

# Making maps of the cosmic microwave background: The MAXIMA example

Radek Stompor,<sup>1,2,3</sup> Amedeo Balbi,<sup>4</sup> Julian D. Borrill,<sup>5,1</sup> Pedro G. Ferreira,<sup>6</sup> Shaul Hanany,<sup>7,1</sup> Andrew H. Jaffe,<sup>1,8,9</sup> Adrian T. Lee,<sup>9,1,10</sup> Sang Oh,<sup>1,9</sup> Bahman Rabii,<sup>1,9</sup> Paul L. Richards,<sup>1,9</sup> George F. Smoot,<sup>1,10,2</sup> Celeste D. Winant,<sup>1,9</sup> and Jiun-Huei Proty Wu<sup>8</sup>

<sup>1</sup>*Center for Particle Astrophysics, University of California, Berkeley, California 94720*

<sup>2</sup>*Space Sciences Laboratory, University of California, Berkeley, California 94720*

<sup>3</sup>*Copernicus Astronomical Center, Warszawa, Poland*

<sup>4</sup>*Dipartimento di Fisica, Università Tor Vergata, Roma, Italy*

<sup>5</sup>*NERSC, Lawrence Berkeley National Laboratory, Berkeley, California 94720*

<sup>6</sup>*Astrophysics, University of Oxford, NAPL, Oxford, OX1 3RH, United Kingdom*

<sup>7</sup>*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455*

<sup>8</sup>*Department of Astronomy, University of California, Berkeley, California 94720*

<sup>9</sup>*Department of Physics, University of California, Berkeley, California 94720*

<sup>10</sup>*Lawrence Berkeley National Laboratory, Berkeley, California 94720*

(Received 26 June 2001; published 26 December 2001)

This work describes cosmic microwave background (CMB) data analysis algorithms and their implementations, developed to produce a pixelized map of the sky and a corresponding pixel-pixel noise correlation matrix from time ordered data for a CMB mapping experiment. We discuss in turn algorithms for estimating noise properties from the time ordered data, techniques for manipulating the time ordered data, and a number of variants of the maximum likelihood map-making procedure. We pay particular attention to issues pertinent to *real* CMB data, and present ways of incorporating them within the framework of maximum likelihood map making. Making a map of the sky is shown to be not only an intermediate step rendering an image of the sky, but also an important diagnostic stage, when tests for and/or removal of systematic effects can efficiently be performed. The case under study is the MAXIMA-I data set. However, the methods discussed are expected to be applicable to the analysis of other current and forthcoming CMB experiments.

DOI: 10.1103/PhysRevD.65.022003

PACS number(s): 98.70.Vc, 98.80.Bp, 98.80.Es

## I. INTRODUCTION

This paper presents a comprehensive set of data analysis methods aiming at the production of a map of the sky and an accurate estimate of map uncertainty in a case of cosmic microwave background (CMB) mapping experiments. We describe a variety of maximum-likelihood-based map-making methods, and discuss their performance in the analysis of the MAXIMA-I data set.

MAXIMA is a balloon-borne experiment [1] built primarily in Berkeley [2] and designed to make a number of short-duration flights. To date the MAXIMA team has published the results of the first flight of the instrument [3,4], consisting of a high-resolution map of almost 100 square degrees of the microwave sky, together with a power spectrum of the CMB anisotropies observed in the map covering a broad range in  $l$  space from  $l \sim 35$  up to  $\sim 1235$ , corresponding to angular scales from  $5^\circ$  down to 5 arcmin. Such products are final results of an involved data analysis pipeline described in this paper. The complexity and size of this data set have proven to be a significant challenge for data analysis methods, setting demanding requirements for both their precision and speed, a challenge which our methods and tools are designed to meet. With other complex and advanced CMB experiments in progress and anticipated [including the satellite missions, the Microwave Anisotropy Probe (MAP) [5] and Planck [6]], these tools and methods can be expected to be of wider interest and applicability. Describing the details of the MAXIMA-I data analysis is another goal of this paper.

The structure of this work is as follows: in Sec. II we deal with the data in the time domain, focusing on data pre-processing and noise estimation, including an outline of the basic features of the MAXIMA-I data set, and of the simulation tools used to test our map-making pipeline. Section III is devoted to the description and comparison of a suite of different map-making methods. Those simultaneously produce both a map and a corresponding pixel-pixel noise correlation matrix. Although the algorithms are all based on the maximum likelihood approach, they differ in the way they attempt to optimize the balance between accuracy and speed. We demonstrate their performance in analyzing MAXIMA-like simulations as well as the actual MAXIMA-I data set. In Sec. IV we discuss ways of handling systematic effects within the general framework of maximum likelihood map making. Although such systematics are inevitable in real CMB data sets, they are rarely considered in more theoretical accounts of CMB data analysis (e.g., [7–15]). In Sec. V we combine these elements, and consider some practical aspects of recently-proposed iterative algorithms for time-domain noise estimation [11,16]. In Sec. VI, we complete our presentation with a description of the numerical tests we have developed to check consistency of our analysis.

The inter-dependencies of different sections of this paper are depicted in Fig. 1.

In this paper we do not consider issues related to the subsequent statistical investigation of these maps, such as tests for Gaussianity or power spectrum estimation. Our map-making methods are intended to be as general as pos-

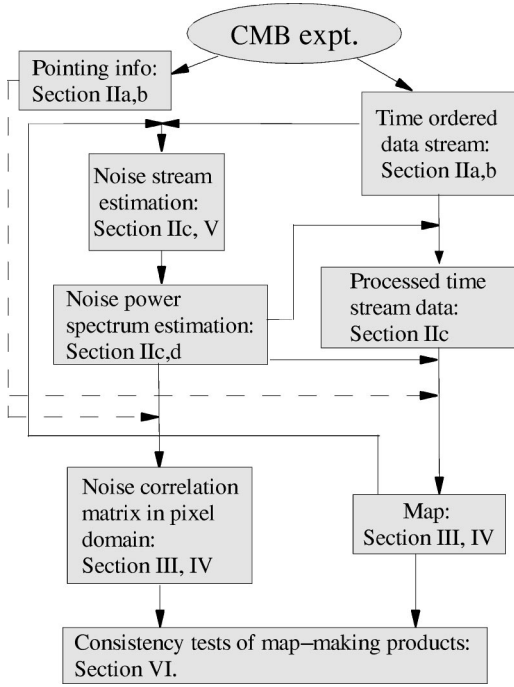


FIG. 1. A flow chart showing the layout and inter-dependencies of the sections of this paper, as well as mutual relations of different stages of the data analysis pipeline as described in the text.

sible, and because they provide both a map and a pixel-domain noise correlation matrix, they do not restrict the subsequent choice of statistical tool. Methods for obtaining an angular power spectrum from, or for searching for non-Gaussianity in, such maps have been described in a number of recent papers (e.g., [17–23] and [24–27], respectively). The power spectra shown in this paper have been computed using a generic uncustomized version of a quadratic estimator as implemented in the publicly-available MADCAP package [28,29].

A comprehensive discussion of the maps and power spectra produced by the MAXIMA-I experiment can be found elsewhere [3,4,30]; their cosmological implications are discussed in [26,31,32].

Hereafter, we denote vectors and scalars with bold and non-bold lower case letters (either Roman or Greek) respectively, while matrices and operators are denoted with either bold or caligraphic upper case letters. Vector and matrix components are indexed in parentheses, rather than by subscript; subscripts and superscripts are used to distinguish between different variants of a given quantity, e.g.,  $p$  and  $t$  denote a pixel and a time domain quantity respectively. A tilde over a quantity denotes its Fourier transform, i.e.,

$$\tilde{g}(f) \equiv \int dt g(t) \exp(2\pi i f t).$$

Occasional failures of these good intentions are also acknowledged. A summary of the most frequently used symbols is given in Table I.

TABLE I. A summary of the notation used in this paper.

Symbol	Description	Section
$\Delta$	data sampling interval	II A
$d(t)$	time stream data	II B
$d_F(t)$	time stream data convolved with filter $F$	II B
$F$	instrumental filters (all together)	II B
$F_{low}$	ac low-pass electronic filter	II B
$F_{high}$	ac high-pass electronic filter	II B
$F_{bolo}$	bolometer low-pass filter	II B
$t_{sky}$	sky signal	II B
$x(\alpha)$	$\alpha$ -synchronous signal	II B
$n_t$	time stream noise	II B
$P(f)$	time domain noise power spectrum	II B
$P_W(f)$	prewhitened time domain noise spectrum	II C
$W$	prewhitening filter	II C
$\lambda_c$	noise correlation length in time domain	II C
$S(f)$	spectrum smoothing window function	II D
$N_t$	time domain noise correlation matrix	II D
$N_{Ct}$	circulant part of $N_t$	II D
$N_{St}$	sparse part of $N_t$	III C
$A$	pointing matrix of the experiment	III A
$m_p$	pixelized sky $\equiv$ map	III A
$N_p$	noise correlation matrix in pixel domain	III A
$N_{Cp}$	circulant part of $N_p$	III C
$N_{Sp}$	sparse part of $N_p$	III C
$n_{pix}$	number of pixels in a map $m_p$	III A
$n_s$	length of the time stream segment	III A
$B, \mathcal{B}$	pointing matrices of synchronous effects	IV A
$x_q$	extra fictitious pixels	IV A
$\mathcal{A}$	generalized pointing matrix	IV A
$\hat{m}_p$	generalized map	IV A
$\mathcal{N}$	generalized noise correlation matrix	IV A
$\tau_t$	time domain template	IV B
$\delta^K$	Kronecker delta	IV B
$u_p$	vector of ones	IV C
$v_p$	singular pixel domain eigenvector	IV C, IV D
$C_l$	CMB anisotropy power spectrum	IV C, IV F
$M_p$	total signal+noise correlation matrix	IV C, IV F
$S_p(C_l)$	CMB signal correlation matrix	IV C, IV F
$N_p^{-1/2}$	Cholesky factor of matrix $N_p$	III D, VI
$w_p$	decorrelated map	VI

## II. TIME ORDERED DATA

### A. The MAXIMA-I data set

The MAXIMA-I data and instrument have been described in [1,3,4]. The MAXIMA-I data set consists of approximately 2300000 measurements for each of 16 photometers and 4 dark channels used to monitor the experiment. To date only data from six of the detectors have been analyzed. These include four photometers sensitive to CMB photons (3 with frequency bandwidths centered on 150 GHz and 1 on 240 GHz), one photometer monitoring atmosphere and foregrounds at 410 GHz, and one “dark” bolometer (screened from incident photons) used to search for systematic prob-

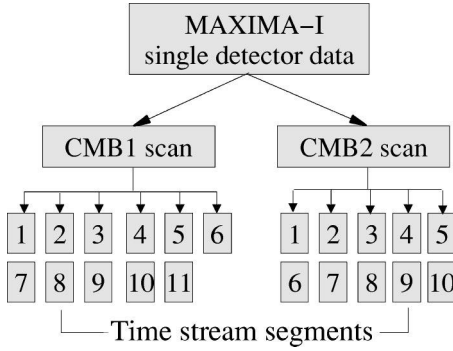


FIG. 2. A structure of the MAXIMA-I data set.

lems. Each of the data streams for each of these detectors is divided into two parts, hereafter called the CMB1 and CMB2 scans respectively (see Fig. 2). These scans were taken at different elevations and were separated in time by  $\sim 1.5$  h. Projected on the sky, they largely overlap one another, creating a well-crosslinked map. Each of the two scans is further subdivided into 11 (CMB1 scan) and 10 (CMB2 scan) data subsets whose lengths range from 30000 to 250000 time samples. These disjoint stretches of contiguous data—hereafter referred to as time stream segments—are defined by the requirement that the noise within a segment be approximately stationary. The noise correlations between the segments are guaranteed to be negligible by discarding a sufficiently long section of the time stream data between neighboring segments. In addition, each of the segments has an overall offset and gradient subtracted independently.

Not all measurements within a segment are to be included in the further analysis. Measurements compromised by glitches—cosmic rays hits, telemetry drop-outs, or other short transients—are flagged as “bad” and constitute gaps in a segment. We usually determine about 2–3 % of the time samples as gaps.

The signal and noise are subjected to several filters before being recorded, including a low-pass filter due to the detector time constant and subsequent ac low- and high-pass filters in the readout electronics. These filters are phase-shifting (i.e., their Fourier space representation is by complex numbers) and they define the temporal frequency response of the instrument to a band between  $f \gtrsim 0.01$  Hz and  $f \lesssim 20$  Hz. This frequency response together with the scan strategy give MAXIMA-I sensitivity to the sky features in the range of angular scales from  $\sim 5^\circ$  down to  $\sim 5$  arcmin. The instrumental filters must be deconvolved from the time ordered data in the course of the data analysis. The functional form of the electronic filters is accurately measured in the laboratory. The detector time constant is measured using the in-flight response of the detectors to a known source—in the case of MAXIMA-I, Jupiter. For our purposes here it is assumed that the instrumental filters are known to a negligible uncertainty in the range of frequencies of interest, i.e.,  $\lesssim 50$  Hz. (Note that the bolometer time constant has an uncertainty of  $\sim 0.5$ – $1.0$  ms. This error is included in the analysis of the MAXIMA-I data [3,4,30], but we neglect it for the purpose of this paper.) The data sampling interval is fixed throughout the entire observation at  $\Delta = 0.0048$  s. Hereafter, we identify

an observation in the time domain (including gaps) by a global sample-number index.

The MAXIMA scanning strategy includes an azimuthal primary mirror chop with a frequency  $\sim 0.45$  Hz superimposed on a slow azimuthal balloon gondola oscillation with a characteristic frequency of  $\sim 0.01$  Hz. Both the position of the primary mirror with respect to the gondola and the azimuthal position of the gondola were recorded during flight, enabling the tracking of primary mirror and/or gondola synchronous systematic effects.

The MAXIMA pointing is reconstructed using the observations of guide stars made throughout the flight. With an rms accuracy better than  $\sim 1'$ , pointing uncertainty is neglected during the map-making stage, although it can, and indeed should, be included as a systematic uncertainty of the final power spectrum (e.g., [4]).

The time stream data are assigned to pixels prior to map-making. The relatively small angular size of the sky patch observed by MAXIMA-I allows us to use simple square pixels on a grid whose rows are at constant declination. Because of computational constraints, most of the tests discussed in this paper have been performed using 8 arcmin pixels, although the highest resolution MAXIMA-I results have been computed using 3 arcmin pixels [4].

The independence of a pixel’s size from its sky position enables us to account for the extra smoothing due to the pixelization by using a simple, albeit approximate, pixel window function [33]. This approximation usually breaks down at the smallest angular scales, due to the lack of uniformity of the sky sampling on these scales. Ways of alleviating this problem are discussed elsewhere [4]. The methods discussed below are independent of the assumed pixelization scheme.

## B. Time stream modeling and simulations

Drawing from the MAXIMA-I experience, we now explicitly list all the features of the time stream data which we have found to be important in the data analysis. We also briefly explain how we incorporate these features into our simulations, which are then used for tests of our data analysis tools.

### 1. Time stream model

We denote the entire raw time stream from a single detector, including the effect of the instrumental filters, as  $\mathbf{d}_F$ . As noted above, this is subdivided into a number ( $n_{seg}$ ) of disjoint segments, so that we can write  $\mathbf{d}_F \equiv \bigcup_{i=1}^{n_{seg}} \mathbf{d}_F^i$ .

Every measurement contains contributions from both the sky and the instrument, and is modeled as,

$$\mathbf{d}_F^i(i) = \sum_j \mathbf{F}(i,j)[t_{sky}(\gamma(j)) + \mathbf{x}(\alpha(j))] + \mathbf{n}_i(i). \quad (1)$$

$\mathbf{F}$  denotes the effect of all of the instrumental filters, and for MAXIMA is therefore a convolution of three filters— $\mathbf{F} \equiv \mathbf{F}_{high} \star \mathbf{F}_{low} \star \mathbf{F}_{bolo}$ —corresponding to the ac high-pass, ac low-pass, and bolometer low-pass filters.  $t_{sky}(\gamma(i))$  is the temperature of the sky in the direction  $\gamma(i)$  observed at time  $i$ .  $\mathbf{x}$  is any extra systematic “ $\alpha$ -synchronous” effect (e.g., primary mirror chop synchronous noise) which only depends

on a known parameter  $\alpha$  (e.g., the mirror position). The dependence of  $\mathbf{x}(\alpha)$  on time is, therefore, exclusively a result of the time dependence of the parameter  $\alpha$ .  $\mathbf{n}_i(i)$  denotes the total (Gaussian, correlated) instrument noise in observation  $i$ . In fact, independent noise components are introduced into the time stream at four different stages in the instrument and, more precisely, the total noise is represented as

$$\mathbf{n}_i = \mathbf{F}_{high}[\mathbf{F}_{low}[\mathbf{F}_{bolo}\mathbf{n}_{i1} + \mathbf{n}_{i2}] + \mathbf{n}_{i3}] + \mathbf{n}_{i4}. \quad (2)$$

Here  $\mathbf{n}_{i1}$ ,  $\mathbf{n}_{i2}$ ,  $\mathbf{n}_{i3}$  and  $\mathbf{n}_{i4}$  denote the independent noise components added to the signal prior to the bolometer low-pass, ac high-pass, ac low-pass filtering, and signal digitization, respectively.  $\mathbf{n}_{i1}$  and  $\mathbf{n}_{i2}$  components together are expected to dominate the total instrumental noise with only  $\mathbf{n}_{i1}$  component displaying a  $1/f$  behavior at low frequencies attributable to temperature fluctuations of the detector. Though all such features are of importance for proper forecasting and simulations of a performance of an experiment, none of these needs to be assumed in the noise estimation described in Sec. II C, which directly estimates the total noise,  $\mathbf{n}_i$ . The instrumental noise is assumed to be Gaussian and stationary within each segment, and is described by a (segment-dependent) noise power spectrum,  $P(f)$ . Each segment is assumed to consist of measurements evenly spaced in time. However, the data constituting gaps are not to be included in a final map.

## 2. Simulations

In order to test our data analysis pipeline we want to be able to simulate the time stream of a MAXIMA-I like experiment. The simulation is designed to incorporate all the important features of the actual data set listed above, including the gap and segment structure, scanning strategy, and an approximate (symmetric) beam [33].

The simulated time stream is described by Eq. (1). The sky signal ( $\mathbf{t}_{sky}$ ) is generated by applying the known pointing solution to a simulated CMB sky, generated as a high-resolution Gaussian realization given some fiducial cosmological parameters. We also include a primary mirror chop synchronous systematic effect,  $\mathbf{x}(\alpha)$ , in our simulations, varying both its functional dependence on the primary mirror position and its amplitude. The instrumental filters are then applied to the simulated time stream as described by Eq. (1). Their functional form follows that of actual MAXIMA filters.

Finally we add Gaussian correlated noise to each time sample. This is modeled according to Eq. (2) assuming that each component,  $\mathbf{n}_{ii}$ , has a power spectrum in a form given by

$$P_{sims}(f) = \sigma_{sims}^2 \left( 1 + \frac{f_{knee}}{f} \right). \quad (3)$$

This approach means that our simulations mimic the range of floating point operations required in analyzing the real data, giving us some insight into possible numerical error accumulation in our data analysis pipeline.

## C. Time stream processing and noise estimation

We now describe the procedure for simultaneously estimating the time-domain noise power spectrum and restoring the stationarity of time stream noise by re-filling the flagged-data gaps in the time stream.

Typically, map-making methodologies assume that the time-domain noise power spectrum is precisely known, even though in practice it has to be estimated from the time stream data itself. The error in this estimation is therefore not included in general in the end-to-end data analysis error budget (although see [11] for a possible way of tackling this issue). This fact clearly highlights the need for a high level of precision at this stage.

The time stream model described by Eq. (1) is quite complex. Some of its features are of less importance for this section so, for simplicity, we start by assuming that the time stream data contains only Gaussian noise (i.e., no sky or systematic signals,  $\mathbf{t}_{sky}$ ,  $\mathbf{x} \ll \mathbf{n}_i$ ) and focus on resolving problems in the noise estimation arising from the presence of gaps in the time stream and of noise correlations on time scales comparable to the length of an entire data segment.

The presence of discontinuities (gaps) in the time stream poses a two-fold problem. It is an obstacle both to estimating the time stream noise spectrum and to devising a fast and efficient map making algorithm (see Sec. III). It is therefore desirable to restore time stream continuity without compromising the stationarity of the noise or sacrificing too much of the data. These two goals, estimating the noise power spectrum and restoring the continuity of the time stream, are achieved by a procedure described below (see Fig. 3 for its synopsis).

### 1. A pure noise time stream

The initial input at this stage consists of the time stream data with the instrumental filters convolved,  $\mathbf{d}_F$ . The filters can be deconvolved in the frequency domain,

$$\mathbf{d}(t) = \int df \tilde{\mathbf{d}}_F(f) \tilde{F}^{-1}(f) \exp(-2\pi i f t). \quad (4)$$

This deconvolution could be done immediately, prior to any further time stream manipulation, but this is not recommended due to the high-frequency noise it adds to the time stream (see Figs. 4 and 5). Although this noise would not be a problem for our formalism in theory, it can be a source of significant numerical error in any practical implementation, possibly biasing the final results. This numerical error is usually incurred while Fourier transforming the noise power spectrum. Instrumental filter deconvolution causes the noise power at the highest frequencies to dominate the total power of the time stream noise by many orders of magnitude, and consequently a small error in the total noise power estimate translates into a substantial error in the noise power estimate in the lower range of frequencies of most interest. The remedy is either to leave the instrumental filter deconvolution until after the noise estimation stage, or to perform the deconvolution at the beginning but concurrently convolving an extra (real) filter,  $\tilde{X}_{re}(f)$ , to suppress the high frequency noise, i.e.,



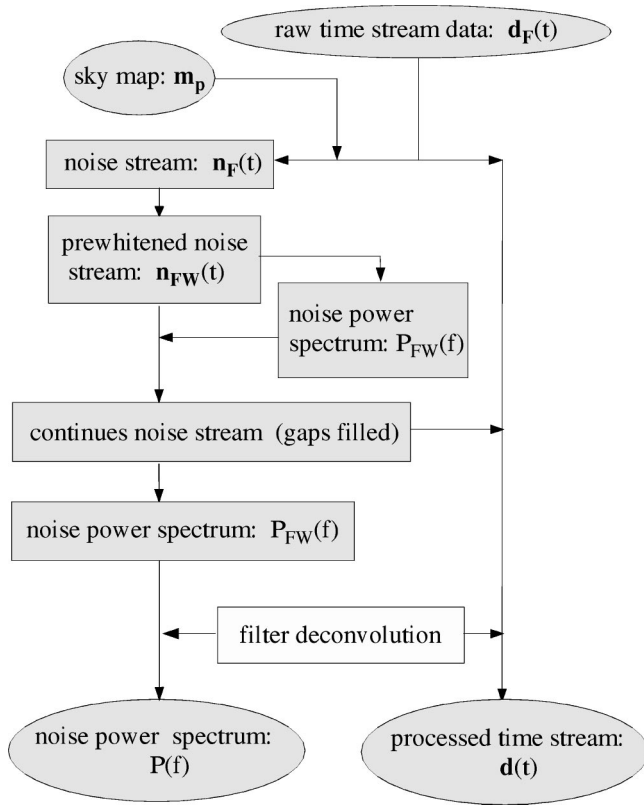


FIG. 3. A flow chart of the noise estimation procedure outlined in Sec. II C. Objects inscribed in ovals are either input or output products of the procedure. This procedure can be incorporated as a part of the iterative noise estimation as described in Sec. V. In this case, its products (noise power spectrum and a processed time stream) are used to generate a sky map, which is one of input objects as shown in this chart.

$$d(t) = \int df \tilde{d}_F(f) \tilde{X}_{re}(f) \tilde{F}^{-1}(f) \exp(-2\pi i f t). \quad (5)$$

In this case, a conventional choice is to apply a real filter given in the frequency domain as the amplitude of the total instrumental filter, i.e.,  $\tilde{X}_{re}(f) = |\tilde{F}(f)|$ . In some map-making algorithms this filter can be deconvolved from the time stream data once the noise power has been estimated and the noise stationarity restored so that effectively no filtering has been applied to the time stream data due the course of data analysis. This is the case when the noise correlation matrix in the time domain does not have to be computed explicitly, as in, e.g., the approximate map-making algorithms discussed below. In these cases the precise shape of the extra filter is not important. If, however, a full noise correlation matrix in time domain is needed, as in exact map-making algorithms, such a deconvolution should be avoided. In this case the precise form of the extra filter is very important, and should be designed to suppress the high frequency noise power sufficiently, while leaving the widest possible range of lower frequencies unaffected.

In the following we set the initial width of all the gaps to be no less than twice the effective width of the instrumental filters (which we conservatively choose to be 100 time

samples, see Fig. 4), reflecting our ignorance about the origin of some of the gaps. Similarly, we always widen gaps on both sides by an effective filter width when (de)convolving any filter from the raw time stream.

The noise estimation algorithm described here is applied to each time stream segment separately and consists of the following steps:

(1) *Reducing the noise correlation length,  $\lambda_c$ :*

The time stream segment is convolved with a prewhitening filter,

$$\tilde{d}_{FW}(f) \equiv \tilde{W}(f) \tilde{d}_F(f), \quad (6)$$

and simultaneously the gaps are widened to account for the width of this filter. The prewhitening filter ( $\tilde{W}$ ) is usually a half-differencing real filter [9]:

$$\tilde{W}(f) \equiv \sin^\beta(\phi/2), \quad (7)$$

with a phase  $\phi = \pi f \Delta$  and a sampling interval  $\Delta$ . The prewhitening index,  $\beta$ , reflects the expected low frequency behavior of a noise power spectrum, i.e.,  $P(f) \sim f^{-2\beta}$  for  $f \rightarrow 0$ . It is adjusted for each segment separately (Fig. 4), so that the prewhitened power spectrum,  $P_W(f) \sim \text{const}$  for  $f \rightarrow 0$ . The functional form of the prewhitening filter is tested in the subsequent stages of the data analysis, and, if necessary, it may be refined and the entire procedure repeated.

(2) *Estimating the noise power spectrum of the prewhitened time stream,  $P_{FW}(f)$ :*

Sections of the time stream segment are located that have no gaps, are longer than the expected noise correlation length, and are separated from each other by at least this distance. These are used to compute a series of statistically independent estimates of the noise power spectra using standard fast-Fourier-transform- (FFT)-based methods (e.g., [34]). These estimates are then checked for stationarity and, if consistent with one another, averaged. If they do not exhibit stationarity then the original time stream segments have to be redefined and made appropriately shorter. The final—average—power spectrum of the stationary noise in a segment can be additionally smoothed and extra(inter)polated to the frequencies corresponding to the full segment.

(3) *Restoring time stream continuity:*

The gaps are filled with a constrained realization of the noise [35] which is assumed to be Gaussian with a power spectrum,  $P_{FW}(f)$ , as determined on the previous step and corresponding noise correlation matrix [see Eq. (10)], denoted as  $N_t$  below. The noise within each gap is constrained only by good data samples in its vicinity, i.e., those which are within the noise correlation length,  $\lambda_c$ , of the gap edges,

$$d_{FW}(i) = \xi(i) + \sum_{m,n} N_t(i,m) N_t'^{-1}(m,n) [d_{FW}(m) - \xi(m)]. \quad (8)$$

Here,  $N_t'$  is a  $2\lambda_c \times 2\lambda_c$  matrix describing noise correlations between the time samples just outside of a given gap,  $\xi$  is a vector of Gaussian, correlated (but unconstrained) random variables, and  $i$  indexes the time samples within a gap. The

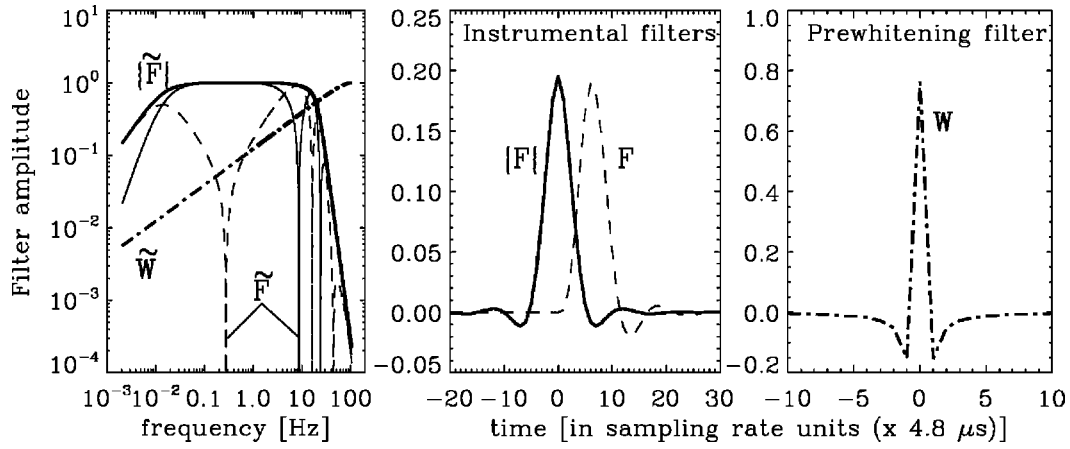


FIG. 4. Filters used in the analysis of MAXIMA-I data. Left panel shows a frequency domain representation of the filters. Thin solid and dashed lines show the (absolute values of) real and imaginary parts of the total instrumental filter,  $\tilde{F}(f)$ , and the thick line its complex amplitude,  $|\tilde{F}(f)|$ . The dot-dashed line depicts the real prewhitening filter,  $\tilde{W}(f)$ , as defined in Eq. (7) ( $2\beta=1.0$ ). The time domain representation of these filters shown in middle panel: solid line—total instrumental filter and dashed line its amplitude; right panel—the prewhitening filter. The effective time domain widths of these filters assumed in the analysis are 100 time samples for both instrumental filters and 50 time samples for the prewhitening filter.

sum is over  $m$  and  $n$ , both of which index the time samples outside of a given gap which are being used to constrain the noise realization within the gap. Due to required inversion of the matrix  $N'_i$  the computational cost of the procedure is  $\mathcal{O}(\lambda_c^3)$  per gap, and it is therefore crucial that  $\lambda_c$  is sufficiently short.

(4) *Re-estimating the noise power spectrum:*

The entire time stream segment, with gaps filled, is now used to re-estimate the noise power spectrum. At this stage the low frequency behavior of the noise can be tested; if it is

found to be significantly non-flat (i.e., non-white), or to differ from the spectrum estimated at stage (2), then the prewhitening filters used at the stage (1) must be adjusted and the entire sequence repeated.

(5) *Deconvolving the filters:*

Once the noise power spectrum has been estimated, and the gaps filled with the constrained noise realization, the instrumental and prewhitening filters can be deconvolved from both the time stream and the noise power spectrum. In general, this applies to both the instrumental (or  $|\tilde{X}_{re}(f)|$ ) and

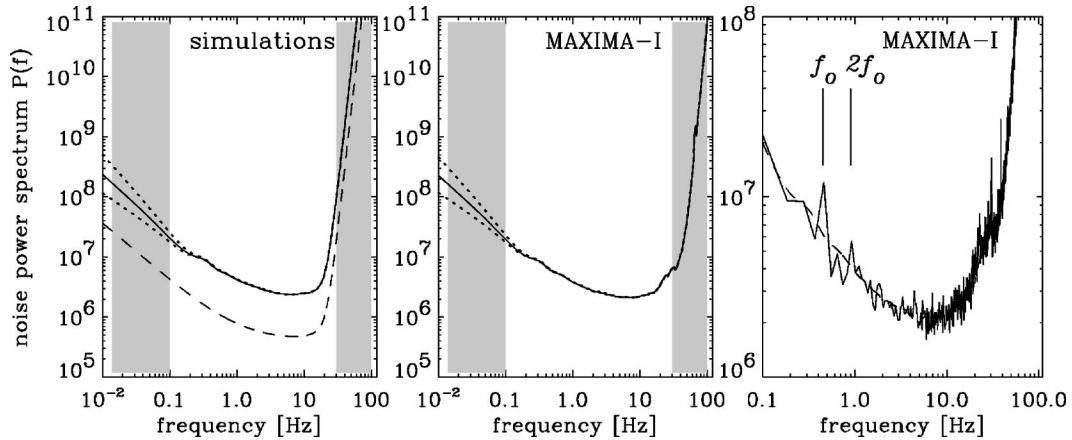


FIG. 5. Estimated noise power spectra for simulated and real data. In left panel, the dashed line shows an effective noise power spectrum with  $2\beta=1.0$  used for a simulation (its amplitude is scaled down by a factor of 5 to avoid overlapping with solid line). Solid and dotted lines show noise power spectra recovered from this simulation using a prewhitening filter as given by Eq. (7) with  $2\beta=1.0$ —solid line;  $2\beta=1.25$  (top) and  $2\beta=0.75$  (bottom) dotted lines. Middle panel shows noise power spectra for a single segment of the actual MAXIMA-I data. Again the recovery has been made using various prewhitening filters as in the left panel. In all the results displayed in these panels aggressive smoothing have been applied to facilitate better comparison. For actual calculations, such an approach does not have to be the best option. Right panel shows a noise spectrum obtained just as a result of averaging at the second step of the noise estimation procedure, prior to any smoothing. Two small spikes at  $\sim 0.45$  Hz and  $0.9$  Hz correspond to a primary mirror chop frequency and its first harmonic. The overplotted dash line is a smoothed version of this spectrum as shown in the middle panel with a solid line. Shaded areas in both left and middle panels show a region of frequencies we marginalize over to avoid either spurious high frequency noise or a noise estimation uncertainty at the low frequency end.

prewhitening filters. As mentioned above, the deconvolution of the instrumental filters is not always advisable since it may introduce numerical errors. The deconvolution of the prewhitening filter is also sometimes postponed, and accounted for only in the algebra of the map-making algorithm [9]. Such an approach attempts to make use of the shortened correlation length of the prewhitened noise to cut the number of floating point operations required to make the map. In practice we find that it is difficult to take advantage of this in any realistic map-making algorithm (see Sec. III), while, as we argue below, performing the deconvolution at the outset of the noise estimation procedure helps to alleviate a number of problems.

This procedure attempts to estimate the ensemble average noise power spectrum using just one realization of the time stream segment. This is clearly not enough, especially in the presence of correlations on time scales comparable with the length of the segment. Prewhitening helps to alleviate this problem, yet it requires some extra knowledge about the functional form of the prewhitening filter. In this approach, an educated guess, followed by iterative refinement, is then tested *a posteriori*. For a prewhitening filter of the form given in Eq. (7), we usually find that the power-law index  $\beta$  can be uniquely determined with an error not bigger than  $|\Delta\beta| \leq 0.1$  by comparing the power spectra computed at the 2nd and 4th stages of the procedure (see Fig. 5).

Computing and averaging the noise power spectra of the independent, gap-free, sections of a segment [step (2) above] helps to recover a better ensemble average spectrum at intermediate and high frequencies. Smoothing in the frequency domain can also be applied to get even closer to the ensemble average. However, in both cases we need to assume that the real noise power spectrum does not display any significant variation on scales smaller than the smoothing or frequency re-sampling scale.

The result of this procedure is an estimate of the ensemble average noise power spectrum in the time domain which is reliable at sufficiently high frequencies (for MAXIMA-I for  $f \gtrsim 0.1$  Hz), but sample-error dominated at the low frequency end ( $f \lesssim 0.1$  Hz). To minimize the effect of the low frequency uncertainty, in the subsequent analysis we marginalize over the low frequency part of the spectrum (see Sec. IV B).

The entire procedure restores the continuity of the time stream segments: the noise is stationary over an entire segment, including gaps, and the sky signal in gaps is zero. This is important in the subsequent stages of the data analysis; we might worry that adding a random (albeit constrained) signal to the data introduces an extra degree of freedom and undermines the uniqueness of the result. Although this is true, while different realizations of the random component may indeed lead to slightly different results, all of these results are necessarily statistically equivalent, with no bias introduced. Moreover, the impact of the extra randomness is significantly reduced by the gap pixel marginalization described in Secs. IV A and IV B. Consequently, we find that, for MAXIMA-I, the final maps and their power spectra are robust and not affected by the random element of this procedure.

Generating constrained noise realizations to fill the gaps is the most time consuming part of this procedure. How important this is for the consistency and accuracy of the entire data analysis pipeline depends on the frequency, size, and regularity of the gaps in a time stream segment. In the case of a handful of narrow, randomly scattered, gaps, an unconstrained, uncorrelated, random noise realization with an rms determined from the rest of the time stream may serve as a convenient first approximation for the analysis.

## 2. A time stream with non-negligible sky signal

If the sky signal present in the time stream data is not negligible, but still sub-dominant, then the noise estimation has to be performed iteratively, following an algorithm proposed by Ferreira and Jaffe [11]. In this case we distinguish between the full (noise + signal) time stream and the noise-only time stream, where the latter is the noise in the time stream data. At each step of the iteration, a maximum-likelihood map estimated on the previous step is subtracted from the time stream giving the current best estimate of the noise stream (see Sec. V). The above noise power spectrum procedure is applied in full only to the noise time stream; only the instrumental filters deconvolution and related gap-widening are performed on the actual time stream (see Fig. 3). Once the noise estimation has been accomplished, only the noise stream continuity is restored. Therefore the gap time samples from the noise time stream data are used to replace their analogues in the full (unprocessed) time stream. This avoids unnecessarily wasting or biasing good data samples which happen to be in a vicinity of a gap.

A simple extension of this iterative approach also allows us to account for problems related to presence of synchronous systematic signals in the data. We discuss the appropriate algorithms and related issues in some detail in Sec. V.

## D. The time domain noise correlation matrix

Formally, maximum likelihood map making requires the full time-time noise correlation matrix, rather than just a noise power spectrum. For an idealized time stream the former is just the Fourier transform of the latter,

$$N_{Ct}(i, j) \equiv \int df P(f) \exp(-2\pi i f(i-j)\Delta). \quad (9)$$

For a finite time stream this expression leads to a circulant matrix denoted by the subscript  $C$  [36]. If the correlation length in the time stream is less than half of the time stream length, then an alternative approximate correlation matrix,  $N_t$ , is given by

$$N_t(i, j) \approx \begin{cases} N_{Ct}(i, j), & \text{if } |i-j| < \lambda_c, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

In practice, numerical error in the estimation (and Fourier transform) of the noise power spectrum means that  $N_t$  computed in this way may not even be positive definite. Such problems can be alleviated by the introduction of an extra power spectrum smoothing window,  $S(f)$ , designed to trun-

cate smoothly the correlations once their amplitude reaches the limits of numerical accuracy. Again, it is prudent to apply the window function prior to deconvolving the prewhitening filter. Equation (9) then assumes a somewhat more complicated form,

$$N_{Cr}(i, j) = \int df |\tilde{W}(f)|^{-2} \exp(-2\pi i f(i-j)\Delta) \times \int df' S(|f-f'|) P_W(f'), \quad (11)$$

where  $\tilde{W}(f)$  is given by Eq. (7). In the following, we make use of this expression whenever computing  $N_{Cr}$ . Its inverse is approximated by an analogous formula [36],

$$N_{Cr}^{-1}(i, j) = \int df |\tilde{W}(f)|^2 \exp(-2\pi i f(i-j)\Delta) \times \int df' S(|f-f'|) P_W^{-1}(f'). \quad (12)$$

Our usual choice for  $S(f)$  is a Gaussian with an appropriately tuned width (typically of 1000–5000 time samples for MAXIMA-I).

### III. MAP MAKING: FORMALISM

#### A. The basic framework

The procedure described above provides the input for the map making *per se*; here, we take as given the time stream data with instrumental filters deconvolved and gaps filled, and the corresponding noise power spectrum for each segment.

In this section we consider a case of a somewhat idealized time stream, leaving a discussion of the fully realistic case to Sec. IV. We consider a time stream consisting of a single segment, neglect most of the systematic contributions, but do admit the presence of (now filled) gaps in the data. The simplified Eq. (1) reads now

$$d_t = A m_p + n_t. \quad (13)$$

Here  $A$  is a pointing matrix assigning each time sample to an appropriate pixel (or a set of pixels in a case of differencing experiments) on the sky (as observed at the given time) with the sky signal given by a pixelized map:  $m_p$ .  $n_t$  is the (Gaussian) noise time stream with correlations given by the power spectrum estimated as in the previous section.

In this case we can write a closed form solution for the map [7,9] (but see [42]). The maximum likelihood estimates of the map,  $m_p$ , and the pixel-pixel noise correlation matrix,  $N_p$ , then read

$$m_p = (A^T M A)^{-1} A^T M d_t, \quad (14)$$

$$N_p = (A^T M A)^{-1} (A^T M N_t M A) (A^T M A)^{-1}. \quad (15)$$

Here,  $M$  is a positive-definite symmetric matrix, and  $N_t$  the time domain noise correlation matrix [Eq. (10)]. If  $M$

$= N_t^{-1}$ , then Eqs. (14) and (15) give minimum variance estimates of  $m_p$  and  $N_p$ . Other choices trade a larger variance for increased computational speed. In particular Tegmark [9] proposed the choice for  $M$  of the circulant part of the  $N_t$  matrix. We discuss this option in Sec. III C below.

In the following we present some different approaches and their application to a realistic MAXIMA-like data set.

#### B. Minimum variance approaches

##### 1. Exact methods

With  $M = N_t^{-1}$ , we get the minimum variance estimates,

$$m_p = (A^T N_t^{-1} A)^{-1} A^T N_t^{-1} d_t, \quad (16)$$

$$N_p = (A^T N_t^{-1} A)^{-1}. \quad (17)$$

The exact implementation of these equations seems a daunting task. The time stream length may easily reach many tens and hundreds of million of samples, making exact inversion of the time domain noise correlation matrix prohibitive. However, for a MAXIMA-I like experiment with segments lengths reaching only  $\mathcal{O}(10^5)$  samples an exact implementation is feasible on a moderately powerful workstation.

At the core of the implementation lies the observation that for a gap-free time stream of the length  $n_s$  with stationary noise the time domain noise correlation matrix is Toeplitz, with  $N_t(i, j) = N_t(|i-j|, 0)$ . The inversion of a Toeplitz matrix can be performed in as few as  $\mathcal{O}(n_s^2)$  operations—clearly feasible for  $\mathcal{O}(10^5)$  time samples—and even faster algorithms are possible (e.g., [36]) bringing that number down to  $\mathcal{O}(n_s \log^2 n_s)$ . The number of operations can be further reduced if the noise correlation length is shorter than the time stream segment length and the noise correlation matrix is therefore band-diagonal.

If gaps are present in the time stream then, in principle, the Toeplitz (stationary) character of the time domain noise correlation matrix is lost. However the gap-filling procedure described above is explicitly designed to restore the stationarity of the noise in the time stream. Since the gap samples contain no sky signal (by construction) they can be treated as data taken while observing a fictitious signal-free pixel in the sky map  $m_p$ . Once the map and its noise matrix are estimated, this extra gap pixel is rejected from the map and the pixel domain noise correlation matrix (effectively marginalizing over it). This is a special application of the extra pixel method discussed in detail below (see Sec. IV A and also [37]).

The computational effort then scales with a number of pixels  $n_{pix}$  and a number of time samples  $n_s$  as:  
noise inverse in time domain:

$$N_t \rightarrow N_t^{-1}: \mathcal{O}(n_s^2);$$

noise inverse in pixel domain:

$$N_t^{-1}, A \rightarrow A^T N_t^{-1} A: \mathcal{O}(n_s^2);$$



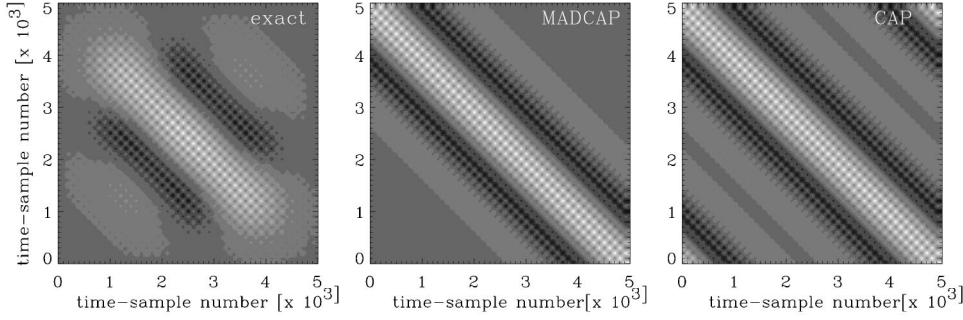


FIG. 6. Inverse time-domain noise correlation matrices computed using MADCAP (middle panel) and circulant (right) approximations and compared with the exact result (left). For visualization purposes the time-stream segment length assumed here is only 5000 samples and the correlation length,  $\lambda_c \approx 1000$ .

noise weighted map:

$$N_t^{-1}, A, d_t \rightarrow A^T N_t^{-1} d_t: \mathcal{O}(n_s^2);$$

noise matrix in pixel domain:

$$N_p^{-1} \rightarrow N_p: \mathcal{O}(n_{pix}^3);$$

final map:

$$N_p, A^T N_t^{-1} d_t \rightarrow N_p (A^T N_t^{-1} d_t): \mathcal{O}(n_{pix}^2).$$

For the first three items substantial savings can be made if the inverse time-time noise correlation matrix,  $N_t^{-1}$ , is assumed to be band-diagonal—an approximation usually well fulfilled for the inverse of a band-diagonal noise correlation matrix. If this is the case, and, in addition, the time domain correlation length is short, then the inverse pixel-pixel noise correlation matrix,  $N_p^{-1}$ , is often rather sparse. In this case additional savings can be made by using the sparse matrix algorithms, e.g., Sherman-Morrison-Woodbury formula (e.g., [34]) to calculate both the map and the noise correlation matrix in the pixel domain.

The memory requirements are also reduced by considering the various matrices' structure. Since we only need to keep the first row of the Toeplitz  $N_t$  matrix and we do not need  $N_t^{-1}$  explicitly, the major memory requirements are set by the size of the  $A^T N_t^{-1}$  matrix, which is  $\mathcal{O}(n_{pix} n_s)$ , and of the noise correlation for the map,  $\mathcal{O}(n_{pix}^2)$ . Note that the first of these limits can be reduced at the expense of the operation counts, since there is no need to save all  $n_{pix} n_s$  elements of this matrix provided we are prepared to re-calculate them as needed.

## 2. Approximate methods

One way to speed up map making, while at the same time providing a good approximation to the optimal minimum variance map was proposed by Ferreira and Jaffe [11] and incorporated into the MADCAP package by Borrill [29]. Rather than explicitly inverting the  $N_t$  matrix we instead use the approximation,

$$N_t^{-1}(i, j) \approx \begin{cases} N_{Cr}^{-1}(i, j), & \text{if } |i - j| \leq \min(n_s/2, \lambda_c), \\ 0, & \text{else.} \end{cases} \quad (18)$$

Here  $n_s$  and  $\lambda_c$  are the time stream and correlation lengths respectively, and the subscript  $C$  denotes the circulant part of

the noise correlation matrix [Eq. (11)]. Apart from this approximation, the remaining steps follow precisely as before. Hereafter we refer to this approach as the MADCAP approximation.

Inversion of a circulant matrix can be accomplished using FFTs [Eq. (12)] in only  $\mathcal{O}(n_s \log n_s)$  operations. This approach is therefore designed to reap the benefits of fast Fourier methods, while at the same time providing a good approximation to the exact solution. If  $n_s \gg 4\lambda_c$  the fraction of  $N_t^{-1}$  matrix elements seriously misestimated (i.e., affected by the segment “edge” effects, see Fig. 6) by this procedure should only be  $\mathcal{O}(4\lambda_c^2/n_s^2)$ , and its performance in the large  $n_s$  limit should be satisfactory. However, because the inverse of a Toeplitz matrix is Toeplitz only when it is also circulant this is the only case when this approach is exact. We demonstrate how well this approximation fares in reproducing the pixel domain map and noise correlation matrix in realistic cases below (Sec. III D).

The operation count changes only for the two items in the list, which now read,

noise inverse in time domain:

$$N_t \rightarrow N_t^{-1}: \mathcal{O}(n_s \log n_s);$$

noise weighted map:

$$N_t^{-1}, A, d_t \rightarrow A^T N_t^{-1} d_t: \mathcal{O}(n_s \log n_s).$$

The scaling of the noise weighted map operation count above assumes the use of FFTs for the Toeplitz matrix multiplication—a trick explained in [36]. The memory requirement is set either by the larger of the size of the noise matrix in the pixel domain  $n_{pix}^2$  or the time stream length  $n_s$ . If an efficient, fast— $\mathcal{O}(n_s \log^2 n_s)$ —implementation of the Toeplitz matrix inversion is viable then, at least in the common case of an approximately band-diagonal inverse noise correlation matrix  $N_t^{-1}$ , the major computational advantage of the approximate method over the exact one would vanish, and the memory requirement would remain as its only asset.

## C. Circulant approaches

If we take  $M = N_{Cr}^{-1}$  then the noise correlation matrix in the pixel domain can be expressed as [9]

$$N_p = N_{Cp} + N_{Sp},$$

$$N_{Cp} \equiv (A^T N_{Cr}^{-1} A)^{-1}, \quad (19)$$

$$N_{Sp} \equiv N_{Cp}(A^T N_{Ct}^{-1} N_{St} N_{Ct}^{-1} A) N_{Cp},$$

where the time domain noise correlation matrix has been decomposed into its circulant ( $N_{Ct}$ ) and sparse ( $N_{St}$ ) parts,

$$N_t \equiv N_{Ct} + N_{St}. \quad (20)$$

The sparse term compensates for the off-diagonal corners of the circulant matrix. As in the case of the MADCAP approximation, the only inversion required in the time domain is that of a circulant matrix. Hence it has been argued that, because the  $N_{St}$  matrix should be very sparse, the operation count in this approach should be significantly lower than in the exact minimum variance approach [9].

The disadvantage of this approach is not in the non-minimum-variance map that results (since the loss of precision is usually insignificant) but rather in implementing its more complicated algebra. If no assumption is made about band-diagonality, then the computational costs scale as:

noise inverse in time domain:

$$N_{Ct} \rightarrow N_{Ct}^{-1}: \mathcal{O}(n_s \log n_s);$$

noise inverse in pixel domain:

circulant part:

$$N_{Ct}^{-1}, A \rightarrow A^T N_{Ct}^{-1} A: \mathcal{O}(n_s^2);$$

sparse part:

$$N_{St}, N_{Ct}^{-1}, A \rightarrow A^T N_{Ct}^{-1} N_{St} N_{Ct}^{-1} A: \mathcal{O}(n_{pix} n_s^2);$$

noise weighted map:

$$N_{Ct}^{-1}, A, d_t \rightarrow A^T N_{Ct}^{-1} d_t: \mathcal{O}(n_s \log n_s);$$

noise matrix in pixel domain:

$$N_p^{-1} \rightarrow N_p: \mathcal{O}(n_{pix}^3).$$

Note that the operation count is dominated by the sparse inverse computation ( $n_s > n_{pix}$ ). The operation count can be reduced if  $N_t$  and  $N_{Ct}$  are assumed to be band diagonal. However, even then, for experiments like MAXIMA the operation count remains  $\mathcal{O}(n_s^2)$ , usually with a large prefactor. Hence this approach is far from competitive with those discussed above, although it can be comparable for very short correlation length data. The memory required again scales as  $\mathcal{O}(n_{pix} n_s)$ .

A possible approximation, which we advocate hereafter, is to neglect the sparse correction completely in the expression for the pixel-pixel noise correlation matrix. Clearly the operation count drops to  $\mathcal{O}(n_s^2)$  or  $\mathcal{O}(n_{pix}^3)$ —whichever is larger—and is usually  $\mathcal{O}(n_{pix}^3)$  especially if  $N_{Ct}^{-1}$  is assumed to be band diagonal. In this case the method's computational requirements are comparable with those of the MADCAP algorithm [29]. In the following we refer to this approach as the circulant approximation (CAP). This approximation is clearly very similar to the MADCAP approach discussed above, although it is strictly exact for a diagonal noise correlation matrix as well as for an infinitely long time stream.

#### D. Comparison and assessment

Although the various map-making methods outlined above are algebraically similar, being derived from the same underlying equations, their implementations and generalized extensions to fully realistic time streams differ considerably.

The approximate methods are of particular interest, because they achieve the speed of fast Fourier transforms. If that is attained without sacrificing the necessary precision is to be tested and may depend on the particular case at hand. The important parameters are the noise correlation length ( $\lambda_c$ ) and the time stream length ( $n_s$ ). If these numbers are comparable the differences between the methods can be significant, but they should disappear whenever  $\lambda_c \ll n_s$  and “edge” effects are unimportant. Given the presence of a so-called  $1/f$  noise component in the time stream, the assumption that  $\lambda_c$  is independent of the time stream length may seem to be plainly wrong. Instead the correlations should be present on time scales comparable with the time stream length. However, a low frequency cutoff (e.g., ac high passing filter in the MAXIMA experiment) usually limits the correlation length independent of the time stream length.

Here we address and illustrate those issues using the MAXIMA-I experiment as an example. For this experiment the average length of the time stream segments is around 50000 samples, with some segments as short as 35000 samples. The very gradual decay of the correlations with time makes it difficult to determine the correlation length precisely. However, we have found that the dependence of the resulting map on the value of  $\lambda_c$  assumed vanishes before  $\lambda_c$  reaches 10000 samples, and have therefore used  $\lambda_c = 10000$  in all computations. Hence, for MAXIMA-I,  $\lambda_c$  is less than but not negligible to  $n_s$ , and prior to taking advantage of the speed of the approximate methods we need to demonstrate their applicability.

Note that comparing the different methods is not entirely straightforward. Assumptions about the band-diagonality of either the time domain noise correlation matrix or its inverse, as required by the different methods, are clearly not equivalent. The relation between the assumed bandwidths of these matrices is also not obvious. It may seem that a fair comparison would be to take the bandwidths to be as large as possible in all methods. However, we then run into the problem of numerical inaccuracies unavoidably present in the calculation of the noise correlation matrix (or its inverse), especially at large time lags, which can result in a non-positive definite (unphysical) noise correlation matrix in pixel domain. Although power spectrum windowing [see Eqs. (11), (12)] can be used to alleviate numerical problems of this sort, this may also have different consequences for each of the methods.

This point is particularly conspicuous when comparisons are made with the exact circulant method. We find that this approach is very sensitive to both the choice of  $\lambda_c$  and the windowing technique. For this reason we do not quote precise numbers for the comparison of this method with the others. By adjusting the free parameters of the method we can reproduce other methods' results quite accurately, and within the expected statistical uncertainty of the maps.

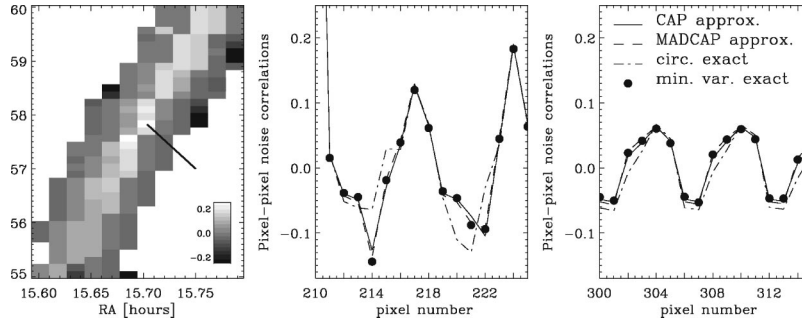


FIG. 7. An example of pixel domain noise correlations estimated for a single segment of MAXIMA-I data set. In this figure, we show correlations of a single selected pixel with its neighbors. This pixel is marked in left panel with an arrow. The color scale shows the correlation level relative to the rms value of the noise in the selected pixel. The scanning direction was from the left bottom corner to the top right, leaving a smudge of strong correlations in pixel domain. Pixels are numbered row by row from the left to the right and from the bottom to the top as shown in the left panel. Note a strongly distorted aspect ratio of the figure due to a scan elongation in a declination direction. These correlations as a function of a pixel number are also displayed in the middle and right panels. They show the relative correlations of the selected pixel with its closer (middle) and more distant neighbors (right panel). The chosen pixel has a number 210. Different lines correspond to different map-making methods used for the correlation estimations. Filled circles show results of the exact minimum variance, solid line shows approximate circulant (CAP), dashed line shows approximate MADCAP and dot-dashed line shows exact circulant methods.

In the case of the other comparisons we have applied a Gaussian window function [Eqs. (11),(12)] while computing the noise power spectrum and kept similar correlation lengths for all three methods. In any case, the numbers quoted below should be looked at as indicative rather than as the best case possible.

We choose to compare maps and noise correlation matrices directly in pixel domain, notwithstanding the fact that the approximations are really applied when the inverse of the noise correlation matrix in the time domain is computed (see Fig. 6). It is the quality of the estimate of the map and its noise correlation matrix which matter for any subsequent analysis. Such an approach also seems to be more general than assessing the quality of a map through the application of a specific statistical tool.

A sample of results is shown in Fig. 7. The middle and right panels show two regimes of pixel-domain noise correlations as estimated using different methods. The left panel shows the correlations of a selected pixel with all the others as projected on the sky. These were computed for a single  $\sim 40000$ -sample segment of the CMB1 scan covering an area of  $\sim 6$  square degrees on the sky, corresponding to  $\sim 400$  square 8 arcmin pixels. Neither filtering nor marginalization has been applied to the time ordered data. The agreement

between results from the different methods is generally quite good. More quantitatively,  $\sim 98\%$  of the matrix elements show relative differences less than 20% when computed using the exact minimum variance approach and the MADCAP approximation, and  $\sim 50\%$  show differences less than 5%. Similar numbers are found for the comparison of this exact method and the CAP approximation. Usually, both approximate methods tend to overestimate high positive correlations. The MADCAP approach also seems to underestimate the amplitude of negative correlations, yet frequently recovers weak correlations more precisely than the other approximation. The relative differences of the maps are of the order of 10% (Fig. 8), which, on average, amounts to no more than 10% of the rms noise level predicted for a given pixel (Fig. 9).

To test if the discrepancies are due to numerical errors, rather than differences in the algorithms, we use the exact method with an explicitly circulant noise correlation matrix, and compare the result with that calculated using the approximate circulant approach. In the absence of numerical inaccuracies both results should be identical. We find that the differences are predominantly at the 2–3 % level. We obtain similar error estimates when analyzing a diagonal noise correlation matrix using all three methods.

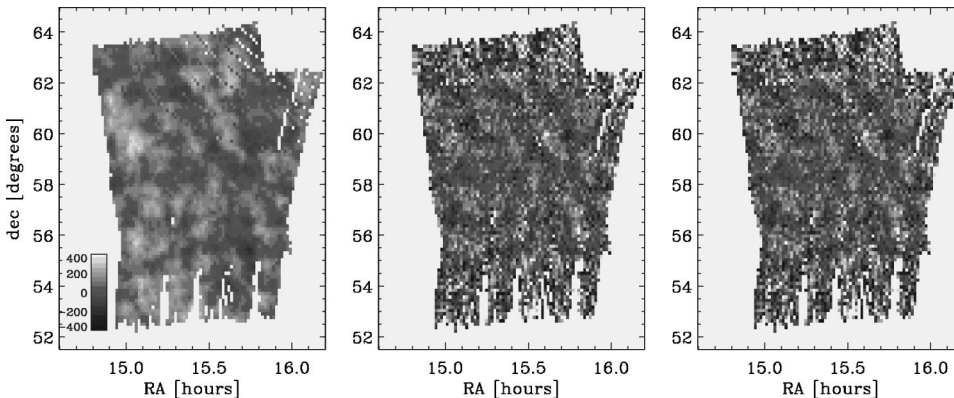


FIG. 8. Simulated MAXIMA-I-like maps of the CMB anisotropy. Left panel shows the simulated map used for creating the time stream data of a single detector of a MAXIMA-like experiment. The time stream was subsequently used to recover the map of the sky applying the approximate CAP (middle) and MADCAP algorithms.



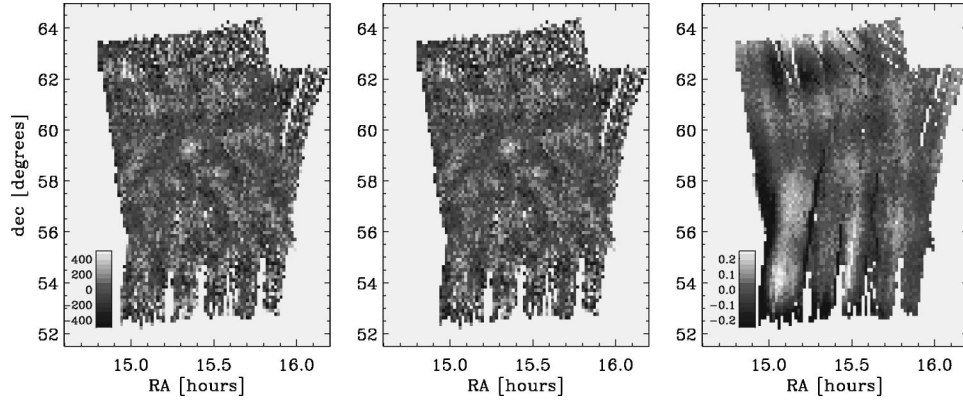


FIG. 9. Comparison of two map-making algorithms applied to the real MAXIMA-I data of a single photometer. The left panel shows the map computed using the approximate circulant (CAP) approach. The middle panel shows the map recovered using MADCAP approach. The difference of both divided by an estimated rms of the pixel noise is shown in the right panel. Note that the color stretch limits are from  $-0.25$  to  $0.25$  in the right panel.

We can also ask if there is any systematic error incurred as a consequence of the approximations which bias the results of a particular approach. One way to address this issue with simulated maps is to check whether the actual noise in an estimated map is properly described by a concurrently estimated noise correlation matrix. This is clearly the case for the exact methods, which are derived to obey such a requirement. To test the approximate approaches we define  $\mathbf{m}_{\text{sky}}$  as the true noiseless map of the sky used for a simulation.  $\mathbf{m}_c$ ,  $N_c$ , and  $\mathbf{m}_m$ ,  $N_m$  are maps and pixel-pixel noise correlation matrices as estimated by the CAP and MADCAP approximate methods respectively. For each of the noise matrices we introduce a matrix  $N_i^{-1/2}$  such that  $N_i^{-1/2}(N_i^{-1/2})^T = N_i^{-1}$ . Consequently the variable  $y_i$ , defined as

$$y_i \equiv N_i^{-1/2}(\mathbf{m}_i - \mathbf{m}_{\text{sky}}), \quad i = c, m, \quad (21)$$

should be a vector of uncorrelated, Gaussian variables with a dispersion equal to unity if no systematic problem is introduced by an approximation. This can be tested using, e.g., the Kolmogorov-Smirnov test (see also Sec. VI). Such a test is comfortably passed by both methods for noise in the pixel domain ranging from  $\sim 100 \mu\text{K}$  per 8 arcmin pixel (the MAXIMA-I single channel level) down to  $\sim 60 \mu\text{K}$  (the MAXIMA-I 4-channel combined level).

The differences between the results of the various methods, as summarized above, are found to be rather small, and indicate that all the methods may be equivalent in practice. Although the problem is difficult to tackle in a more quantitative and general way, it can be addressed from the point of view of specific statistics derived from the maps. An example of such a test, comparing the power spectrum statistic, is shown in Fig. 10. In this case the power spectra of the MAXIMA-I maps computed using both map-making methods show very good agreement.

In practice, the approximate methods have an advantage over the exact methods of being easily implementable. In fact, both can be straightforwardly applied using the MADCAP package [28]. The differences in the results provide some insight into the sensitivity of the results to the treatment of

time domain noise correlations. More to the point, the result of any statistical test applied to the map, which depends on the choice of the approximate method used during map-making, should be treated with suspicion. We have demonstrated that for the MAXIMA-I data analysis and, in particular, for power spectrum estimation, the differences between the methods are negligible.

In summary, all four map-making methods are broadly consistent at the noise level of the MAXIMA-I maps. This concordance is only expected to improve if longer time stream segments are considered. Numerical errors can readily be kept under control even with as many as  $\mathcal{O}(10^5 - 10^6)$  time samples, at least for MAXIMA-like data sets, although further tests may be needed for low noise cases. Both approximate methods are easy to implement, with their speed being limited by the pixel domain noise correlation matrix inversion. The exact minimum variance method provides a useful cross-check on both the validity of the approximations

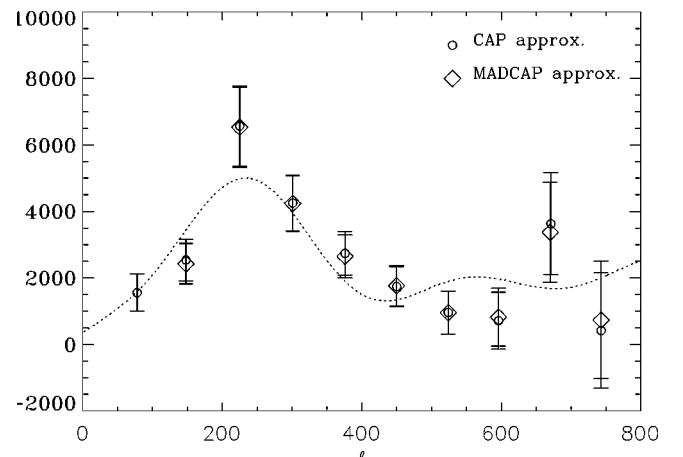


FIG. 10. Comparison of power spectra computed for the maps made with two different map-making algorithms as shown in two left panels of Fig. 9. The data are those of a single MAXIMA-I detector. Circles show the results for the map made using approximate circulant (CAP) algorithm, and diamonds show the results for the map made using the MADCAP approximation.



and numerical error propagation, and, with further improvements may be able to achieve the computational speed of the approximate approaches.

We note again that all of this assumes a continuous time stream with stationary noise resulting from application of the gap filling algorithm; the restoration of the Toeplitz character of the noise correlations is crucial for the feasibility of the exact minimum variance approach, as well for as the accuracy of the approximate methods.

#### IV. MAP MAKING: AMENDMENTS

So far we have been assuming an “optimistic” model for the signal in the time stream,  $\mathbf{d}_t$  [Eq. (13)]. Commonly, unwanted contributions (e.g., the overall average temperature, primary mirror- or gondola- synchronous noise) are present in the data [Eq. (1)] and have to be dealt with if a reliable result is to be obtained.

Let us consider a case of an arbitrary time domain contribution of known temporal behavior. We take these to be given by a set of  $m$  template vectors  $\boldsymbol{\tau}_t^{(i)}$ , of unknown amplitudes,  $x^{(i)}$ , so the resulting time stream equation is

$$\mathbf{d}_t = \mathbf{A}\mathbf{m}_p + \mathbf{n}_t + \sum_{i=1}^m x^{(i)} \boldsymbol{\tau}_t^{(i)}. \quad (22)$$

The additional number  $x^{(i)}$  can be formally treated as an extra (fictitious) pixel “observed” with a “pointing matrix” as given by  $\boldsymbol{\tau}_t^{(i)}$ . These additional numbers can be estimated during the course of a map-making procedure and then marginalized over. Alternatively, the entire extra term,  $\sum_i x^{(i)} \boldsymbol{\tau}_t^{(i)}$ , can be thought of as a part of the noise term and directly marginalized over in the course of map making. Clearly these two approaches are just different implementations of the same idea, yet, depending on the particular problem at hand, one or the other may be more appropriate. In the following, we refer to them as the *extra pixels* and *marginalization* methods, respectively, and discuss them and some of their applications.

##### A. Extra pixels

Let us define a vector  $\mathbf{x}_q$  as

$$\mathbf{x}_q^T \equiv [x^{(1)}, \dots, x^{(m)}], \quad (23)$$

and a matrix  $\mathbf{B}$  such that

$$\mathbf{B} \equiv [\boldsymbol{\tau}_t^{(1)}, \dots, \boldsymbol{\tau}_t^{(m)}]. \quad (24)$$

We can now rewrite Eq. (22) as

$$\mathbf{d}_t = \mathbf{A}\mathbf{m}_p + \mathbf{B}\mathbf{x}_q + \mathbf{n}_t. \quad (25)$$

From this it is clear that the extra time stream contribution can be thought of as a set of extra (fictitious) pixels ( $\mathbf{x}_q$ ) “observed” with a pointing matrix given by  $\mathbf{B}$ . If we, furthermore, recast this equation as

$$\mathbf{d}_t = [\mathbf{A}, \mathbf{B}] \begin{bmatrix} \mathbf{m}_p \\ \mathbf{x}_q \end{bmatrix} + \mathbf{n}_t \equiv \mathcal{A} \hat{\mathbf{m}}_p + \mathbf{n}_t, \quad (26)$$

we find that it is formally identical to Eq. (13). Here, we have introduced a generalized pointing matrix,  $\mathcal{A}$ , and a generalized map,  $\hat{\mathbf{m}}_p$ . Those correspond to an “ordinary” map and a pointing matrix, but now extended to incorporate also the extra pixels. The map-making methods discussed above can all be applied in the present case, with all their caveats and strengths. The appropriate equations preserve the form of Eqs. (14) and (15) but the pointing matrix,  $\mathbf{A}$ , the map,  $\mathbf{m}_p$ , and the pixel domain noise correlation matrix,  $\mathbf{N}_p$ , need to be replaced by their generalized counterparts, i.e.,  $\mathcal{A}$ ,  $\hat{\mathbf{m}}_p$  and  $\mathcal{N}_p$  respectively. As a result, the map-making procedure provides an estimate of a generalized map and a generalized noise correlation matrix in pixel domain.

In addition to the minimum-variance considerations described above, the formalism so far described can also be thought of as describing the likelihood function for the data: the distribution of the underlying (generalized) map is Gaussian with mean and variance as given by Eqs. (16) and (17) (or approximations to these). Thus, the generalized map is a simultaneous estimate of both the real map and the “extra” pixels. Knowledge of the amplitudes of these extra pixels are often useful for diagnostic purposes. However, we are in the end interested in the real map itself; hence we wish to marginalize over the  $\mathbf{x}_q$ . With a uniform prior, we find the usual Gaussian results:

the marginalized map is just the generalized map with the extra pixels removed; and

the marginalized noise matrix is just the generalized noise matrix with the rows and columns corresponding to the extra pixels removed.

Clearly, only well-understood features of the time stream, for which the pointing operator  $\mathbf{B}$  is known, can be accounted for using this approach. Moreover, if the extra component is indistinguishable from the CMB temperature map, i.e., is sky stationary, then the clean separation of the map into CMB and non-CMB components is impossible. That is manifested as a singularity of the  $\mathcal{A}^T \mathbf{M} \mathcal{A}$  [see Eqs. (14) and (15)] matrix. In some cases, if such a singular pixel-domain mode is determined, it can be accounted for on subsequent stages of the analysis (see Sec. IV C). In general the discussed framework is quite universal and can be applied successfully in a variety of circumstances of practical interest. Specific examples of its applications are time stream gaps, relative offsets between separate segments of the map, and primary mirror synchronous effects. All those were found of importance for the MAXIMA-I data analysis and we discuss them in detail below.

*a. Time-stream gaps, offsets and temporal frequencies.* Perhaps the most straightforward applications of this method are to time stream gaps, time-stream segments offsets, and unwanted temporal frequencies.

The direct application of the extra pixel method to the unprocessed raw time stream in order to properly marginalize over the unknown content of the gaps would require introducing as many extra pixels [parameters  $x_i$  in Eq. (25)] as

time samples in the gaps. Not only is this a computationally significant extension, but also a source of a multitude of possible numerical problems and singularities. With the gap-filling procedure as described earlier, one extra  $x$  parameter (the gap pixel) and a single template,  $\tau_i$ , such as

$$\tau_i(i) = \begin{cases} 1, & i \in \text{gaps}, \\ 0, & \text{otherwise}, \end{cases}$$

suffices to account for all of the gaps of a single segment. That is precisely the gap pixel approach briefly mentioned in Sec. III B.

If a map of more than a single segment is to be produced from a single map-making procedure (an approach adopted in MADCAP), care must be taken with regard to unknown relative offsets between different segments. In total power experiments the actual offsets are spurious and contain no information about the sky. In the parlance of the extra pixel method, the offsets can be considered as amplitudes,  $x^{(I)}$ , of a set of time-domain templates,  $\tau_i^{(I)}$ , defined for each segment  $I$  as

$$\tau_i^{(I)}(i) = \begin{cases} 1, & i \in I, \\ 0, & \text{otherwise}, \end{cases}$$

and therefore they can be straightforwardly incorporated within the framework of the extra pixel method and the map-making formalism [29].

Similarly unwanted temporal frequencies can be described with one extra parameter, their common amplitude, if they are first filtered out from the time stream, and then replaced by a pure Gaussian noise realization with the noise power spectrum as estimated earlier.

*b. Primary mirror synchronous signal.* Periodic motion of the primary mirror and the gondola can potentially become a source of an extra parasitic contribution to the total signal measured by a MAXIMA-like experiment. Due to its origin, such a contribution is likely to be a single-valued function of the position of either the primary mirror or the gondola and therefore can be modeled using the extra pixel method discussed above.

Only the primary mirror synchronous effect has been found to be important for the actual MAXIMA-I data and required this treatment. Below we describe this case in some detail as an example.

Assuming that the primary mirror synchronous contribution is a slowly varying function of the mirror position, we can characterize it using a discrete set of amplitudes (i.e., an extra pixel “map”), each of which describes the magnitude of the parasitic signal corresponding to a narrow range of mirror positions (i.e., an extra pixel). The extra pointing matrix  $B$  assigns then the time samples to these discretized primary mirror positions. The presence of the instrumental filters introduces an additional complication. They induce phase shifts in the time stream and therefore modify the pointing matrix in a way depending on the precise location that the synchronous signal appeared in the on-board electronics chain (which we do not know *a priori*). In principle, there are four possible choices for the correct pointing matrix

of the primary mirror synchronous signal. However, the effective, total, phase shift, caused by the instrumental filters, is strongly dominated by the ac low pass filter, leading, therefore, to only two significantly different choices for the combined pointing matrix of this synchronous effect,  $B$ . We choose those to be

$$B = B,$$

or

$$B = F_{bolo}^{-1} F_{low}^{-1} F_{high}^{-1} B.$$

Here we use  $B$  to denote the discretized mirror position and denote instrumental filters as in Sec. II B. The first choice above corresponds to the case in which the synchronous signal is added prior to ac low pass filtering, e.g., a mirror related modulation, and the second one to the case in which the extra contribution arises later on.

A number of extra pixels depends on a particular problem at hand. In the MAXIMA-I case we use a separate set of extra pixels, consisting of 20 to 50 discretized mirror positions, for each of the segments. The relative offset of the primary mirror signal and the CMB map is arbitrary, therefore we constrain the value assigned to the first extra pixel (corresponding to the leftmost position of the mirror) to be zero. That breaks the degeneracy between absolute offsets for both components and avoids a singularity in the generalized noise correlation matrix in pixel domain.

For MAXIMA-I we find no other singularity caused by the inclusion of extra pixels describing the primary mirror synchronous signal in the map-making problem. This is due to the carefully designed MAXIMA-I scanning strategy; the typical time scale for variation of the extra contribution is significantly longer than the time needed for the instrument antenna to cross the pixel on the sky, and the majority of real sky pixels are revisited many times, with the primary mirror position different at each visit.

By using a single set of extra pixels per segment, we make an implicit assumption about the stationarity of the underlying synchronous signal on the time scale of the length of the segment. Such an assumption needs to be tested *a posteriori*. Recovered primary-mirror-synchronous signals are shown in Fig. 11. These results are for simulations with the synchronous signal explicitly assumed to be strictly stationary, and for the actual MAXIMA-I data. In the latter case, the results for different, but adjacent, segments of the same detector are indeed consistent within the error bars, suggesting that the assumption of stationarity is satisfied.

The differences between recovered signals can be the result of actual differences of the instrumental signal, or of the sky signal, which can have a component synchronous with the mirror position, or because of instrumental noise. We find that for the MAXIMA-I experiment, a part of the dipole signal is subtracted from the sky map together with the primary mirror synchronous signal. This effect is mainly due to the monotonic gradient-like morphology of the dipole within the boundaries of the  $\sim 10^\circ \times 10^\circ$  patch observed by MAXIMA-I. For all multipole modes which vary significantly across that patch, i.e., for  $l \gtrsim 20$ , such an effect is expected to be unim-

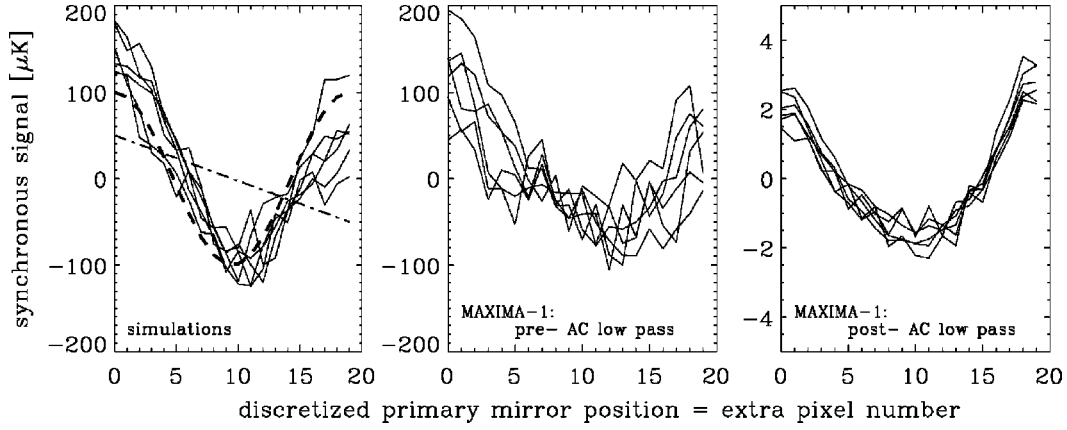


FIG. 11. The primary mirror synchronous signal recovered from the simulation (left panel) and the real MAXIMA-1 data as a function of the discretized position of the primary mirror with respect to the gondola. In each of the panels, the recovered signal for each of the six segments of the CMB1 scan is shown with thin solid line. In the left panel, the thick dashed line shows the primary mirror synchronous signal as used for the simulation. The apparent asymmetry of the recovered signal in contrast with the symmetry of the function used for simulation is due to a residual CMB dipole contribution shown with a thin dot-dashed line. A very similar, slightly asymmetric, shape of the functional dependence is found for the signal recovered from the real data as shown in the middle panel. That panel shows the estimated amplitude of the signal synchronous with the mirror position, which was added to the time stream data prior to the ac low-pass filtering and therefore also includes the sky signal, such as the CMB dipole. The rightmost panel shows the estimated primary mirror synchronous signal added to the time stream after ac low pass filtering, and therefore originating in the instrumental electronics. Note the very good stability of that contribution with time and its negligible amplitude, as compared to the signal shown in the middle panel, as well as to the expected CMB anisotropy. The noise level corresponding to different mirror positions is strongly correlated and for that reason not shown in the figure. A typical level is  $\sim 40 \mu\text{K}$  and  $\sim 1 \mu\text{K}$  for the middle and left panel.

portant. We show effects of the subtraction of the primary mirror synchronous signal on a power spectrum estimation in Fig. 12.

### B. Marginalization

In the extra pixel method, we first determine the generalized map and noise matrix, and then marginalize over the unwanted pixels. In the case of the marginalization approach, we first marginalize over the unwanted temporal modes, and then pass directly to the marginalized map and noise matrix. It is based on the idea that the undetermined amplitude can be treated as a random variable with a prior probability with dispersion  $\sigma_x^2$  [18]. Again, we start with the more complicated time stream of Eq. (22), but for compactness allow only for a single template,  $\tau_t$ . Now, we assign the unknown amplitude,  $x$ , a prior probability density with variance  $\langle x^2 \rangle = \sigma_x^2$ . Then, we marginalize over this unknown amplitude right away, before making the map. We then recover a distribution for the time stream in the form of a Gaussian with an effective time stream noise matrix given by (hereafter  $\otimes$  denotes a tensor product)

$$N'_t = N_t + \sigma_x^2 \tau_t \otimes \tau_t^T. \quad (27)$$

That is, we can consider both the Gaussian noise ( $n_t$ ) and the template ( $x\tau_t$ ) together as a noise-like term with this correlation matrix. The more complicated equation (22) can be then recast in the familiar form of Eq. (13). Solving for the maximum likelihood map gives expressions in the pixel domain mirroring that of Eq. (14), with the noise correlation matrix in time domain replaced now by  $N'_t$ . Marginalizing

over the amplitude,  $x$ , while taking the limit  $\sigma_x^2 \rightarrow \infty$ , is equivalent to marginalization with a uniform prior, as considered above [18]. This is tractable because we can simplify the calculation of  $N'_t{}^{-1}$  using the Sherman-Morrison-Woodbury formula [34], in this context given by

$$N'_t{}^{-1} = N_t{}^{-1} - \frac{(N_t{}^{-1} \tau_t) \otimes (N_t{}^{-1} \tau_t)^T}{\sigma_x^2 + \tau_t^T N_t{}^{-1} \tau_t} \xrightarrow{\sigma_x^2 \rightarrow \infty} N_t{}^{-1} - \frac{(N_t{}^{-1} \tau_t) \otimes (N_t{}^{-1} \tau_t)^T}{\tau_t^T N_t{}^{-1} \tau_t}. \quad (28)$$

As expected we have

$$N'_t{}^{-1} \tau_t = 0. \quad (29)$$

Finally, note that the usual map-making formulas in a minimum variance case [Eqs. (16) and (17)] require only the inverse noise correlation matrix,  $N_t{}^{-1}$ , and remain unchanged if  $N_t{}^{-1}$  is replaced by  $N'_t{}^{-1}$ . Because the correction to  $N_t{}^{-1}$  is additive, we can also think of this as a correction to the inverse pixel noise:

$$N'_p{}^{-1} \rightarrow N_p{}^{-1} - \frac{(A^T N_t{}^{-1} \tau_t) \otimes (A^T N_t{}^{-1} \tau_t)^T}{\tau_t^T N_t{}^{-1} \tau_t}. \quad (30)$$

One might suspect that a weakness of the marginalization method would be the computational overhead involved in computations of the extra term in Eq. (28). However, the products of  $N_t{}^{-1}$  matrix and a template vector  $\tau_t$  requires

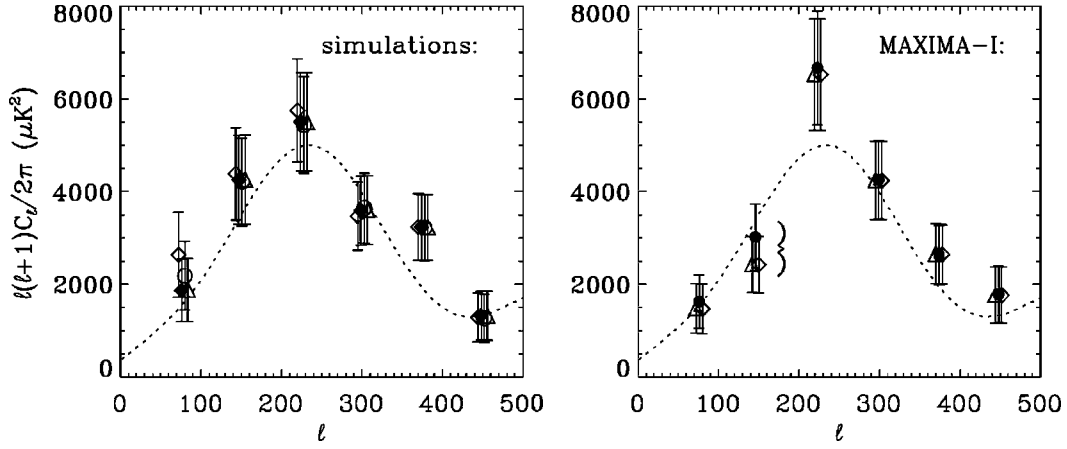


FIG. 12. Power spectra of the CMB anisotropies recovered from the simulations (left panel) and the MAXIMA-I single detector map (right). In the left panel all shown spectra are computed assuming the same CMB sky and noise realization and a MAXIMA-I like observation. We demonstrate here the performance of the extra pixel method in a case of primary synchronous signal. The displayed points correspond to four cases: with a primary synchronous signal absent in the time stream data and either not accounted for during map making (triangles) or accounted for (open circles) using the extra pixel method on the map-making stage. The remaining two cases, with a primary mirror signal present in the data are shown with open diamonds (no extra pixel method applied) and filled diamonds computed for the map produced with the extra pixel method. Right panel shows results of tests of the extra pixels method applied directly to the data. Spectra shown with open symbols have been computed using different definitions of the extra pixels: diamonds are for the standard case with synchronous signal depending only on the position of the antenna, and triangles are for the case when also a dependence on a direction of the primary mirror motion (left or right) is allowed. Filled circles depict the spectrum computed for a map with no extra pixel method applied during map making. Clearly, only the power in the bin centered at  $\sim 150$  has been affected by the primary mirror synchronous signal.

only  $\mathcal{O}(n_s \log n_s)$  operation if  $N_t^{-1}$  is Toeplitz, or  $\mathcal{O}(n_s \lambda_{band})$  if it is (approximately) band-diagonal with a bandwidth,  $\lambda_{band}$ . Also, recalling that what we need in Eq. (30) is  $A^T N_t'^{-1} A$  rather than  $N_t'^{-1}$  itself, we can cut the number of necessary operations by performing the products from outside inwards. That can be done in either  $\mathcal{O}(n_s)$  or  $\mathcal{O}(n_{pix}^2)$  floating point operations, whichever is larger. So the final scaling for any single template is given either by  $\mathcal{O}(n_s \log n_s)$  or  $\mathcal{O}(n_{pix}^2)$ .

If more than one template is required than Eq. (28) needs to be generalized and reads

$$N_t'^{-1} = N_t^{-1} - (N_t^{-1} B) [B^T N_t^{-1} B]^{-1} (N_t^{-1} B)^T \quad (31)$$

where  $B$  is as defined in the previous section [Eq. (24)], and the expression in the square brackets is a square matrix of the size given by a total number of time-domain templates. We have assumed that the template amplitudes are uncorrelated, i.e.,  $\langle x^{(i)} x^{(j)} \rangle = 0$  if  $i \neq j$ . In this case again [cf., Eq. (29)],

$$N_t'^{-1} B = 0, \quad (32)$$

and, hence, also for each template,  $\tau_t^{(i)}$  [cf. Eq. (24)],

$$N_t'^{-1} \tau_t^{(i)} = 0. \quad (33)$$

This approach is fully general and can be used in all the cases already mentioned in the context of the extra pixel method—both methods give identical results. The advantage of the extra pixel method is that it not only marginalizes over the unwanted effect, but also computes its individual

maximum-likelihood estimate. That can be useful for *a posteriori* tests, if some prior assumptions have been made to extract the effect.

Note also that we have constructed a matrix we call  $N_t'^{-1}$ . However, in the limit  $\sigma_x^2 \rightarrow \infty$ ,  $N_t'$  itself no longer exists: by construction it has an infinite eigenvalue corresponding to the eigenvector  $\tau_t$ . (There are also pathological cases where  $N_t'^{-1}$  can be inverted because of numerical effects.) Nonetheless, the pixel-domain noise matrix  $N_p' [\equiv (A^T N_t'^{-1} A)^{-1}]$  may still exist due to mixing of the modes via the operator  $A$ . In fact the resulting  $N_p'^{-1}$  is singular only if  $\tau_t = A v_p$ , where  $v_p$  is a pixel domain template. Then from Eq. (30) it follows that [cf. Eq. (29)]

$$N_p'^{-1} v_p = 0. \quad (34)$$

In cases when  $N_p'$  exists the “extra pixel” approach will work as well. However, the marginalization procedure would work even if  $A^T N_t'^{-1} A$  is singular as we discuss that in Sec. IV C.

Below we elaborate on a particular application of the marginalization method to the removal of specific frequency modes from the time stream.

*Frequency band marginalization.* Here we will use the method of the explicit marginalization, outlined above, to derive a useful concise formula marginalizing over unwanted frequency bands. In such a case a required set of templates is given in the time domain as

$$\tau_{(m)}(j) \equiv \cos\left(\frac{2\pi i_0(j+m)}{n_s}\right). \quad (35)$$



Here,  $i_0$  corresponds to the frequency  $f_0 = i_0/n_s\Delta$ , which is to be marginalized over,  $j$  is a time variable, and  $m = 0, \dots, [n_s/i_0]$  determines the overall phase shift and also numbers the template. As usual  $n_s$  stands for the length of the time stream segment under consideration and  $\Delta$  is the sampling rate. We can now insert these templates into Eq. (31) to obtain a fully general formula for the inverse noise correlation matrix with a frequency mode,  $f_0$ , marginalized over. As explained above, a numerical implementation of such a strict marginalization is feasible within a framework of the minimum variance methods. This expression simplifies significantly if  $N_t^{-1}$  is a circulant matrix,  $N_t^{-1} = N_{Ct}^{-1}$ . Then  $N_t'^{-1}$  is also circulant,  $N_t'^{-1} = N_{Ct}'^{-1}$ , and both are uniquely defined by their power spectra [cf., Eq. (9)]. Here, we denote these as  $\mathcal{P}'(f)$  and  $\mathcal{P}(f)$ . From Eq. (31) we then have

$$\mathcal{P}'(i) = \mathcal{P}(i) - \mathcal{P}(i_0)(\delta^K(i, i_0) + \delta^K(i, -i_0)). \quad (36)$$

Here  $\delta^K$  is the Kronecker delta, and we have made use of the fact that in Fourier space the template  $\tau_{(m)}$  is represented as

$$\begin{aligned} \tilde{\tau}_{(m)}(j) \equiv & \cos\left(\frac{2\pi i_0 m}{n_s}\right) [\delta^K(j, i_0) + \delta^K(j, -i_0)] \\ & + \iota \sin\left(\frac{2\pi i_0 m}{n_s}\right) [\delta^K(j, -i_0) - \delta^K(j, i_0)], \end{aligned} \quad (37)$$

with  $\iota$  denoting an imaginary unit. From Eq. (36), we have  $\mathcal{P}'(i_0) = \mathcal{P}'(-i_0) = 0$ , and therefore, in the case of circulant noise matrices, the extra marginalization term in Eq. (31) just zeros the power of the frequency mode which is to be marginalized over. In fact, such an answer could have been guessed, by noting that the single-frequency modes are eigen-modes of the inverse of the circulant matrix (as well as the circulant matrix itself) and that to introduce a Sherman-Morrison-Woodbury-like correction with respect to any of those modes it is sufficient to set their corresponding eigenvalues to zero [Eq. (33)].

This observation suggests a simple prescription for frequency marginalization in the case of the circulant approximate map-making algorithm discussed before. Because the power spectrum related to the matrix  $N_{Ct}^{-1}$  is the inverse of the noise power spectrum [cf. Eq. (12)], it is possible to marginalize over the amplitude of a frequency mode,  $f_0$ , by setting the inverse of the noise power spectrum corresponding to that frequency to zero prior to calculating  $N_{Ct}^{-1}$  via Eq. (12).

The impact of such a procedure on the final map is clear from Eq. (14): the frequency modes with zeroed power are removed from the map but that is self-consistently accounted for in the corresponding noise correlations matrix.

Similarly, the MADCAP approach can be modified by adopting the following approximation to the inverse noise correlation matrix in time domain [cf. Eq. (18)],

$$N_t'^{-1}(i, j) \approx N_{Ct}'^{-1}(i, j), \quad \text{if } |i - j| < n_s/2, \quad (38)$$

and zero otherwise. The eigenvectors of  $N_t'^{-1}$  are not in fact identical with those of  $N_{Ct}'^{-1}$  and, therefore, the former will not usually have the eigenvalues equal precisely to zero anymore. The removal of unwanted modes from the map in this case is therefore only approximate. Again, this is an effect we have found to be negligible in practice.

One may wish to marginalize over frequency bands which are compromised by a periodic parasitic signal (e.g., synchronous effects). However, in this case this method is less discriminating than the extra pixel method, making no use of the phase information usually available. Marginalization is also useful to minimize the significance of sampling uncertainty present at low frequency and leading to errors in the noise estimation procedure (see Sec. II C). It can also be applied at the high frequency end, where the precise shape of the instrumental filters is not well known. In fact, in the MAXIMA-I case we have applied marginalization to deal with the lowest,  $\lesssim 0.1$  Hz, and the highest frequencies,  $\gtrsim 30$  Hz, for the reasons just mentioned. We have found no strong dependence of our results on the specific choice of bounds, obtaining nearly identical results if these values are set to be 0.2 Hz and 20 Hz respectively.

### C. Singularities and pixel templates

Having accounted for these time stream “templates,”  $\tau_t$ , we can produce the map and the *inverse* pixel-domain noise matrix. As mentioned above, we may not be able to obtain the noise matrix itself, due to zero eigenvalues in the inverse, corresponding to “infinite noise” in some modes.

However, just as the mapmaking procedure *per se* only requires  $N_t'^{-1}$ , subsequent manipulations of the map often can be cast in terms of matrix inverses. (In another language, infinite noise corresponds to zero weight.) For example, the likelihood function if we consider a Gaussian-distributed signal is

$$\mathcal{L}(C_l) = \frac{1}{|2\pi\mathbf{M}_p|^{1/2}} \exp\left[-\frac{1}{2}\mathbf{m}_p^T \mathbf{M}_p^{-1} \mathbf{m}_p\right], \quad (39)$$

where  $\mathbf{m}_p$  is the pixelized map as before, and  $\mathbf{M}_p$  is the variance corresponding to the uncorrelated sum of the CMB signal,  $S_p(C_l)$ , and the pixel-domain noise correlation matrix, i.e.,  $\mathbf{M}_p \equiv S_p(C_l) + \mathbf{N}_p$ . If, due to time domain marginalization,  $N_p^{-1}$  has zero eigenvalue corresponding to a pixel-domain template [cf. Eq. (34)],  $\mathbf{v}_p = (\mathbf{A}^T N_t'^{-1} \mathbf{A})^{-1} \mathbf{A}^T N_t'^{-1} \tau_t$ , then we compute  $N_p$  as

$$N_p = (N_p^{-1} + \epsilon \mathbf{v}_p \otimes \mathbf{v}_p^T)^{-1} - \epsilon^{-1} \mathbf{v}_p \otimes \mathbf{v}_p^T, \quad (40)$$

where  $\epsilon$  is a small positive number and the inversion on the right-hand side (RHS) is to be performed directly. This procedure replaces an infinite eigenvalue of the  $N_p$  (corresponding to the eigenmode  $\mathbf{v}_p$ ) with zero, leaving all other eigenvalues unaffected. The total inverse correlation matrix is now to be understood as

$$\begin{aligned} \mathbf{M}_p^{-1} &= \lim_{\sigma_v^2 \rightarrow +\infty} (\mathbf{S}_p(C_I) + \mathbf{N}_p + \sigma_v^2 \mathbf{v}_p \otimes \mathbf{v}_p^T)^{-1} \\ &= (\mathbf{S}_p(C_I) + \mathbf{N}_p)^{-1} - \frac{\hat{\mathbf{v}}_p \otimes \hat{\mathbf{v}}_p^T}{\mathbf{v}_p^T \hat{\mathbf{v}}_p}, \end{aligned} \quad (41)$$

where  $\hat{\mathbf{v}}_p \equiv (\mathbf{S}_p(C_I) + \mathbf{N}_p)^{-1} \mathbf{v}_p$  and the last term is a usual Sherman-Morrison-Woodbury term. This expression assumes that  $\mathbf{S}_p(C_I) + \mathbf{N}_p$  is invertible. However, any zero eigenvalue modes, additional to  $\mathbf{v}_p$ , can be treated in the same way as  $\mathbf{v}_p$ , once a singular mode has been determined.

Even if we have not explicitly accounted for effects that may leave us with a singular  $\mathbf{N}_p$ , we can use a similar technique, if we know the pixel-domain pattern of the responsible modes. In this case, in analogy with the time stream case [Eq. (25)], we can write the map as

$$\mathbf{m}_p = \mathbf{s}_p + a \mathbf{v}_p + \mathbf{n}_p = \mathbf{s}_p + \mathbf{n}'_p. \quad (42)$$

Here,  $\mathbf{s}_p$  is the signal,  $\mathbf{n}_p$  is the noise, and  $\mathbf{v}_p$  describes the shape of the unknown mode, with unknown amplitude  $a$ , over which we will marginalize. Once again we can make an use of Eq. (41), with  $\mathbf{N}_p$  this time computed in the standard way.

As an example, the total offset of the map is spurious and undetermined (and it is often numerically convenient to set it to zero). That is, the detectors are only sensitive to temperature variations, rather than absolute temperatures. In the case of a lack of correlation between the noise and the underlying map (as we have assumed all along), the inverse of the noise correlation matrix computed for such a map should be singular by construction. The eigenvector corresponding to the zero eigenvalue is just a constant function of a pixel number, i.e.,  $\mathbf{v}_p^T \equiv [1, \dots, 1]^T$ . Knowing that, it is straightforward to perform the “inversion” as in Eq. (40). In this specific case such information can be included while computing the power spectra using the MADCAP package [29]. An analogous problem has been addressed by the Cosmic Background Explorer (COBE) Differential Microwave Radiometer (DMR) team [38]. This equation can be straightforwardly generalized for cases with more complicated eigenvectors as discussed below.

Equation (40) simply sets to zero an infinite eigenvalue of an inverse matrix. That formal procedure does not “solve” the problem of singularity; rather it gives a compact expression for the noise correlation matrix, which together with the knowledge of the singular modes provides complete information needed for statistically sound exploitation of the map.

If such a singularity is identified, it usually can be dealt with efficiently, producing a statistically valid result. Therefore, determination of singular modes present in the map has to be a part of a robust map-making procedure. Numerical inaccuracies often obscure singularities, making them difficult to find. The presence of such an undiscovered singular mode does not necessarily invalidate the outcome. In fact, in some applications, the final result can be still correct, while such a computation may accidentally duplicate the (approximate) numerical marginalization technique discussed by

Bond, Jaffe and Knox [18] and, e.g., implemented in MADCAP. However, it is still advisable to first determine the singular modes prior to applying such a method.

We also note in passing that this same formalism can be used to “marginalize over” other sorts of template amplitudes at the map stage, rather than the time stream templates considered earlier. An important example of this are templates corresponding to known sources of foreground emission, such as galactic dust, whose spatial morphology is well-known from studies at other wavelengths, and are subdominant but potentially important contaminants at CMB wavelengths [18].

For other easily identifiable singularities see the next section.

#### D. Combining maps of time-stream segments

The map-making formalism as presented so far can be applied to a number of statistically independent segments simultaneously. However, it may be advantageous to analyze each of these separately. In this case one needs to combine the separate segments together at the end. For Gaussian noise this would be quite straightforward, were it not for arbitrariness in the offset of each segment. The latter introduces possible relative offset shifts between the segments. This can be resolved for partially overlapping segments if we require them to display, within the noise uncertainty, the same underlying pattern in the common region. This introduces an extra complication to the well-known formula (e.g., [37]) for the optimal co-addition of two maps. The maximum-likelihood problem can be solved in a standard manner or using an approach analogous to that of Sec. IV B. In consequence, on defining, for each time-stream segment,  $\mathbf{u}_{p(I)}$  as a pixel-domain vector of ones and  $\hat{\mathbf{u}}_{p(I)} \equiv \mathbf{N}_{p(I)}^{-1} \mathbf{u}_{p(I)}$ , and introducing a corrected inverse noise correlation matrix,  $\mathbf{N}'_{p(I)}{}^{-1}$ , such as [cf. Eqs. (28),(41)],

$$\mathbf{N}'_{p(I)}{}^{-1} \equiv \mathbf{N}_{p(I)}^{-1} - \frac{\hat{\mathbf{u}}_{p(I)} \otimes \hat{\mathbf{u}}_{p(I)}^T}{\mathbf{u}_{p(I)}^T \hat{\mathbf{u}}_{p(I)}}, \quad (43)$$

we can express a final full map and a corresponding noise correlation matrix in a familiar manner,

$$\mathbf{N}_p^{\text{tot}} = \left\{ \sum_I \mathbf{N}'_{p(I)}{}^{-1} \right\}^{-1}, \quad (44)$$

$$\mathbf{m}_p^{\text{tot}} = \mathbf{N}_p^{\text{tot}} \sum_I \mathbf{N}'_{p(I)}{}^{-1} \mathbf{m}_{(I)}. \quad (45)$$

Here the sum is over maps of all segments to be combined. As discussed above the undetermined absolute offsets of each of the segments separately are reflected in the singularity of each of the redefined inverse noise correlation matrices,  $\mathbf{N}'_{p(I)}{}^{-1}$ . The final inversion in Eq. (44) again has to be understood as in Eq. (40) and the singularity needs to be accounted for in subsequent stages of the data analysis, as, for instance, shown in Eq. (41).

A MAXIMA-I based example of an application of this procedure is shown in Fig. 13. As in the case of a single con-

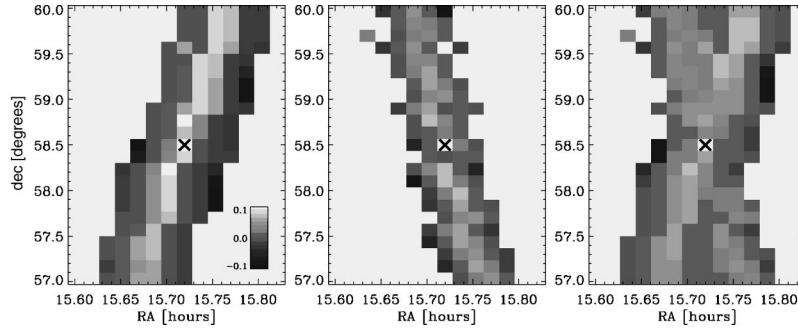


FIG. 13. Pixel domain correlations of the noise projected on the sky. All three panels show a level of the correlations relative to the rms value of the noise for the same pixel, which is marked with an “x.” Color coding is shown in left panel. This pixel has been observed twice during the MAXIMA-I flight. The noise correlations for the first observation are shown in the left panel, and these for the second one in the middle panel. Right panel shows the final co-added noise correlations [see Eq. (44)]. Clearly, due to the MAXIMA-I scanning strategy and the presence of the noise correlations in the time domain, the noise correlation pattern in pixel domain is highly anisotropic and strongly correlated as a result of any single observation of a pixel. The combined noise, for all pixels which were observed twice, is however significantly less correlated and more isotropic.

tinuous time stream segment [9], re-observing the same patch of the sky along the different scanning direction not only suppresses the level of the noise per pixel but also weakens and isotropizes the correlation pattern in pixel domain.

Note that the expression in the curly brackets of Eq. (44) becomes singular also, if not every segment map is connected (directly or through the number of intermediaries) with the others. If no such link exists the relative offset of such a segment map with respect to the rest of the map remains unknown giving rise to a singular mode in addition to the one related to the absolute offset. However, Eqs. (44) and (45) can still be applied if the inversion of the singular matrix is interpreted as in Eq. (40). The singular eigenvector  $\mathbf{v}_p$  has to now be appropriately replaced.

For instance, in a case of a single disjoint segment the two singular modes emerging as a result of the unknown total offset of the full map and a relative offset of the disjoint part can be chosen as having ones for every pixel belonging to one disjoint part and zeros elsewhere (or vice versa). With the inversion now described by the straightforward generalization of Eq. (40) for the case of multiple singular eigenvectors, the final product of the operation given by Eq. (45) can be a map composed of many disconnected regions with the uncertainty due to our ignorance of their relative offsets incorporated into the total noise correlation matrix,  $\mathbf{N}_p^{\text{tot}}$ , as given by Eq. (44).

Numerically one may encounter nearly singular cases whenever the overlapping region between two segments is too limited or the noise per pixel in the overlapping area too high to provide any useful constraint on the free offset. A practical and safe way of dealing with such a problem is to reject a (small by assumption) number of common pixels to make a given part genuinely disconnected and to account in a mathematically strict manner for the arising singularity of the noise correlation matrix.

A power spectrum of such an unconnected map can be subsequently computed as explained in Sec. IV C. However, that requires more involved algebra than that currently implemented in the MADCAP version of the quadratic estima-

tor. For that reason we have rejected all the disconnected segments from the MAXIMA-I maps while computing their power spectra. In that way the number of segments used for the final analysis decreased to 14 per detector [3,4].

#### E. Combining maps of different photometers

The additional difficulty here in comparison with the previous section is introduced by the possible relative calibration difference between maps produced for different detectors. Calibration uncertainty is generically difficult to be included into a maximum likelihood framework due to its multiplicative character. For the MAXIMA-I data set we have found that the relative calibration between different photometers' maps are with a high precision correct if the mean dipole based calibration is adopted for each of the maps [30]. Therefore rather than seeking a general solution to the problem, we combined maps as given by the Eq. (45) and used the largest single detector error for the calibration uncertainty of the combined map.

#### F. Low- $l$ aliasing

A potential bias of the final anisotropy power spectrum (but also for other statistics, see, e.g., [27]) resulting from this kind of the data analysis is due to an incomplete (and modest in a case of all balloon-borne experiments) sky coverage. That induces the correlation between otherwise uncorrelated (for a statistically isotropic sky)  $l$  modes. As a consequence, power contained in the low- $l$  modes beyond the detection capability can leak to the higher  $l$  modes which are to be estimated. Because the amplitude of the anisotropy power spectrum usually decays with increasing  $l$ —and is many orders of magnitude higher in monopole and dipole than in any other mode—there is a potential for biasing of the low- $l$  end of the estimated power spectrum.

One possible approach is to consider the unwanted modes as pixel-space templates as discussed above (Sec. IV C):  $\mathbf{v}_p(i_p) = Y_{lm}(i_p)$  for all the  $(l, m)$  modes to be marginalized over. ( $i_p$  is here a pixel number.) We then explicitly marginalize over them, while estimating, e.g., the anisotropy power

spectrum [Eq. (41)]. This is closely analogous to the time stream frequency marginalization of Sec. IV B.

Another, approximate, way to deal with the problem also can be applied directly on the power spectrum estimation stage is implemented in the MADCAP package [29,39] as described at the end of this section.

An alternative exact solution is based on Górski's idea [40]. In this approach the unwanted low- $l$  modes are removed from the map prior to further statistical analysis and the corresponding noise correlation matrix in pixel domain is appropriately corrected to account for the additional uncertainty. Unlike the other just-mentioned options this method produces a “cleaned” version of the sky map to be used henceforth.

Let us start by defining a scalar product of two functions  $f$  and  $g$  defined at each pixel  $i_p$  of our map  $\mathbf{m}_p$  as (hereafter  $\star$  stands for a complex conjugate),

$$(f|g) \equiv \frac{1}{n_{\text{pix}}} \sum_{i_p} f^*(i_p) g(i_p). \quad (46)$$

Also we denote by  $\{y\}$  a subset of the  $(l_0+1)^2$  spherical harmonics of the order not higher than  $l_0$  which are to be removed from the map. In general  $\{y\}$  is neither a linearly independent nor a complete basis on the map  $\mathbf{m}_p$ . However, we can construct a set of orthonormal functions  $\{\psi\}$  spanning the space of the spherical harmonics included in  $\{y\}$ . The construction can be performed by using Singular Value or Cholesky decomposition of the Kowalewski-Gram determinant of  $\{y\}$  functions (as, e.g., [40]) and rejecting all the singular modes. The resulting set of functions though orthonormal is clearly not complete. To achieve completeness we can supplement it with the functions from the another orthonormal (and complete) set of functions defined on the map and retain from the latter—through a standard Gram-Schmidt orthonormalization procedure—only those functions (or their linear combination) which are orthogonal to all the  $\{\psi\}$  functions. Practically, that part of the procedure can be encoded using singular value decomposition. A convenient choice of the extra complete functional basis is just a pixel basis,  $\{p\}$ ,

$$p_{(i)}(j_p) \equiv \delta^K(i, j_p), \quad \text{where } i, j_p = 1, \dots, n_{\text{pix}}. \quad (47)$$

On the successful completion of the entire procedure we end up with the set of the  $n_{\text{pix}}$  orthonormal functions. By construction, all the  $\{\psi\}$  functions are included in the final basis, and they span all the spherical harmonics of the order  $\leq l_0$  on the map  $\mathbf{m}_p$ . Hence all the remaining functions of the final basis (denoted hereafter  $\{\xi\}$ ), which are orthogonal to the functions  $\psi$  by construction, are also orthogonal to the all spherical harmonics of that order.

Algebraically the described procedure consists of a pair of linear transformations, which in a pixel representation can be written down as

$$\psi \equiv K y, \quad (48)$$

$$\xi \equiv L p. \quad (49)$$

Here  $K$  changes the basis from that of spherical harmonics,  $\{y\}$ , to  $\{\psi\}$ .  $L$  transforms the pixel basis,  $\{p\}$ , to the orthogonal basis made of basis vectors either parallel or perpendicular to  $\{\psi\}$  and retains only the latter subset,  $\{\xi\}$ . These functions complement  $\{\psi\}$  and both sets together form an orthonormal and complete basis  $\{\psi, \xi\}$  on the map, such as

$$(y|\xi) = 0. \quad (50)$$

The map purged of all the contribution from the low ( $\leq l_0$ ) order spherical harmonics is then given as

$$\mathbf{m}'_p = \sum_j (\mathbf{m}_p | \xi_{(j)}) \xi_{(j)}, \quad (51)$$

and is therefore uniquely represented by a vector  $\mathbf{m}_\xi$  defined as

$$\mathbf{m}_\xi(i) \equiv (\mathbf{m}_p | \xi_{(i)}). \quad (52)$$

The total (signal plus noise) correlation matrix for  $\mathbf{m}_\xi$  is then given as

$$\langle \mathbf{m}_\xi \otimes \mathbf{m}_\xi^T \rangle = [L(S_p + N_p)L^T], \quad (53)$$

where  $\langle \dots \rangle$  denotes an ensemble average.  $S_p$  and  $N_p$  are signal and noise correlation matrices computed for the complete map,  $\mathbf{m}_p$ , prior to any mode removal, i.e.,  $S_p + N_p \equiv \langle \mathbf{m}_p \otimes \mathbf{m}_p^T \rangle$ . If the sky signal contains only CMB anisotropy then the signal term,  $S_p$ , is given by

$$S_p = \frac{1}{4\pi} \sum_l P_l(2l+1) C_l, \quad (54)$$

where  $C_l$  is the power spectrum of the CMB fluctuations, and  $P_l$  denote matrices of the Legendre polynomials computed for all the pairs of the pixels in the given map, i.e.,

$$P_l(i_p, j_p) \equiv P_l[\cos(\chi(i_p) \cdot \chi(j_p))]. \quad (55)$$

Here  $\chi(i_p)$  is a unit vector pointing at the center of the pixel  $i_p$  and  $P_l$  a standard Legendre polynomial of an order  $l$ .

The correlation matrix of the map without low- $l$  modes, as defined in Eq. (53), can also be rewritten as a sum of a noise-like,  $N_\xi$ , and signal-like,  $S_\xi$ , term,

$$L[S_p + N_p]L^T = S_\xi + N_\xi. \quad (56)$$

Furthermore, on introducing the redefined Legendre polynomials matrices  $P'_l \equiv L P_l L^T$  and using Eq. (54), we can rewrite the signal part of Eq. (56) as

$$\begin{aligned} S_\xi &= \frac{1}{4\pi} \sum_l L P_l L^T (2l+1) C_l \\ &= \frac{1}{4\pi} \sum_l P'_l (2l+1) C_l. \end{aligned} \quad (57)$$

The last equation shows that the signal correlation matrix of the “cleaned” map is related to the angular power spectrum of the CMB anisotropies,  $C_l$ , in a way which is formally



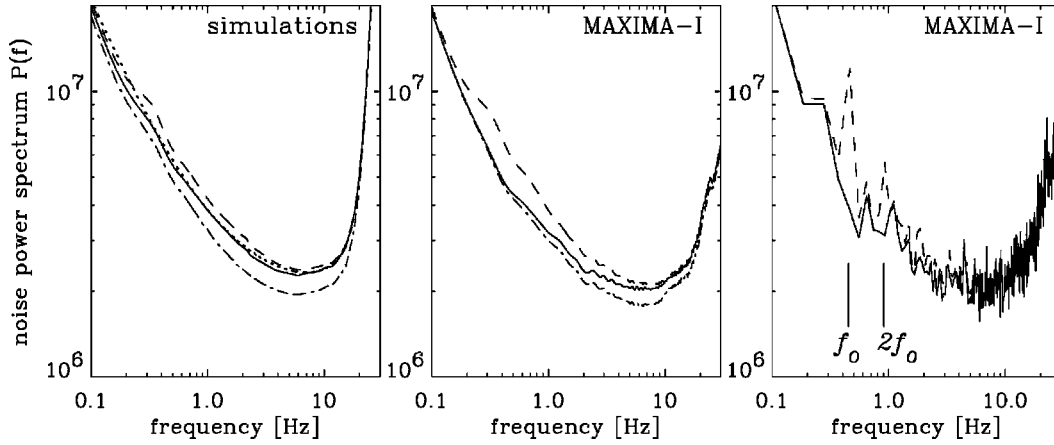


FIG. 14. Estimated noise power spectra for simulated and real data using the iterative approach of Ferreira and Jaffe [11]. In left panel, dashed line shows a noise power spectrum computed as in Sec. II C without iterative corrections. The dotted line depicts the noise spectrum used for a simulation, that almost perfectly overlaps with a solid line corresponding to a noise power spectrum calculated using 4 iterative steps and a map with 8 arcmin pixels. The dash-dotted line shows a result after 4 iteration but using a map with 3 arcmin pixels. That demonstrates the bias as described in the text. Middle panel shows noise power spectra for a single segment of the actual MAXIMA-I data. The power spectrum has been recovered using various approaches as in the left panel: the solid line shows the spectrum computed with 4 iterative steps and 8 arcmin pixels, dashed line shows the uniterated estimate, and dash-dotted line shows a result of 4 iterative steps with 3 arcmin pixels. The right panel shows noise spectra obtained just as a result of averaging on the second step of the noise estimation procedure, prior to any smoothing (see Sec. II C and the rightmost panel of Fig. 5). Dashed line corresponds to a spectrum obtained with no iteration, displaying characteristic two small spikes at  $\sim 0.45$  Hz and  $0.9$  Hz corresponding to a primary mirror chop frequency and its first harmonic. The overlaid solid line is a spectrum after 4 iteration using 8 arcmin pixels. Clearly in the latter spectrum the spikes are corrected as a result of iterations.

identical to that for the complete map, cf. Eq. (54). Therefore we can use usual quadratic estimator algorithms and, in particular, the MADCAP package [29,39] to estimate the CMB power spectrum having at disposition only the map [and the corresponding noise correlation matrix, Eq. (56)] with low- $l$  modes removed. To use that package in this case one just needs to replace matrices  $P_l$  (which constitute an intermediate output of MADCAP) with matrices  $P'_l$  and use the cleaned map in the  $\xi$  representation,  $m_\xi$ , and the noise correlation matrix,  $N_\xi$ , as an input, instead of usual  $m_p$  and  $N_p$ .

Though the matrices  $K$  and  $L$  can be chosen to be sparse (nearly triangular) the entire procedure is quite involved and the computational cost scales as  $\mathcal{O}(n_{pix}^3)$ . The same scaling applies to the computations of products of  $L$  and  $P_l$  matrices. However, while only the low  $l$  part of the angular power spectrum is likely to be affected by that kind of aliasing, the results of such a test applied to a map with relatively large pixels can provide a useful estimate of the size of the expected effect. In the MAXIMA-I case we applied this test to the map with 10 ft pixels (i.e., less than 4000 pixels in total) and compared it with the simple approximate template procedure of [18] as implemented in the MADCAP package of weighing out the monopole and dipole from the map, and introducing an extra low- $l$  bin to the recovered anisotropy power spectrum which is *a posteriori* rejected. In the MAXIMA-I case this bin extended from  $l=2$  up to  $l=34$ . Its rejection corresponds to the marginalization over that bin power and would be formally strict only if the likelihood for the power spectrum bin amplitudes were precisely Gaussian. Though such an approach is less general, because of these approximations and because it explicitly assumes the isotropy of the sky signal in

the rejected modes, we found no difference between the results of both methods. Therefore, we conclude that the MAXIMA-I final anisotropy power spectrum, in the published to date range of multipoles,  $35 \leq l \leq 1235$ , is not affected by any appreciable contribution due to aliasing of the low- $l$  mode power.

## V. ITERATIVE NOISE ESTIMATION

To date MAXIMA is one of the most sensitive experiments in terms of the noise level per measurement achieving the level of  $\sim 1500 \mu\text{K}$ , i.e.,  $100 \mu\text{K}\sqrt{s}$ , for some of the detectors. In spite of that the expected CMB and foreground signal for the observed patch of sky is still expected to be a subdominant part of a single measurement ( $\ll 10\%$  of the total power). The sky-related contribution to the power confined within some of the frequency bands can be, however, much higher ( $\leq 10\%$ ). Similarly, the non-stationary effects, though they may appear to be quite small, can be limited to a number of narrow frequency bands, dominating the power in there. For instance, in the MAXIMA-I case the primary mirror synchronous signal with its amplitude of  $200 \mu\text{K}$  dominates occasionally the power in the narrow frequency bands centered on the fundamental mode of primary mirror chop and its few lowest harmonics (see Fig. 14).

The important assumption behind the noise power spectrum estimation as presented in Sec. II C, i.e., that noise dominates the time stream measurements, though not clearly breached needs to be, therefore, tested.

To account for that effect we follow the iterative approach of [11]. It attempts to recover the noise component of the

entire time stream, which is subsequently used in the noise estimation procedure (see Sec. II C). The starting point of the iterative procedure is an approximation that  $\mathbf{n}_t^{(0)} \approx \mathbf{d}_t$ . Given this, it proceeds to the noise estimation and then to the map making. The resulting (zeroth order) map is used as an estimate of the signal in the time stream and is subtracted from the time stream on the next iterative step [11,16]. If we denote the noise contribution to the time stream on the  $i$ th step as  $\mathbf{n}_t^{(i)}$  then

$$\mathbf{n}_t^{(i)} = \mathbf{d} - \mathbf{A}\mathbf{m}^{(i-1)} - \mathbf{B}\mathbf{x}^{(i-1)} = \mathcal{A}^{(i-1)}\hat{\mathbf{m}}_p^{(i-1)}, \quad (58)$$

where  $\mathbf{m}^{(i-1)}$ ,  $\mathbf{x}^{(i-1)}$  and  $\hat{\mathbf{m}}_p^{(i-1)}$  are a map and a primary mirror signal and a generalized map respectively as estimated on the previous step. As shown by [11] the differences between the maps and noise correlation matrices estimated on the subsequent steps of the iteration decrease very quickly and the required precision is achieved rather rapidly. Usually we have found that at most four iterative steps were needed to reach the accuracy of few percent. The iterative approach can be significantly sped up if no explicit inversion of the pixel-pixel noise correlation matrix is performed on each iterative step but the map is calculated using an iterative linear system solver [14,16,41].

It is important to notice that the method is only asymptotically—i.e., in the limit of the large number of effective degrees of freedom—unbiased as guaranteed by its maximum likelihood origin. That limit is achieved, for instance, when a number of time samples increases, but a number of pixels is fixed.

The ensuing bias can be estimated as follows. Let us assume that we know the noise correlation function,  $N_t$ . Though the noise estimation goal seems to have been achieved, the iteration, as given by Eq. (58), can go on. The time stream estimate on the next step will contain noise only but composed of two components: a true time stream noise  $\mathbf{n}_t = \mathbf{n}_t^{(i)}$  and a pixel domain noise,  $\mathbf{n}_p$ , projected back to the time domain via the pointing matrix (and therefore also non-stationary):

$$\mathbf{n}_t^{(i+1)} = \mathbf{n}_t^{(i)} - \mathcal{A}\mathbf{n}_p = \mathbf{n}_t - \mathcal{A}\mathbf{n}_p. \quad (59)$$

The correlations of the latter quantity are computable and given by

$$\begin{aligned} \langle \mathbf{n}_t^{(i+1)}(i_t) \mathbf{n}_t^{(i+1)}(j_t) \rangle &= N_t(i_t, j_t) - \sum_{i_p, j_p} \mathcal{A}(i_t, i_p) \mathcal{N}_p(i_p, j_p) \\ &\quad \times \mathcal{A}^T(j_t, j_p). \end{aligned} \quad (60)$$

Here the summation is over all pairs of pixels and  $\mathcal{N}_p$  stands for the noise correlation matrix in the (generalized) pixel domain corresponding to the true noise correlation in time domain  $N_t$ .  $\langle \dots \rangle$  denotes an average over the statistical ensemble of the noise realizations.

Interestingly, due to existing correlations between noise in the time and pixel domain “adding” the extra noise to the time stream as described by Eq. (59) results in the underestimation of the actual noise power in the time domain. Though the above formula does not really help to unbiased the

procedure it gives a useful criterion of its applicability. For practical use it is useful to consider the overall power suppression due to the bias. Taking the trace of Eq. (60) we get

$$\begin{aligned} \sum_{i_t} \langle \mathbf{n}_t^{(i+1)}(i_t) \mathbf{n}_t^{(i+1)}(i_t) \rangle &= \sum_{i_t} N_t(i_t, i_t) \\ &\quad - \sum_{i_p} n_t(i_p) \mathcal{N}_p(i_p, i_p). \end{aligned} \quad (61)$$

Two limiting cases are evident. If each of the pixels is observed once then the noise iteration has no meaning, as we have no means to distinguish between the noise and the sky signal, and its result is fully biased, i.e., the left hand side of the above equation is zero. If there is no correlation in the time domain then pixel noise is just proportional to the number of times a given pixel has been observed. Inserting that into Eq. (61) renders

$$\frac{1}{n_s} \sum_{i_t} \langle \mathbf{n}_t^{(i+1)}(i_t) \mathbf{n}_t^{(i+1)}(i_t) \rangle = \sigma_t^2 \frac{(n_s - n_{pix})}{n_s}, \quad (62)$$

where  $\sigma_t^2$  is a diagonal element of  $N_t$ . This formula, which in the white noise case can be also derived directly from the maximum likelihood considerations, quantifies the introduced fractional bias as equal to  $n_{pix}/n_s$ . If the number of pixels is fixed then in the limit of the increasing number of measurement the bias disappears as expected.

If the correlations in the time streams are not negligible then a similar expression can be concocted with the number of pixels replaced by an “effective” number of the parameters  $n_{eff}$ , which is to be determined case by case,

$$\frac{1}{n_s} \sum_{i_t} \langle \mathbf{n}_t^{(i+1)}(i_t) \mathbf{n}_t^{(i+1)}(i_t) \rangle \approx \sigma_t^2 \frac{(n_s - n_{eff})}{n_s}. \quad (63)$$

The intuitive meaning of  $n_{eff}$  is clear from the expression

$$n_{eff} \equiv \sum_{i_p} n_s(i_p) \frac{\mathcal{N}_p(i_p, i_p)}{\sigma_t^2}. \quad (64)$$

In the MAXIMA-I case  $n_{eff}$  is not smaller than  $n_{pix}$ , and hence correlations tend to increase the bias of the noise iteration procedure.

It is important to notice that due to our assumption of stationarity, the bias depends only on how on average the measurements are divided between pixels and is therefore robust to the presence of the poorly sampled pixels.

From Eq. (61) it is clear that the bias is a result of the noise presence in the pixel domain. It is therefore advantageous to use all available information to minimize the noise of the estimated map, including all the time stream data during which a given patch of the sky has been observed.

In the MAXIMA-I case we have found (Fig. 14) that it is still advantageous to perform noise iterations for pixels as small as 8 arcmin. For smaller pixels, the resulting map is occasionally too noisy, and the likely bias larger than an

expected gain. Therefore, in those cases, we either restrict iterative corrections to deal only with the primary mirror synchronous signal or use noise power spectrum estimates obtained for the 8 arcmin pixels [4]. The latter approach is helpful in correcting for the low-angular-scale power, which should constitute the bulk of the sky signal present in the time stream. However, one may worry that it also introduces spurious correlations at the time lags corresponding to the characteristic crossing time of the big pixel. In practice, we have found that the results rendered by both these approaches are in very good agreement.

Clearly, invoking some kind of a map-denoising method and/or applying on this stage more aggressive filtering can be useful to extend the applicability of iterative noise estimation. We leave this issue for future research.

## VI. CONSISTENCY TESTS

A number of assumptions and approximations are involved in the computations of a map and a corresponding noise correlation matrix, so it is desirable to test the consistency of the final map-making products. Clearly that is difficult for the maps containing still-to-be-determined cosmological contributions without any prior assumptions. However, it can be done for the maps containing just noise. Those can be either the projections on the sky of the so called “dark” bolometers usually incorporated in experiments to track instrumental effects in the data [1], or just differences of the maps computed for single detectors. If, as expected, the maps recovered for a single photometer contain only the sky signal and the noise, the subtraction removes the sky component leaving a map of the noise only. Under the assumption of Gaussian time stream noise, the noise maps are also Gaussian, and their correlations are given by the noise correlation matrices produced in parallel by a map-making procedure (Sec. III). Any failure to meet such a requirement would suggest either a failure of the basic assumptions or some other problem with the data, the data analysis methods, or both.

The noise maps, due to correlations and inhomogeneity of the noise, are described by the multi-dimensional probability distributions with a number of dimensions equal to a number of pixels  $n_{pix}$  of a map under consideration. Therefore, notwithstanding the large number of pixels, a simple histogram of the pixel temperatures is likely to display significant deviations from the 1-dimensional Gaussian distribution even for the truly Gaussian case. To alleviate this problem we first “prewhiten” the map performing a linear transformation in order to decorrelate the measurements in the different pixels of the map

$$\mathbf{w}_p \equiv \mathbf{N}_p^{-1/2} \mathbf{m}_p. \quad (65)$$

Here  $\mathbf{N}_p^{-1/2}$  is a “square root” of the noise correlation matrix as estimated from the map-making, satisfying a relation

$$\mathbf{N}_p^{-1} \equiv \mathbf{N}_p^{-1/2} [\mathbf{N}_p^{-1/2}]^T. \quad (66)$$

In the following we take  $\mathbf{N}_p^{-1/2}$  to be a Cholesky triangular matrix (e.g., [36]). If the noise correlations are estimated

correctly (or at least consistently), then the components of the vector  $\mathbf{w}_p$  (a “prewhitened map”) are uncorrelated and their variances are equal to unity. Thus the multi-dimensional problem reduces to the well-defined one-dimensional test. Moreover, if the noise in the pixel domain is Gaussian then each of the components of  $\mathbf{w}_p$  is randomly drawn from the Gaussian 1-dimensional distribution with the unit variance. That is the hypothesis which we test. We apply a one dimensional Kolmogorov-Smirnov test. Its results give an estimate of how often the one point distribution function such as the one actually measured can be obtained from the Gaussian distribution with a unit variance. For MAXIMA-I we have four single-detector maps, giving us six independent difference maps. We test all of these noise maps, finding that the KS significance is always higher than  $\sim 10\%$  and usually as high as  $\sim 50\text{--}60\%$ , confirming the very good consistency of our map-making products. The sample of the results is also shown in Fig. 15. Clearly, a histogram of the unprewhitened map shows a significant deviation from the Gaussian curve near the peak (lower, left panel of the figure) and in the tails of the distribution. As anticipated, prewhitening largely resolves the discrepancies, allowing us to recover nearly perfectly Gaussian curves.

Though such KS-like tests usually provide a weak diagnostic, they are powerful consistency tests when passed. Sources of possible failure are abundant. Those can be either problems of the data set like an extra photometer-dependent parasitic signal in the time stream leaving its imprint in the difference maps, or non-Gaussianity of the time domain noise, or cross correlation between maps used for the creation of the difference maps. Moreover, the fact that one of our maps was actually made of the data obtained by the photometer centered at a higher frequency ( $\sim 240$  GHz) than the others frequency ( $\sim 150$  GHz) also suggests the lack of a substantial frequency-dependent sky (e.g., foreground) signal, as expected from the choice of the low contamination contrast patch observed during the MAXIMA-I flight. Alternatively, the source of the problem may lay with the data analysis, such as noise misestimation both in the time domain and pixel domain as a result of the involved approximations. In fact we have found that the results of KS tests depend on the precise map-making procedure we choose to apply to the real data. In particular, we derive somewhat different numbers if, e.g., no noise iteration has been performed, or no primary-mirror-synchronous signal has been removed. In both cases the differences are mainly due to the differences in the long wavelength modes present in the maps, which are the most susceptible to the details of the map-making algorithm. Nevertheless, that shows that the KS test possesses sensitivity which makes it a useful tool in tracking realistic problems in the maps and/or procedure. The KS test can be also applied to the maps with the low- $l$  multipoles removed as described in the previous section. Hence the lack of an indication of the problem with the KS test results is quite encouraging and may serve as a fairly comprehensive validation of the final results.

In addition, for every pair of the noise maps we can consider a two-dimensional probability distribution of two noise maps calculated as differences of two actual single detector

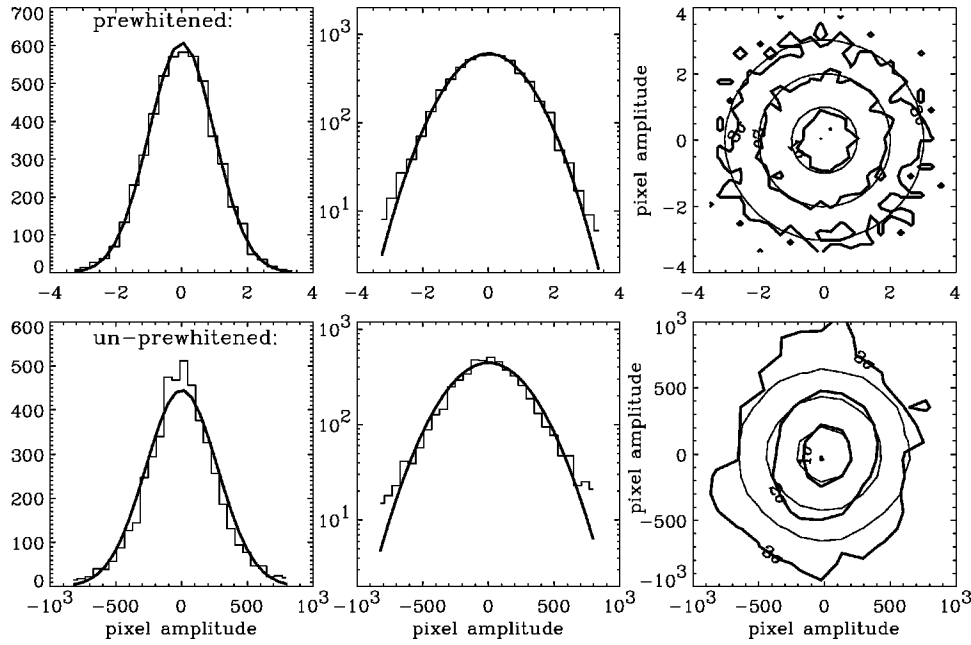


FIG. 15. Probability distribution functions of the pixel amplitudes for the noise maps computed as a difference of two maps made for two different photometers. The upper row shows the results using prewhitened maps. Their corresponding unprewhitened versions are displayed in the lower row. Histograms show the results for the real data and smooth lines the Gaussian curves with a dispersion equal to one (upper panels) or fitted to best match the actual results. The two right panels show contour plots of the histograms of joint two-dimensional probability distributions of the pixel temperature for two “noise” maps. These were computed as a difference of the sky maps recovered for two independent pairs of the photometers. The lower panel shows the results for the actual maps with the correlated pixel noise, and upper panel for the maps prewhitened prior to histogramming. Concentric ellipses marked with thin solid lines show the Gaussian expectation. Those were computed assuming no correlations between both maps and were discretized the same way as data histograms and setting the dispersions either to unity (upper panel) or to the best fit values (lower panel).

maps. Here we use a difference map of the first and second detector,  $\mathbf{m}_{p(12)}$ , and of the third and fourth,  $\mathbf{m}_{p(34)}$  and consider the probability distribution  $\mathcal{P}(\mathbf{m}_{p(12)}, \mathbf{m}_{p(34)})$ .

If we assume that the signal detected by different detectors is uncorrelated, the correlation matrix for a pair  $(\mathbf{m}_{p(12)}, \mathbf{m}_{p(34)})$  has a block diagonal structure with blocks given by the correlation matrices of each of the difference map separately. Prewhitening in such a case is simply given by

$$\begin{bmatrix} \mathbf{w}_{p(12)} \\ \mathbf{w}_{p(34)} \end{bmatrix} \equiv \begin{bmatrix} N_{p(12)}^{-1/2} & \mathbf{0} \\ \mathbf{0} & N_{p(34)}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{m}_{p(12)} \\ \mathbf{m}_{p(34)} \end{bmatrix}. \quad (67)$$

The two dimensional probability distribution  $\mathcal{P}(\mathbf{w}_{p(12)}, \mathbf{w}_{p(34)})$  is then bound to have a unit variance and the Gaussian shape unless the cross-correlation term is indeed present or any of the previously mentioned reasons or assumptions is not satisfied. The obtained results (right hand panels of Fig. 15) seem to agree well with the expectations and therefore confirming the consistency of our analysis.

## VII. SUMMARY

Recent CMB data sets have set new challenges for the data analysis. Problems are related to the sheer size of new data sets and also to the quality of the analysis tools, which can make a full use of increasing power of data.

This paper describes how those challenges were met in

the analysis of the MAXIMA-I data set. The successful production of the final results, as published in the recent papers [3,4], required us to improve on the existing and develop and test new methods and tools. Though some of them had to be significantly customized to be efficient, the others seem to be of a more general character and applicability extending to data sets as big as that of the forthcoming satellite missions [5,6].

We have focused here on time ordered data manipulation techniques and map-making algorithms. We have presented a comprehensive, consistent approach allowing us to recover a map of the sky and to estimate its error matrix in realistic circumstances of an actual CMB experiment. The highlights include:

- the time stream noise estimation procedure coupled together with the gap filling method through constrained noise realization;

- an exact version of the map making code;

- the statistically sound methods of dealing with the time stream filtering of contaminated frequencies and time-domain templates.

The entire suite of the methods presented here amounts to a self-consistent approach to building a final map out of smaller parts in an efficient way. The construction can be halted when the largest map lending itself to the statistical analysis (e.g., power spectrum estimation) has been built. Dealing with only subsets of an entire data set seems to be the only efficient way of searching for, understanding and



removing any systematic problems.

CPU time-wise, the presented methods are limited by the noise correlation matrix inversion, which requires  $\mathcal{O}(n_{pix}^3)$  operation. This is the price to be paid if no assumption is made about possible symmetries present in the map and a pixel-pixel noise correlation matrix is required. This obstacle may not be insurmountable, even without sacrificing the generality of the approach. In most of the realistic situations such a matrix is expected to be rather sparse. Moreover, from the knowledge of a scanning strategy and characteristic correlation length in the time domain, it can be guessed *a priori* which matrix elements need to be computed. We leave both issues for further investigation, noting only here that extensive use of a supercomputer and the MADCAP package facilitates computations of the maps containing up to  $\sim 40000$  [4].

## ACKNOWLEDGMENTS

R.S. and S.H. acknowledge support of NASA Grant NAG5-3941. A.H.J. and J.H.P.W. acknowledge support from NASA LTSA Grant No. NAG5-6552 and NSF KDI Grant No. 9872979. P.G.F. acknowledges support from the RS. B.R. and C.D.W. acknowledge support from NASA GSRP Grants No. S00-GSRP-032 and S00-GSRP-031. R.S. also acknowledges help (in Poland) of Polish State Committee for Scientific Research Grant No. 2P03D01719. Computing resources were provided by the National Energy Research Computing Center at Lawrence Berkeley National Laboratory. MAXIMA is supported by NASA Grant NAG5-4454 and by the NSF through the Center for Particle Astrophysics at UC Berkeley, NSF cooperative agreement AST-9120005.

- 
- [1] A. T. Lee *et al.*, Proceedings of the 3K cosmology meetings, 1999, astro-ph/9903249.
  - [2] <http://cfpa.berkeley.edu/maxima>
  - [3] S. Hanany *et al.*, *Astrophys. J. Lett.* **545**, L5 (2000).
  - [4] A. T. Lee *et al.*, *Astrophys. J. Lett.* **561**, L1 (2001).
  - [5] <http://map.gsfc.nasa.gov/>
  - [6] <http://astro.estec.esa.nl/SA-general/Projects/Planck/>
  - [7] E. L. Wright, astro-ph/9612006.
  - [8] M. Tegmark, *Astrophys. J. Lett.* **480**, L87 (1997).
  - [9] M. Tegmark, *Phys. Rev. D* **56**, 4514 (1997).
  - [10] S. P. Oh, D. N. Spergel, and G. Hinshaw, *Astrophys. J.* **510**, 551 (1999).
  - [11] P. G. Ferreira and A. H. Jaffe, *Mon. Not. R. Astron. Soc.* **312**, 89 (2000).
  - [12] R. Stompor *et al.*, Proceedings of the Garching meeting “Mining the sky,” 2000, astro-ph/0012418.
  - [13] G. Hinshaw, Proceedings of the Garching meeting “Mining the sky,” astro-ph/0011555.
  - [14] O. Doré, R. Teyssier, F. R. Bouchet, D. Vibert, and S. Prunet, *Astron. Astrophys.* **374**, 358 (2001).
  - [15] P. Natoli, G. deGasperi, C. Gheller, and N. Vittorio, *Astron. Astrophys.* **372**, 346 (2001).
  - [16] S. Prunet, C. B. Netterfield, E. Hivon, and B. P. Crill, astro-ph/0006052.
  - [17] M. Tegmark, *Phys. Rev. D* **55**, 5895 (1997).
  - [18] J. R. Bond, A. H. Jaffe, and L. Knox, *Phys. Rev. D* **57**, 2117 (1998).
  - [19] J. D. Borrill, *Phys. Rev. D* **59**, 027302 (1999).
  - [20] I. Szapudi, S. Prunet, D. Pogosyan, A. S. Szalay, and J. R. Bond, *Astrophys. J. Lett.* **548**, L115 (2000).
  - [21] B. Wandelt, E. Hivon, and K. M. Górski, astro-ph/0008111.
  - [22] O. Doré, L. Knox, and A. Peel, *Phys. Rev. D* **64**, 083001 (2001).
  - [23] E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen, astro-ph/0105302.
  - [24] P. G. Ferreira, J. Magueijo, and K. M. Górski, *Astrophys. J. Lett.* **503**, L1 (1998).
  - [25] J. H. P. Wu, *Phys. Rev. Lett.* **87**, 251303 (2001).
  - [26] J. H. P. Wu *et al.*, astro-ph/0104248, 2001.
  - [27] L. Cayón *et al.*, *Mon. Not. R. Astron. Soc.* **326**, 1243 (2001).
  - [28] <http://cfpa.berkeley.edu/borrill/MADCAP>
  - [29] J. D. Borrill, astro-ph/9911389.
  - [30] R. Stompor *et al.* (in preparation).
  - [31] A. Balbi *et al.*, *Astrophys. J. Lett.* **545**, L5 (2000).
  - [32] R. Stompor *et al.*, *Astrophys. J. Lett.* **561**, L7 (2001).
  - [33] J. H. P. Wu *et al.*, *Astrophys. J., Suppl. Ser.* **132**, 1 (2001).
  - [34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, England, 1992).
  - [35] Y. Hoffman and E. Ribak, *Astrophys. J. Lett.* **380**, L5 (1991).
  - [36] H. Golub and F. van Loan, *Matrix Operations* (Johns Hopkins University Press, Baltimore, 1987).
  - [37] A. de Oliveira-Costa *et al.*, *Astrophys. J. Lett.* **509**, L77 (1998).
  - [38] E. L. Wright, G. Hinshaw, and C. L. Bennett, *Astrophys. J. Lett.* **458**, L53 (1996).
  - [39] J. D. Borrill, P. G. Ferreira, A. H. Jaffe, and R. Stompor, Proceedings of the Garching meeting “Mining the sky,” 2000.
  - [40] K. M. Górski, *Astrophys. J. Lett.* **430**, L85 (1994).
  - [41] S. Prunet *et al.*, Proceedings of the Garching meeting “Mining the sky,” 2001, astro-ph/0101073.
  - [42] We have omitted the extra (prewhitening) filter matrix,  $\mathbf{D}$ , introduced in [9] to reduce the number of floating point operations in the matrix multiplications of Eq. (14). In fact the band diagonality of “prewhitened” matrices (i.e.,  $\mathbf{M}' \equiv \mathbf{DMD}^T$  and  $\mathbf{N}_i'^{-1} \equiv \mathbf{D}\mathbf{N}_i^{-1}\mathbf{D}^T$ ) can be exploited only if the products in Eq. (15) are performed from outside inwards, and though that may speed up computation of  $\mathbf{A}^T\mathbf{M}'$  the computational cost of the subsequent products, i.e.,  $(\mathbf{A}^T\mathbf{M}')\mathbf{N}_i'^{-1}$  and  $(\mathbf{A}^T\mathbf{M}')\mathbf{N}_i'^{-1}(\mathbf{A}^T\mathbf{M}')^T$ , offsets any advantages on the earlier stage.