Fisher matrix element

$$F \overset{\text{def}}{=} \frac{1}{2} \sum_{\ell_i m_i} b_{\ell_1 \ell_2 \ell_3} b_{\ell_4 \ell_5 \ell_6} \mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} \mathcal{G}_{m_4 m_5 m_6}^{\ell_4 \ell_5 \ell_6} C_{\ell_1 m_1, \ell_4 m_4}^{T-1}$$
$$\times C_{\ell_2 m_2, \ell_5 m_5}^{T-1} C_{\ell_3 m_3, \ell_6 m_6}^{g-1}. \quad \text{(B11)}$$

[This expression generalizes the Fisher matrix for all-sky isotropic noise previously considered in Eq. (43) to an arbitrary noise covariance.] Note that we have computed the normalization explicitly; a short calculation shows that the estimator in Eq. (B8) has unit response to the bispectrum $b_{\ell_1 \ell_2 \ell_3}$, so that the estimator is normalized and does not need the $1/\mathcal{N}$ prefactor.

The proof will depend on the assumption of weak non-Gaussianity; specifically we will assume that the fields are sufficiently close to Gaussian that the estimator variance can be approximated by its Gaussian contribution.

First, we can show using the Cramer-Rao inequality that any unbiased estimator $\mathcal{E}$ has variance $\geq 1/F$, where F is the Fisher matrix in Eq. (B11). This is proved using the method of [62,104], expanding the likelihood function for $a_{\ell m}$, $g_{\ell m}$ around its Gaussian limit using the Edgeworth expansion.

Now consider the variance $\text{Var}(\hat{\mathcal{C}})$. We are assuming that this variance may be calculated using Gaussian statistics, so that Wick's theorem gives

$$\text{Var}(\hat{\mathcal{C}}_3, \hat{\mathcal{C}}_3) = F + f^T (C^g)^{-1} f$$
$$\text{Cov}(\hat{\mathcal{C}}_3, \hat{\mathcal{C}}_1) = \text{Cov}(\hat{\mathcal{C}}_1, \hat{\mathcal{C}}_1) = f^T (C^g)^{-1} f, \quad \text{(B12)}$$

where we have defined

$$f_{\ell m} = \frac{1}{2} \sum_{\ell_i m_i} b_{\ell_1 \ell_2 \ell} \mathcal{G}_{m_1 m_2 m}^{\ell_1 \ell_2 \ell} C_{\ell_1 m_1, \ell_2 m_2}^{T-1}. \quad \text{(B13)}$$

Putting this together, we get $\text{Var}(\hat{\mathcal{C}}) = 1/F$, i.e. the Cramer-Rao inequality is saturated. This completes our proof that the estimator is optimal, under the assumption of weak non-Gaussianity.

When is this assumption satisfied for lensing? Roughly speaking, weak non-Gaussianity starts to break down when the instrumental sensitivity becomes good enough that a high signal-to-noise detection of CMB lensing can be achieved. More precisely, consider the case of full-sky

coverage and isotropic noise. This allows us to make contact with the results of [58], where an unbiased estimator $\hat{\phi}_{\ell m}$ is defined for each multipole of the lensing potential, with full-sky noise power spectrum $N_\ell^{\phi\phi}$ given previously in Eq. (4). In this notation, one can show that our filtered field $\tilde{\phi}_{\ell m}$ is equal to $(N_\ell^{\phi\phi})^{-1}\hat{\phi}_{\ell m}$, and our estimator is given by

$$\hat{\mathcal{C}} = \sum_{\ell m} C_\ell^{\phi g} \left(\frac{\hat{\phi}_{\ell m}^*}{N_\ell^{\phi\phi}}\right)\left(\frac{g_{\ell m}}{C_\ell^{gg} + N_\ell^{gg}}\right). \qquad (B14)$$

The first improvement that can be made to this estimator is to make the replacement

$$\hat{\mathcal{C}} \rightarrow \hat{\mathcal{C}}' = \sum_{\ell m} C_\ell^{\phi g} \left(\frac{\hat{\phi}_{\ell m}^*}{C_\ell^{\phi\phi} + N_\ell^{\phi\phi}}\right)\left(\frac{g_{\ell m}}{C_\ell^{gg} + N_\ell^{gg}}\right) \qquad (B15)$$

to incorporate the nonzero sample variance of the lenses. Our estimator $\hat{\mathcal{C}}$ is optimized assuming Gaussian covariance among modes of the CMB, and does not "know" that there is extra sample variance hidden in the problem. However, it is unclear how to generalize $\hat{\mathcal{C}}'$ to the case of sky cuts and inhomogeneous noise, as we have done for $\hat{\mathcal{C}}$, allowing an arbitrary noise covariance matrix N.

The estimators $\hat{\mathcal{C}}, \hat{\mathcal{C}}'$ agree when $C_\ell^{\phi\phi} \ll N_\ell^{\phi\phi}$, i.e. when the reconstruction noise in the lensing potential dominates the signal, considered one mode of the potential at a time. This condition holds for WMAP, as can be seen from the direct comparison in Fig. 2, left panel. However, the estimator $\hat{\mathcal{C}}$ which we have constructed would start to become suboptimal for future surveys with sufficient sensitivity to reconstruct the lensing potential with signal-to-noise ~ 1 per mode. For even more futuristic sensitivity levels, even the improved estimator $\hat{\mathcal{C}}'$ would become suboptimal; the three-point estimator could be improved by using a maximum likelihood formalism which incorporates information from higher-point correlation functions of all orders [59].

In addition to these optimality issues for future surveys, there are other ways in which our estimator might be extended. First, we have not considered CMB polarization, which is ultimately expected to provide more sensitivity to lensing than temperature [105]. Second, by using full-blown C^{-1} filtering, we have ensured optimality of the estimator, but it would be interesting to determine whether a simpler filter can be found which achieves near-optimal power spectrum uncertainties. As we have remarked in Appendix A, the C^{-1} operation seems prohibitively expensive for Planck with existing preconditioners, so finding such a filter may be a practical necessity for future experiments.

APPENDIX C: RESOLVED POINT SOURCES

We give a proof of a statement made in Sec. VII C: correlations between the mask and the galaxy field cannot fake the lensing signal, i.e. the expectation value

$$\langle \hat{C}_b^{\phi g} \rangle_{T,G,M} = 0 \qquad (C1)$$

in the absence of CMB lensing. We have introduced the notation $\langle \cdot \rangle_{T,G,M}$ to denote an expectation value taken over realizations of the CMB T, galaxy counts G, and mask M (where the last two are assumed correlated).

In the proof, we will denote the quadratic reconstruction $\tilde{\phi}$ which appears in the lensing estimator by $\tilde{\phi}(T, M)$ to emphasize that it depends on both the CMB realization T and the mask M. We will analogously denote the filtered galaxy field by $\tilde{g}(G, M)$. In this notation, the lensing estimator can be written

$$\hat{C}_b^{\phi g} = \sum_{\ell m} [\tilde{\phi}(T, M) - \langle\tilde{\phi}(T', M)\rangle_{T'}]_{\ell m}^* \tilde{g}(G, M)_{\ell m}, \quad (C2)$$

where we have written the one-point term as an average over CMB realization T' with the mask M fixed. Taking the expectation value $\langle \cdot \rangle_{T,G,M}$ on both sides, we obtain

$$\langle \hat{C}_b^{\phi g} \rangle_{T,G,M} = \left\langle \sum_{\ell m} \langle \tilde{\phi}(T, M)_{\ell m}^* \rangle_T \langle \tilde{g}(G, M)_{\ell m}\rangle_G \right.$$
$$\left. - \langle \tilde{\phi}(T', M)_{\ell m}^* \rangle_{T'} \langle \tilde{g}(G, M)_{\ell m}\rangle_G \right\rangle_M = 0. \quad (C3)$$

In the first line, we have used the fact that, in the absence of lensing, the CMB realization T is independent of the galaxy realization G once the mask M has been specified, to bring the expectation value $\langle \cdot \rangle_T$ inside the sum. This completes the proof that the expectation value in Eq. (C1) vanishes in the absence of CMB lensing, i.e. mask-galaxy correlations cannot fake the lensing signal.

It is interesting to note that this proof would break down if the one-point term were omitted from the lensing estimator $\hat{C}_b^{\phi g}$. In this case, we would obtain

$$\langle \hat{C}_b^{\phi g} \rangle_{T,g,M} = \left\langle \sum_{\ell m} \langle \tilde{\phi}(T, M)_{\ell m}^* \rangle_T \langle \tilde{g}(G, M)_{\ell m}\rangle_G \right\rangle_M \quad (C4)$$

which cannot be simplified further: the map $\langle \tilde{g}(G, M)\rangle_G$ can be nonzero if there are mask-galaxy correlations, and the map $\langle \tilde{\phi}(T, M)\rangle_T$ is generally nonzero in the presence of a mask.

APPENDIX D: BEAM ASYMMETRY

To include beam asymmetry in our simulation pipeline, we need an expression for the beam-convolved temperature $\tilde{T}(x)$ in each pixel x, in terms of three quantities: the beam profile, the scan strategy, and the unconvolved CMB $T(x)$.

We represent the beam profile in real space as $G(\theta, \varphi)$ or in harmonic space as

$$G(\theta, \varphi) = \sum_{\ell s} g_{\ell s} Y_{\ell s}(\theta, \varphi). \qquad (D1)$$

Following [46], Appendix B, the scan strategy will be represented by the following quantity:

$$w(x, \alpha) = 2\pi \frac{\sum_{i \in x} \delta(\alpha - \alpha_i)}{\sum_{i \in x} 1}, \qquad \text{(D2)}$$

where the angle α parametrizes beam orientations at the pixel x, relative to an arbitrarily chosen reference angle. The sum in Eq. (D2) runs over timestream samples i which fall in pixel x with beam orientation α_i. Note that $w(x, \alpha)$ depends on the choice of reference direction, or local frame at x.

We briefly recall the theory of spin-s fields; for more details see [106]. A spin-s field ($-\infty < s < \infty$) is a function $(_s f)$ whose value at x depends on a choice of frame, or pair of orthonormal basis vectors $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ at x. Under the right-handed rotation

$$\begin{aligned} \hat{\mathbf{e}}_1' &= (\cos\theta)\hat{\mathbf{e}}_1 + (\sin\theta)\hat{\mathbf{e}}_2 \\ \hat{\mathbf{e}}_2' &= -(\sin\theta)\hat{\mathbf{e}}_1 + (\cos\theta)\hat{\mathbf{e}}_2, \end{aligned} \qquad \text{(D3)}$$

$(_s f)$ must transform as $(_s f)' = e^{-is\theta}(_s f)$. One can define spin-$s$ spherical harmonics $_s Y_{\ell m}(\theta, \varphi)$; these are an orthonormal set of basis functions for spin-s fields, with properties that are similar to the ordinary (spin-0) spherical harmonics $Y_{\ell m}$.

If we Fourier transform the frame-dependent quantity $w(x, \alpha)$ in the angle α,

$$w(x, \alpha) = \sum_{s=-\infty}^{\infty} (_s w(x))^* e^{is\alpha}, \qquad \text{(D4)}$$

then $_s w(x)$ will be a spin-s field as suggested by the notation.

Now we can write an expression for the beam-convolved CMB temperature $\tilde{T}(x)$:

$$\tilde{T}(x) = \int d^2 x' T(x') \int \frac{d\alpha}{2\pi} w(x, \alpha)_{2P} G(\theta_{xx'}, -\alpha), \quad \text{(D5)}$$

where $\theta_{xx'}$ denotes the angle between points x, x', and the subscript "2P" on any frame-dependent quantity [such as $w(x, \alpha)$] indicates the "two-point" frame: the reference direction $\hat{\mathbf{e}}_1$ at x points toward x'.

Equation (D5) simply states that the beam-convolved temperature at x is given by averaging over scan directions α, with the beam profile rotated through angle α before it is applied. To simplify this expression, we plug in Eqs. (D1) and (D4), obtaining

$$\tilde{T}(x) = \int d^2 x T(x) \sum_{s\ell} (_s w(x)_{2P})^* g_{\ell s} Y_{\ell s}(\theta_{xx'}, 0). \quad \text{(D6)}$$

Now use the identity

$$Y_{\ell s}(\theta_{xx'}, 0) = \sqrt{\frac{4\pi}{2\ell + 1}} \sum_m (_s Y_{\ell m}(x))_{2P} Y_{\ell m}^*(x') \qquad \text{(D7)}$$

to obtain

$$\tilde{T}(x) = \sum_{s\ell m} \sqrt{\frac{4\pi}{2\ell + 1}} (_s w(x))^* g_{\ell s} a_{\ell m} (_s Y_{\ell m}(x)). \qquad \text{(D8)}$$

This is our desired expression for $\tilde{T}$. The final result is a spin-0 quantity, so we have dropped the 2P.

Equation (D8) is a sum over beam multipoles s multiplied by the spin-s component of the scan strategy. Note that the spin-0 component $[_0 w(x)]$ is equal to 1 by construction [Eq. (D2)], so that the $s = 0$ term in Eq. (D8) does not depend on the scan strategy and is simply given by convolving $\{a_{\ell m}\}$ with the beam transfer function $b_\ell = \sqrt{4\pi/(2\ell + 1)} g_{\ell 0}$. The higher-spin terms do depend on the scan strategy and represent corrections to the symmetric-beam approximation. If the beam is azimuthally symmetric ($g_{\ell s} = 0$ for $s > 0$), or the beam is arbitrary but the scan is isotropic in each pixel [$_s w(x) = 0$ for $s > 0$], then the higher-spin terms do not contribute and the symmetric-beam approximation is exact. For WMAP, we find that the sum over s in Eq. (D8) converges rapidly so that truncating at $s_{\text{smax}} = 16$ fully incorporates beam asymmetry.

A fast algorithm for evaluating Eq. (D8) may be given by noting that each term in the s sum is simply a spin-s spherical transform. In an isolatitude coordinate system, a spin-s transform may be performed with computational cost $\mathcal{O}(\ell_{\max}^3)$ by using the recursion

$$\rho_{\ell m}^s (_s Y_{\ell m}) = \left(z + \frac{sm}{\ell(\ell - 1)} \right) _s Y_{\ell-1,m} - \rho_{\ell-1,m}^s (_s Y_{\ell-2,m}) \qquad \text{(D9)}$$

on each isolatitude ring, where we have defined $\rho_{\ell m}^s = \sqrt{(\ell^2 - m^2)(\ell^2 - s^2)/(4\ell^2 - 1)}/\ell$. Thus the total computational cost of incorporating beam asymmetry via Eq. (D8) is $\mathcal{O}(s_{\text{smax}} \ell_{\max}^3)$.

Finally, we include a detail which is specific to WMAP. The preceding treatment has assumed that there is one beam $g_{\ell s}$ and one scan $_s w(x)$ for each simulated map. In WMAP, we have one simulated map per differencing assembly, obtained as the difference of A-side and B-side measurements. In this case, one makes the replacement

$$(_s w(x))^* g_{\ell s} \rightarrow (_s w^A(x))^* g_{\ell s}^A + (_s w^B(x))^* g_{\ell s}^B \qquad \text{(D10)}$$

in Eq. (D8), where $g_{\ell m}^A$, $g_{\ell m}^B$ are the A-side and B-side beams, and w^A, w^B are defined by

$$w^A(x, \alpha) \stackrel{\text{def}}{=} 2\pi \frac{\sum_{a \in x} \delta(\alpha - \alpha_a)}{(\sum_{a \in x} 1) + (\sum_{b \in x} 1)} \qquad \text{(D11)}$$

$$w^B(x, \alpha) \stackrel{\text{def}}{=} 2\pi \frac{\sum_{b \in x} \delta(\alpha - \alpha_b)}{(\sum_{a \in x} 1) + (\sum_{b \in x} 1)}, \qquad \text{(D12)}$$

where $\sum_{a \in x}$, $\sum_{b \in x}$ denote sums over A-side and B-side timestream samples which fall in pixel x.

[1] C. L. Bennett *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **148**, 1 (2003).

[2] D. N. Spergel *et al.* (WMAP Collaboration), arXiv:astro-ph/0603449.

[3] W. C. Jones *et al.*, Astrophys. J. **647**, 823 (2006).

[4] A. C. S. Readhead *et al.*, Astrophys. J. **609**, 498 (2004).

[5] C.-L. Kuo *et al.*, arXiv:astro-ph/0611198.

[6] A. G. Riess *et al.* (Supernova Search Team), Astron. J. **116**, 1009 (1998).

[7] S. Perlmutter *et al.* (Supernova Cosmology Project), Astrophys. J. **517**, 565 (1999).

[8] D. G. York *et al.* (SDSS Collaboration), Astron. J. **120**, 1579 (2000).

[9] M. Colless *et al.* (2DFGRS Collaboration), Mon. Not. R. Astron. Soc. **328**, 1039 (2001).

[10] U. Seljak, A. Slosar, and P. McDonald, J. Cosmol. Astropart. Phys. 10 (2006) 014.

[11] H. Hoekstra, H. K. C. Yee, and M. D. Gladders, Astrophys. J. **577**, 595 (2002).

[12] P. Astier *et al.* (SNLS Collaboration), Astron. Astrophys. **447**, 31 (2006).

[13] A. Refregier, Annu. Rev. Astron. Astrophys. **41**, 645 (2003).

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

[15] J. Kovac *et al.*, Nature (London) **420**, 772 (2002).

[16] F. Piacentini *et al.*, Astrophys. J. **647**, 833 (2006).

[17] A. C. S. Readhead *et al.*, arXiv:astro-ph/0409569.

[18] D. Barkats *et al.*, Astrophys. J. **619**, L127 (2005).

[19] T. E. Montroy *et al.*, Astrophys. J. **647**, 813 (2006).

[20] B. R. Johnson *et al.*, arXiv:astro-ph/0611394.

[21] A. Lewis and A. Challinor, Phys. Rep. **429**, 1 (2006).

[22] R. A. Sunyaev and Y. B. Zel'dovich, Astrophys. Space Sci. **7**, 20 (1970).

[23] R. A. Sunyaev and Y. B. Zel'dovich, Annu. Rev. Astron. Astrophys. **18**, 537 (1980).

[24] R. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).

[25] A. Kosowsky, New Astron. Rev. **47**, 939 (2003).

[26] J. E. Ruhl *et al.* (SPT Collaboration), Proc. SPIE **5498**, 11 (2004).

[27] M. Birkinshaw, S. F. Gull, and H. Hardebeck, Nature (London) **309**, 34 (1984).

[28] W. L. Holzapfel *et al.*, arXiv:astro-ph/9702224.

[29] S. J. LaRoque *et al.*, arXiv:astro-ph/0204134.

[30] K. S. Dawson, W. L. Holzapfel, J. E. Carlstrom, M. Joy, and S. J. LaRoque, Astrophys. J. **647**, 13 (2006).

[31] N. Afshordi, Y.-T. Lin, D. Nagai, and A. J. R. Sanderson, arXiv:astro-ph/0612700.

[32] P. Fosalba, E. Gaztanaga, and F. Castander, Astrophys. J. **597**, L89 (2003).

[33] N. Afshordi, Y.-S. Loh, and M. A. Strauss, Phys. Rev. D **69**, 083524 (2004).

[34] C. Hernandez-Monteagudo, R. Genova-Santos, and F. Atrio-Barandela, Astrophys. J. **613**, L89 (2004).

[35] S. Boughn and R. Crittenden, Nature (London) **427**, 45 (2004).

[36] M. R. Nolta *et al.* (WMAP Collaboration), Astrophys. J. **608**, 10 (2004).

[37] S. P. Boughn and R. G. Crittenden, New Astron. Rev. **49**, 75 (2005).

[38] S. J. Maddox, G. Efstathiou, W. J. Sutherland, and J.

[39] R. Scranton *et al.* (SDSS Collaboration), arXiv:astro-ph/0307335.

[40] P. Fosalba and E. Gaztanaga, Mon. Not. R. Astron. Soc. **350**, L37 (2004).

[41] N. Padmanabhan *et al.*, Phys. Rev. D **72**, 043525 (2005).

[42] A. Cabre, E. Gaztanaga, M. Manera, P. Fosalba, and F. Castander, Mon. Not. R. Astron. Soc. Lett. **372**, L23 (2006).

[43] J. D. McEwen, P. Vielva, M. P. Hobson, E. Martinez-Gonzalez, and A. N. Lasenby, arXiv:astro-ph/0605122.

[44] C. M. Hirata, N. Padmanabhan, U. Seljak, D. Schlegel, and J. Brinkmann, Phys. Rev. D **70**, 103501 (2004).

[45] J. J. Condon *et al.*, Astron. J. **115**, 1693 (1998).

[46] G. Hinshaw *et al.*, arXiv:astro-ph/0603451.

[47] http://lambda.gsfc.nasa.gov/.

[48] N. Jarosik *et al.* (WMAP Collaboration), arXiv:astro-ph/0603452.

[49] N. Jarosik *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **148**, 29 (2003).

[50] L. Page *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **148**, 39 (2003).

[51] K. M. Gorski *et al.*, Astrophys. J. **622**, 759 (2005).

[52] For more information on HEALPIX visit, see http://www.eso.org/science/healpix/.

[53] A. Blanchard and J. Schneider, Astron. Astrophys. **184**, 1 (1987).

[54] U. Seljak, Astrophys. J. **463**, 1 (1996).

[55] F. Bernardeau, Astron. Astrophys. **324**, 15 (1997).

[56] W. Hu, Phys. Rev. D **64**, 083005 (2001).

[57] W. Hu, Astrophys. J. **557**, L79 (2001).

[58] T. Okamoto and W. Hu, Phys. Rev. D **67**, 083002 (2003).

[59] C. M. Hirata and U. Seljak, Phys. Rev. D **67**, 043001 (2003).

[60] A. Kogut *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **148**, 161 (2003).

[61] E. Komatsu, D. N. Spergel, and B. D. Wandelt, Astrophys. J. **634**, 14 (2005).

[62] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 05 (2006) 004.

[63] K. M. Smith and M. Zaldarriaga, arXiv:astro-ph/0612571.

[64] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001).

[65] M. Bartelmann and P. Schneider, Phys. Rep. **340**, 291 (2001).

[66] J. R. Bond, A. H. Jaffe, and L. Knox, Phys. Rev. D **57**, 2117 (1998).

[67] S. P. Oh, D. N. Spergel, and G. Hinshaw, Astrophys. J. **510**, 551 (1999).

[68] J. Jewell, S. Levin, and C. H. Anderson, Astrophys. J. **609**, 1 (2004).

[69] B. D. Wandelt, D. L. Larson, and A. Lakshminarayanan, Phys. Rev. D **70**, 083511 (2004).

[70] A. Cooray, M. Kamionkowski, and R. R. Caldwell, Phys. Rev. D **71**, 123527 (2005).

[71] C. Blake and J. Wall, Mon. Not. R. Astron. Soc. **329**, L37 (2002).

[72] C. Blake, P. G. Ferreira, and J. Borrill, Mon. Not. R. Astron. Soc. **351**, 923 (2004).

[73] J. S. Dunlop and J. A. Peacock, Mon. Not. R. Astron. Soc. **247**, 19 (1990).

Loveday, Mon. Not. R. Astron. Soc. **242**, 43 (1990).

[74] S. P. Boughn and R. G. Crittenden, Phys. Rev. Lett. **88**, 021302 (2002).

[75] M. J. Rees and D. W. Sciama, Nature (London) **217**, 511 (1968).

[76] C. Bennett *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **148**, 97 (2003).

[77] D. P. Finkbeiner, M. Davis, and D. J. Schlegel, Astrophys. J. **524**, 867 (1999).

[78] D. P. Finkbeiner, arXiv:astro-ph/0301558.

[79] C. G. T. Haslam *et al.*, Astron. Astrophys. **100**, 209 (1981).

[80] L. Toffolatti *et al.*, Mon. Not. R. Astron. Soc. **297**, 117 (1998).

[81] Y.-T. Lin and J. J. Mohr, Astrophys. J. Suppl. Ser. **170**, 71 (2007).

[82] K. Coble *et al.*, arXiv:astro-ph/0608274.

[83] P. Giommi *et al.*, arXiv:astro-ph/0703150.

[84] A. Cooray and R. K. Sheth, Phys. Rep. **372**, 1 (2002).

[85] A. Amblard, C. Vale, and M. J. White, New Astron. Rev. **9**, 687 (2004).

[86] K. M. Huffenberger and U. Seljak, New Astron. Rev. **10**, 491 (2005).

[87] V. Springel, M. J. White, and L. Hernquist, arXiv:astro-ph/0008133.

[88] M. J. White, L. Hernquist, and V. Springel, Astrophys. J. **579**, 16 (2002).

[89] U. Seljak, J. Burwell, and U.-L. Pen, Phys. Rev. D **63**, 063001 (2001).

[90] R. Scoccimarro and R. K. Sheth, Mon. Not. R. Astron.

[91] P. Monaco, T. Theuns, and G. Taffoni, Mon. Not. R. Astron. Soc. **331**, 587 (2002).

[92] E. Komatsu and U. Seljak, Mon. Not. R. Astron. Soc. **336**, 1256 (2002).

[93] E. Komatsu and U. Seljak, Mon. Not. R. Astron. Soc. **327**, 1353 (2001).

[94] E. Komatsu and T. Kitayama, Astrophys. J. **526**, L1 (1999).

[95] A. F. Heavens, arXiv:astro-ph/9804222.

[96] A. Lewis and S. Bridle, Phys. Rev. D **66**, 103511 (2002).

[97] R. Narayan, Astrophys. J. Lett. **339**, 53 (1989).

[98] T. J. Broadhurst, A. N. Taylor, and J. A. Peacock, Astrophys. J. **438**, 49 (1995).

[99] M. LoVerde, L. Hui, and E. Gaztanaga, Phys. Rev. D **75**, 043519 (2007).

[100] http://www.rssd.esa.int/index.php?project=Planck.

[101] M. Tegmark, Astrophys. J. **480**, L87 (1997).

[102] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1992), 2nd ed., ISBN 0-521-43064-X.

[103] H. K. Eriksen *et al.*, Astrophys. J. Suppl. Ser. **155**, 227 (2004).

[104] D. Babich, Phys. Rev. D **72**, 043003 (2005).

[105] W. Hu and T. Okamoto, Astrophys. J. **574**, 566 (2002).

[106] M. Zaldarriaga and U. Seljak, Phys. Rev. D **55**, 1830 (1997).

Soc. **329**, 629 (2002).